Daoliang Li
Yingyi Chen
(Eds.)

Computer and Computing Technologies in Agriculture V

5th IFIP TC 5/SIG 5.1 Conference, CCTA 2011
Beijing, China, October 2011
Proceedings, Part II

2 **Part II**

🜨 Springer

IFIP Advances in Information and Communication Technology 369

IFIP – The International Federation for Information Processing

IFIP was founded in 1960 under the auspices of UNESCO, following the First World Computer Congress held in Paris the previous year. An umbrella organization for societies working in information processing, IFIP's aim is two-fold: to support information processing within ist member countries and to encourage technology transfer to developing nations. As ist mission statement clearly states,

> *IFIP's mission is to be the leading, truly international, apolitical organization which encourages and assists in the development, exploitation and application of information technology for the benefit of all people.*

IFIP is a non-profitmaking organization, run almost solely by 2500 volunteers. It operates through a number of technical committees, which organize events and publications. IFIP's events range from an international congress to local seminars, but the most important are:

- The IFIP World Computer Congress, held every second year;
- Open conferences;
- Working conferences.

The flagship event is the IFIP World Computer Congress, at which both invited and contributed papers are presented. Contributed papers are rigorously refereed and the rejection rate is high.

As with the Congress, participation in the open conferences is open to all and papers may be invited or submitted. Again, submitted papers are stringently refereed.

The working conferences are structured differently. They are usually run by a working group and attendance is small and by invitation only. Their purpose is to create an atmosphere conducive to innovation and development. Refereeing is less rigorous and papers are subjected to extensive group discussion.

Publications arising from IFIP events vary. The papers presented at the IFIP World Computer Congress and at open conferences are published as conference proceedings, while the results of the working conferences are often published as collections of selected and edited papers.

Any national society whose primary activity is in information may apply to become a full member of IFIP, although full membership is restricted to one society per country. Full members are entitled to vote at the annual General Assembly, National societies preferring a less committed involvement may apply for associate or corresponding membership. Associate members enjoy the same benefits as full members, but without voting rights. Corresponding members are not represented in IFIP bodies. Affiliated membership is open to non-national societies, and individual and honorary membership schemes are also offered.

Daoliang Li   Yingyi Chen (Eds.)

Computer and Computing Technologies in Agriculture V

5th IFIP TC 5/SIG 5.1 Conference, CCTA 2011
Beijing, China, October 29-31, 2011
Proceedings, Part II

Springer

Volume Editors

Daoliang Li
Yingyi Chen
China Agricultural University
China-EU Center for Information & Communication Technologies (CICTA)
17 Tsinghua East Road, P.O. Box 121, Beijing, 100083, P.R. China
E-mail: {dliangl, chenyingyi}@cau.edu.cn

Preface

I would like to express my sincere thanks to all authors who submitted research papers to support the 5th International Conference on Computer and Computing Technologies in Agriculture (CCTA 2011) held in Beijing, China, during October 29–31, 2011.

This conference was hosted by China Agricultural University; the IFIP TC5 Special Interest Group (SIG) on Advanced Information Processing for Agriculture (AIPA); National Natural Science Foundation of China; and China-EU Centre for Information and Communication Technologies (CICTA).

Proper scale management is not only a necessary approach in agro-modernization and agro-industrialization but also required for the development of agricultural productivity. Therefore, the application of different technologies in agriculture has become especially important. 'Informatized agriculture' and the 'Internet of Things' are hot research topics in many countries aiming to scientifically manage agriculture to yield high incomes with low costs. CICTA covers the research and development of advanced and practical technologies applied to agriculture and promotes international communication and cooperation; it has successfully held five International Conferences on Computer and Computing Technologies in Agriculture since 2007.

The topics of CCTA 2011 covered a wide range of the interesting theory and applications of all kinds of technology in agriculture, including the Internet of Things; simulation models and decision-support systems for agricultural production; agricultural product quality testing; traceability and e-commerce technology; the application of information and communication technology in agriculture; and universal information service technology and service system development in rural areas.

We selected the 189 best papers among all those submitted to CCTA 2011 for these proceedings. The papers are divided into three themes. It is always exciting to have experts, professionals and scholars with creative contributions getting together and sharing some inspiring ideas and hopefully accomplishing great developments in high-demand technologies.

Finally, I would like also to express my sincere thanks to all authors, speakers, Session Chairs and attendees, from home and abroad, for their active participation and support of this conference.

October 2011   Daoliang Li

Conference Organization

Sponsors

China Agricultural University
The IFIP TC5 Special Interest Group (SIG) on Advanced Information
   Processing for Agriculture(AIPA)
National Natural Science Foundation of China

Organizers

China-EU Center for Information and Communication Technologies in
   Agriculture (CICTA)

Chair

Daoliang Li

Conference Secretariat

Lingling Gao

Table of Contents – Part II

GIS, GPS, RS and Precision Farming

Study on Soil Nutrients Spatial Variability in Yushu City

Yueling Zhao[1], Haiyan Han[2], Liying Cao[1], and Guifen Chen[1,*]

[1] Jilin Agricultural University, Changchun, Jilin 130118
[2] Changchun University, Changchun, Jilin 130022
zyueling@163.com

Abstract. In order to help farmers to understand soil nutrients, use effectively our soil and protect our environment, at the same time to improve sustainable development of the agricultural system, the topsoil samples were collected form yushu city .The ways of combining Geostatistics with GIS were applied to analyze the spatial variability of soil nutrients, such as Alkaline hydrolysis nitrogen, available phosphorus and available potassium. All results shows that Alkaline hydrolysis nitrogen, available phosphorus and available potassium in the coefficient of variation were from 9% to 46%, and the largest coefficient of variation of available phosphorus was 46%. By using Semi-variogram function of statistical analysis, Alkaline hydrolysis nitrogen, available phosphorus and available potassium show Moderate spatial dependency. The relevance of space about available phosphorus may be the weakest, which is effected mainly by random factors. So the different management and the level of fertilizer may be done according the spatial variability of different area format.

Keywords: Geostatistics, soil nutrients, spatial variability, statistical analysis.

The soil is non-uniform complex. The nutrient properties already receive the restriction from natural environment condition, and influenced by artificial actions. The numerous researches indicated that the soil characteristic value have the obvious difference in the different space position. In the 1970s, Burgess and others used statistical method in soil science research area, they had overcome classics Fisher the statistical idea in the research soil space changeability aspect insufficiency [1,2]. In recent years, with the popularization and development of all kinds of technology, more and more scholars considered, the research soil question from the different angles. Many researchers study the soil nutrient spatial distribution and the management by using GPS, GIS and so on correlation technology, some progress has been gain in this field.

The maize filed of yushu city is the main study objective .by using the technology of land statistical method and the model of Geostatistical Analysis in ARCGIS9.2, soil characteristic of spatial variability were studied, which analysis the variability and dependence of soil nutrients properties. The conclusions can sever to precision agriculture, adjust to management manners, protect our resources.

---

1   Materials and Methods

1.1   Study Area Description

The study area, yushu city is located in the northern of Jilin province (44°30′57″-45°15′02″N, 126°01′44″E - 127°05′09″E), The climate in yushu city belongs to the warm temperate zone, half moist continental monsoon climate, featuring four distinct seasons. Which has cold, long winters and short, mild summer. Yushu city has a mean annual temperature of 4.6℃-5.6 ℃ and 500-mm annual rainfall. Soil parent materials in the city are composed of black soil.
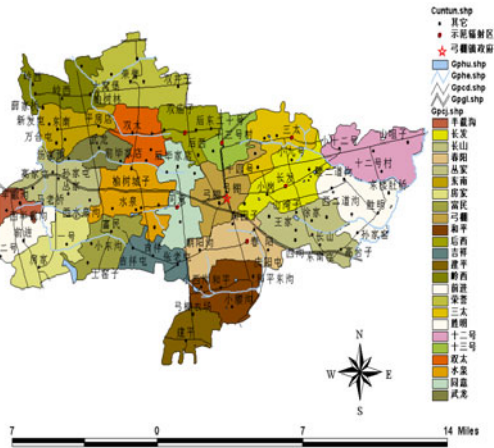
Fig. 1. Location of soil sampling distribution

1.2   Soil Sampling and Analyses of Nutrients

A total of 121 topsoil samples were collected in 2008(Fig.1).All sample sites were recorded using Different Global Positioning System (AgGPS-132),the plot has an extension of 1600m2(80m×80m).All collected samples were put a cool, dry, well ventilated, not special smell and the dust pollution   place.

Alkaline hydrolysis nitrogen was gain by using Alkaline hydrolysis diffusion process; available phosphorus in soil was extracted with 0.5 mol L-1 NaHCO3 and determined using the molybdenum blue colorimetric method; available potassium in soil was extracted with 1mol L-1 NH4OAC and determined using a flamephotometer.

If you have more than one surname, please make sure that the Volume Editor knows how you are to be listed in the author index.

1.3   Geostatistical Analysis

The soil was divided into many unform regions in traditional statistic. It describes some soil nutrients by Calculating   the mean, standard deviation, the variance,

coefficient variation and so on of some samples. Geostatistical analysis, amied to describe the variance between the point values sampled in the field.The variable of a region is the key to geostatistical analysis. The main application of geostatistics in soil science has been the estimation and mapping of soil attributes out of sampled areas [3]. Kriging is a linear geostatistical interpolation technique that provides a best linear unbiased estimator for quantities that vary in space. Kriging estimates are calculated as weighted sums of the adjacent sampled concentrations. If data appear to be highly continuous in space, the points closer to those estimated receive higher weights than those that are farther away [4]. Semivariance analysis was used to estimate the range over which samples of the soil nutrient variables were related .The following is Semi-variogram formulate.

$$r(h) = \frac{1}{2n} \sum_{i=1}^{n} \left[z(xi) - z(xi+h) \right]$$

Where Z(xi)and Z(xi+h)are experimental measures of any two points separated by the vector h ,and n is the number of experimental pairs separated by h.

1.4   Data Treatment with Computer Software

Raw data were analyzed with different software packages. the descriptive parameters and the probability analyses were calculated with spss for window(version 16.0) .All maps were produced using GIS software ArcMap(version 9.2)and its spatial analyst and geostatistical analysis extensions.

2   Results and Discussion

2.1   Descriptive Parameters of the Raw-Data Set

The coefficient of variation (CV) can show the spatial variability in soil science. Normally , if the   CV is ≤0.1 ,the variable was considered weakly dependent; if the CV is between 0.1amd 1,the variable was considered moderately dependent; and if the CV≥1,the variable was considered strongly dependent. The main statistical characteristics of the field data set are reported in table. The CV in the city is between 9% and 46%. The CV of variable phosphorus is the biggest, it gets to 46%. The reason is different farmer have different management Strategies, for example some farmer like to use much more fertilizer, others do not like that.

The CV of the Alkaline hydrolysis nitrogen in the city is 11%, which can be moderately variable. The CV of the available potassium in the city is 9%, which can be weak variable. The reason is that Alkaline hydrolysis nitrogen and available potassium move easy in soil and they abstract easy by plant.

Table 1. Descriptive statistics of soil nutrients

Items （mg..g-1）	Sample points	Min	Max	Kurtosis	Skewness	Means	Median	S.D	C.V
Alkaline hydrolysis N	118	103.	189.2	-0.15	-0.11	144.9	145.8	16.9	0.11
Available P	118	5.8	44.46	9.74	2.63	13.08	11.56	6.44	0.49
Available K	118	80	138	0.68	-0.19	112.91	112	10.5	0.09

2.2   The Test of Soil Nutrient Content Normality

Kriging is based on normal distribution in geostatistics in soil science. If some data are not normal distribution, some manners must be taken to change them to be normal distribution or close to normal distribution. Because of the reason, the soil nutrients information must be test [5.6]. Kolmogorov-Smirnov(K-S) test for goodness –of-fit was performed to test the normality of the selected soil property distributions. The Alkaline hydrolysis nitrogen, available phosphorus and available potassium were all normally distributed. Their value of Asymp.Sig. (2-tailed) are 0.97,0.08,0.16, respectively. Because they are all bigger than 0.05,the data set under test is regarded as following a normal distributions. In order to show clearly their normal distribution, some histograms can be drawn by spss software (v 16.0). Histograms of soil nutrients properties with a normal distribution curve are shown in Fig.2-4 (n=118).
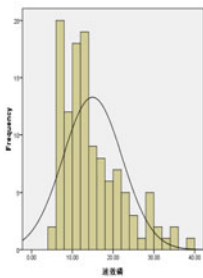
Fig. 2. Histograms of N   **Fig. 3.** Histograms of P   **Fig. 4.** Histograms of K
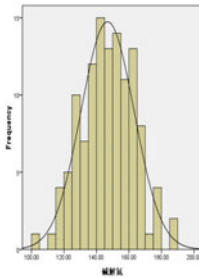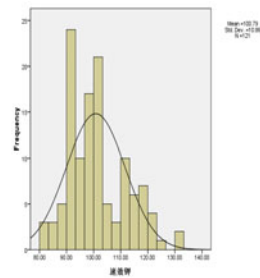
2.3   Semi-variogram Analysis

To further explore the nature of soil variability, a geostatistical analysis was executed based on Geostatistics. The table 2 shows some main parameters.C0 is the nugget, which express the variable condition when Semi-variogram is in zero point. Partial

sill is the sum of C0 and C, which express the total variable in the system including the constitutive variation and the random variation. Partial sill is higher, the total variation is higher. The nugget to sill ratio enables comparison of the relative magnitude of the nugget effect among soil properties [7], especially if sampled at similar scales. If the nugget to sill ratio was <25%,the variable was considered strongly spatially dependent; if the ratio was between 25%-75% ,the variable was considered moderately spatially dependent; and if the ratio was >75% the variable was considered weakly spatially dependent. Based on the nugget-to-sill ratio, soil nutrient properties in Jilin province yushu city, the spherical model is used well to show the spatial structure of Alkaline hydrolysis nitrogen, available phosphorus and available potassium. The nugget-to-sill ratio are 72%, 26%, 61%.They shows the nutrient properties are moderate spatial dependence .the structure and random factors make the nutrients properties variable and show some spatial variability.

Table 2. Best-fitted semi-variogram models of soil nutrients and corresponding parameters

Soil properties	Theoretic model	Nugget C_0	Partial sill C_0+C	C_0/C_0+C	Range(m)
Alkaline hydrolysis nitrogen （ $mg.kg^{-1}$ ）	球状模型	72.213	214.96	0.73	12.0762
AP （ $mg..kg^{-1}$ ）	球状模型	11.192	49.146	0.26	26.0162
AK($mg..kg^{-1}$)	球状模型	100.65	163.89	0.61	12.59

2.4   Analysis of Soil Spatial Distribution Maps

Information generated through semi-variogram was used to calculate sample weighing factors for spatial interpolation by a Kriging procedure. In order to model spatial variability and delineate spatial distribution of soil nutrients, geostatistical variogram analysis and Kriging estimation, which provides the best linear unbiased prediction at un-sampled locations, have been widely used [8]. Nutrients concentration estimates can be used to plan spatially variable fertilizer applications [9,10]. The maps of alkaline hydrolysis nitrogen (mg.kg-1), available phosphorus (mg.kg-1), available potassium (mg.kg-1)were generated by ARCGIS(v.9.2) software in the study. Spatial patterns of Alkaline hydrolysis N and Available K are similar over the study area .they are block shape. The spatial distribution of available P show a strip shape. Understanding of soil nutrient spatial variability and possible areas of soil nutrient deficiency in the city are important because it can serve as a basis for planning management strategies to improve crop yields and reduce environmental impact, particularly, which will benefit the sustainability of the agricultural system.

3   Conclusions

The results shows that Alkaline hydrolysis nitrogen, available phosphorus and available potassium in the coefficient of variation were from 9% to 46%, and the largest coefficient of variation of available phosphorus was 46%. Alkaline hydrolysis nitrogen and available potassium are normally distributed. available phosphorus is nearly normally distributed, which has very big Kurtosis and Skewness. The phenomenon show the data is not normally distributed well. But the soil nutrients properties are in good condition, there are more stronger Potential in next years.

The constitutive factors and the random factors results to soil spatial variability .By using Semivariance function of statistical analysis, Alkaline hydrolysis nitrogen, available phosphorus and available potassium show Moderate spatial dependency. The contribution of constitutive factors in the following order: available phosphorus > available potassium> Alkaline hydrolysis nitrogen. The variability of available phosphorus is mainly effected by constitutive factors. Because the character of phosphorus moving slow in soils. The relevance of space about available phosphorus may be the weakest, which is effected mainly by random factors. The different management and the level of fertilizer may be done according the different spatial variability of area.

We can get the spatial distributions of Alkaline hydrolysis nitrogen, available phosphorus and available potassium generated form their semi-variograms .There are not same nutrient distribution in different farmland form the Kriging maps by GIS. The prediction maps of Alkaline hydrolysis nitrogen, available phosphorus and available potassium were generated using ordinary Kriging methods .The different samples exist different spatial distributions in the maize fields of yushu city, but there are a certain principle in the samples. For example, some show strip shape, other show block shape. Nutrients concentration estimates can be used to plan spatially variable fertilizer applications and plan different management strategies.

References

1. Rodríguez, A., Durán, J., Fernández-Palacios, J.M., Gallardo, A.: Spatial variability of soil properties under Pinus canariensis canopy in two contrasting soil textures. Plant Soil 322, 139–150 (2009)
2. Brocca, L., Morbidelli, R., Melone, F., et al.: Soil moisture spatial variability in experimental areas of central Italy. Journal of Hydrology 333, 356–373 (2007)
3. Goovaerts, P.: Geostatistics in soi science: state-of-the-art and perspectives. Geoderma 89(1-2), 1–45 (1999)
4. Gressie, C.: The origins of Kriging Math. Geol. 22(2), 239–252 (1990)

5. Yang, Z.J., Wei, J.S., He, P., et al.: Spatial variance of potassium in rubber plantation soil in danzhou city of Hainan province. Journal of Northwest Forestry University 25(5), 41–44 (2010)
6. ZhaoL, M., Shi, X.Z., Huang, Y., et al.: Influential factors of spatial heterogeneity of soil nutritious in taihu lake region soils. Soils 40(6), 1008–1012 (2008)
7. Zhao, J., Zhang, J.M., Ming, K., et al.: Spatial heterogeneity of soil nutrients in blacksoil,China —a case study at hailun county. Bulletin of Soil and Water Conservation 24(6), 53–57 (2004)
8. Zhang, M., He, P.F., Chen, W.Q.: Spatio-temporal variability analysis of soil nutrients based on GIS and geostatistic. Journal of Northeast Agricultural University 41(3), 53–58 (2010)
9. Lark, R.M., Ferguson, R.B.: Mapping risk of soil nutrient deficiency or excess by disjunctive and indicator kriging. Geoerama 118, 39–53 (2004)
10. Bai, Y.L., Jin, J.Y., Yang, L.P., et al.: Variability of Soil Nutrients in Field and Fertilizer Recommendation. Plant Nutrition and Fertilizer Science 7(2), 129–133 (2001)

The Application of Classical Least Square Algorithm in the Quantitative Analysis of Lime in Wheat Flour by ATR-MIR Spectroscopy[*]

Dong Wang[1,2,3,**], Donghai Han[1], Zhihong Ma[3], Ligang Pan[3], Liu Zhao[3], Ping Han[3], and Jihua Wang[2,3,***]

[1] College of Food Science and Nutritional Engineering,
China Agricultural University, Beijing 100083, P.R. China
wangd@nercita.org.cn
[2] Agricultural Information Technology,
Beijing Research Center for Information Technology in Agriculture,
Beijing 100097, P.R. China
[3] Beijing Research Center for Agri-food Testing and Farmland Monitoring,
Beijing 100097, P.R. China
wangjh@nercita.org.cn

Abstract. In this thesis, classical least square (CLS) regression was applied in the attenuate total reflection mid-infrared (ATR-MIR) spectra data processing of lime in wheat flour to develop the corrected peak height model of the raw spectra, the corrected peak area model of the raw spectra, the corrected peak height model of the 2nd derivative spectra and the corrected peak area model of the 2nd derivative spectra respectively. The result indicated that the correlation coefficients of the four models mentioned above are 0.9648, 0.9696, -0.9646 and -0.9599 respectively. *F-test* result indicated that a very remarkable correlation exists between the estimated and specified values of the calibration set and external validation set. The detection limits of the four models mentioned above are 3.51 %, 3.21 %, 3.51 % and 3.69 % respectively, which can fulfill the demand for the rapid quality safe screening of wheat flour in the market. This method, to some extent, can provide some references for not only the design and manufacturing of the special MIR instrument for the quality safe control of wheat flour in the market but also the quantitative determination of banned additives in wheat flour.

Keywords: Classical Least Square, Attenuate Total Reflection, Mid-Infrared Spectroscopy, Wheat Flour, Lime.

1   Introduction

Wheat flour is one of the important food sources for human beings, of which the quality safety is highly focused by all quarters of society. However, the quality safety incidents about wheat flour have been reported frequently in recent years, such as lime added in it, which results in not only negative influence but also being harmful to the customers' health. Therefore, the development of rapid, simple, practical and low-cost determination method for detecting lime in wheat flour becomes particularly urgent and very important.

Among the common instrument analytical methods, mid-infrared (MIR) spectroscopy analysis is widely applied in the fields of the identification and characterization of microbes[1], detection and identification of bacteria[2], quality analysis of edible oil [3, 4], quantitative determination of anthocyanins in cherry[5] with the advantages of distinguishable peaks, strong characteristics, high sensitivity, rapid analysis and non-invasive analysis.

However, the traditional MIR analysis involves the sample preparation such as tabletting with KBr, which is difficult for the operators. Attenuated total reflection (ATR) technology can acquire the information of samples according to the reflective signals of the sample surface, which is free of sample preparation. ATR technology has the advantages of high sensitivity, distinguishable peaks, user-friendly control and non-interference by humidity. Based upon the advantages mentioned above, ATR technology is now being widely applied in the fields of analysis of juice sweetness[6], biology structure studies[7], material study[8] and so forth. Therefore, it is convenient to collect the MIR spectra of powder samples, such as wheat flour, by ATR technology.

It is known by the experiment data that lime, such as calcium oxide, calcium hydroxide, has the characteristic absorption peaks in the MIR range, which is the theoretical foundation of the qualitative and quantitative analysis to it. The characteristic absorption peak in MIR range of wheat flour is far away from that of lime, i.e. the interference from wheat flour is not very obvious. Based on the MIR spectral characteristics of wheat flour and lime, CLS algorithm has a certain potential application in the quantitative analysis of lime in wheat flour by MIR spectra data.

Classical least square (CLS) regression is a data processing method applied in the data regression of analytical chemistry commonly[9, 10]. The basic principle of CLS is that to calculate the evaluated values of the regression coefficients between the independents and dependents according to a group of obvious values in order to decrease the errors between the specified values and the predicted values furthest. The obvious errors of CLS obey three hypotheses, i.e. the obvious values are independent each other with the same precision, the obvious errors are non-systematical, the obvious errors are as the law of normal distribution[11]. Compared to partial least square regression and principal component regression, CLS regression is easy to understand relatively. Meanwhile, it is easier to achieve CLS algorithm on the special MIR detection instruments, which is in favor of the promotion of this technology.

In this research, ATR-MIR spectra of the wheat samples mixed with lime were collected. The quantitative determination of lime in wheat flour by ATR-MIR

combined with CLS algorithm was studied. The corrected peak height / area models of the raw spectra and the corrected peak height / area models of the 2nd derivative spectra were developed respectively. The external validation set was employed to evaluate the prediction performance of the models. This method is of great importance to the rapid quantitative screening of lime in wheat flour. This research, to some extent, will provide some references to not only the design and manufacturing the special MIR spectrometer for the quality safety screening of wheat flour, but also the quantitative determination of other banned additives in wheat flour.

2   Experiment

2.1   Instrument and Materials

Instrument: FT-MIR spectrometer, type: Spectrum 400, Perkin Elmer, U.S..

The ATR-MIR spectra are collected by FT-MIR spectrometer combined with ATR accessory, of which, the reference spectrum is collected by air, the wavenumber range is 4000 cm^{-1} - 540 cm^{-1}, the wavenumber resolution is 4 cm^{-1}, the time of accumulation is 240. The reference spectrum is collected once every 30 minutes.

Materials: wheat flour milled by the winter wheat from Changping experimental base of Beijing, without any additives. Calcium oxide and calcium hydroxide: A.R..

2.2   Method

The quantitative calibration models of lime are developed by the corrected peak height and corrected peak area of lime in ATR-MIR spectra combined with the algorithm of classical least square. The external validation set, of which the average and standard deviation are similar to the calibration set, is employed to evaluate the prediction performances of the models. The contents of lime in both the calibration set and external validation set are distributed in the range of 0.6 % - 16.0 % uniformly. The mass of each sample is about 50 grams. The mass data of wheat flour and lime in each sample are recorded according to analytical balance accurately. The information of the content of lime in the samples is showed in table 1.

Table 1. The information of the content of lime in the wheat flour samples (%)

Items	Calibration set	External validation set
Sample size	30	10
Minimum	0.6831	0.9643
Maximum	15.8145	14.1465
Average	5.2319	5.3556
Standard deviation	4.36	4.49

In order to decrease the error of peak height and peak area measurement, the corrected peak height / area models of the 2nd derivative spectra are discussed despite the corrected peak height / area models of the raw spectra. The 2nd derivative [12] is one of the data preprocessing method that is applied in the data preprocessing in MIR data commonly, which can eliminate the baseline drifting and baseline translation.

The corrected height is the distance between the peak top of the target peak and the connecting line of the two base points around the target peak, while the corrected area is the integral value between the spectrum curve and the connecting line of the two base points around the target peak. The sketch maps of the measurement of corrected peak height and corrected peak area of the raw and the 2nd derivative spectra are showed as figure 1.

| A1 | A2 | B1 | B2 |
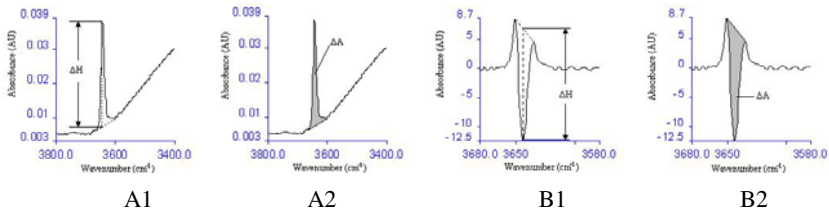
Fig. 1. The sketch maps of the measurement of the corrected peak height (A1) and the corrected peak area (A2) of the raw spectra and the corrected peak height (B1) and the corrected peak area (B2) of the 2nd derivative spectra

Correlation coefficient (*r*) and root mean square error of calibration (*RMSEC*) are regarded as the parameters to evaluate the models' performance; while standard error of prediction (*SEP*) and ratio performance deviation (*RPD*) are regarded as the parameters to evaluate the models' prediction performance; the statistic *F-test* value is regarded as the parameter to evaluate the regression significant of regression equation.

Looking up in the *F-test* value table according to the sample size of the calibration set and external validation set, $F_{(1,28)} = 7.64$, $F_{(1,8)} = 11.26$ when the significance level $\alpha = 0.01$.

3   Result and Discussion

3.1   Development of Calibration Models

Some of the ATR-MIR spectra of the samples from calibration set are showed as figure 2, from which, it can be seen clearly that the characteristic peak of lime exists at 3642 cm^{-1}.

The calibration models of lime in wheat flour are developed by the content of lime, i.e. the specified values, in wheat flour and the corrected peak height of the raw spectra data, the corrected peak area of the raw spectra data, the corrected peak height of the 2nd derivative spectra data and the corrected peak area of the 2nd derivative

spectra data respectively. The sample size of the calibration set is 30. The correlations between the specified values and the corrected peak height of the raw spectra, the corrected peak area of the raw spectra, the corrected peak height of the 2nd derivative spectra and the corrected peak area of the 2nd derivative spectra are showed as figure 3, while the regression equations, correlation coefficients (*r*) and *F-test* values (*F*) of the four calibration models are showed in table 2.
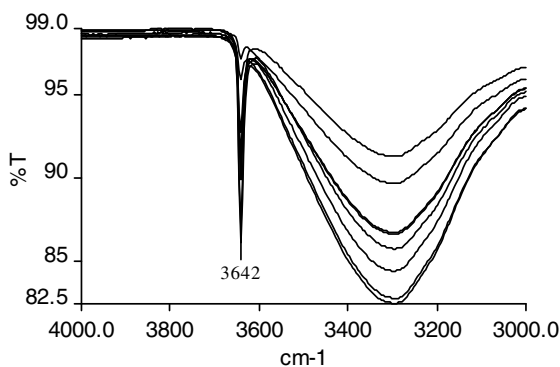
Fig. 2. Some of the ATR-MIR spectra of calibration set

A1   A2   B1   B2
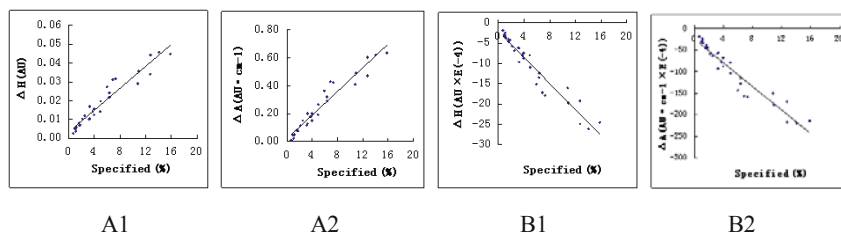
Fig. 3. The correlations between the specified values and the corrected peak height of the raw spectra (A1), the corrected peak area of the raw spectra (A2), the corrected peak height of the 2nd derivative spectra (B1) and the corrected peak area of the 2nd derivative spectra (B2)

Table 2. The regression equations, correlation coefficients (*r*) and *F-test* values (*F*) of the calibration models

Models	Regression equation	*r*	*F*
Corrected peak height of raw spectra	y = 0.0029x + 0.0034	0.9648	377
Corrected peak area of raw spectra	y = 0.0426x + 0.0183	0.9696	440
Corrected peak height of 2nd derivative spectra	y = -1.6070x - 2.1049	-0.9646	375
Corrected peak area of 2nd derivative spectra	y = -13.6530x - 25.6610	-0.9599	328

It can be seen from table 2 that the absolute values of the correlation coefficients of the models are no less than 0.95; the *F-test* values of the models are all much larger than that looked up in *F-test* value table ($F_{(1,28)}$ = 7.64 when the significance level α = 0.01), which indicated that a very remarkable correlation exists between the specified values and the corrected peak heights / areas of the raw spectra or the corrected peak heights / areas of the 2nd derivative spectra, which indicates that CLS regression has a certain potential application in the quantitative analysis of lime in wheat flour by ATR-MIR spectra data.

The correlations between the specified values and estimated values by self-prediction of the models are showed as figure 4.
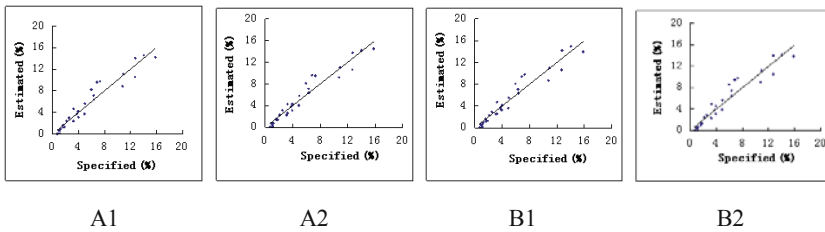
| A1 | A2 | B1 | B2 |

A1: the corrected peak height model of the raw spectra, A2: the corrected peak area model of the raw spectra, B1: the corrected peak height model of the 2nd derivative spectra, B2: the corrected peak area model of the 2nd derivative spectra

Fig. 4. The correlations between the estimated and specified values of the calibration set

It can be seen from figure 4 that, for each model, a good correlation exists between the specified values and estimated values by self-prediction of the model. The regression equations, correlation coefficients (r_c), root mean square errors of calibration (*RMSEC*), *F-test* values (F_c) of the four groups of data by self-prediction are showed as figure 3.

Table 3. The regression equations, correlation coefficients (r_c), *RMSEC* and *F-test* values (F_c) of the calibration set by self-prediction

Models	Regression equation	r_c	RMSEC	F_c
Corrected peak height of raw spectra	y = 1.0020x + 0.0061	0.9649	1.17	378
Corrected peak area of raw spectra	y = 0.9981x + 0.0198	0.9699	1.07	444
Corrected peak height of 2nd derivative spectra	y = 0.9988x + 0.0106	0.9647	1.17	376
Corrected peak area of 2nd derivative spectra	y = 0.9947x + 0.0497	0.9606	1.23	334

It can be seen from figure 3 that for all of the regression equations by self-prediction, the slops are all close to 1, the intercepts are close to 0, the correlation coefficients are no less than 0.96, the *F-test* values are much larger than that looked up in *F-test* value table ($F_{(1,28)}$ = 7.64 when the significance level α = 0.01), which indicates that, for each model, a very remarkable correlation exists between the specified values and estimated values by the self-prediction of the calibration model, i.e. CLS regression has a certain potential application in the quantitative analysis of lime in wheat flour by ATR-MIR spectra data.

3.2   External Validation

The external validation set is employed to evaluate the prediction performance of the calibration models. The sample size of the external validation set is 10. The correlation between the specified values and estimated values by the corrected peak height model of the raw spectra, the corrected peak area model of the raw spectra, the corrected peak height model of the 2nd derivative spectra, the corrected peak area model of the 2nd derivative spectra are showed as figure 5.
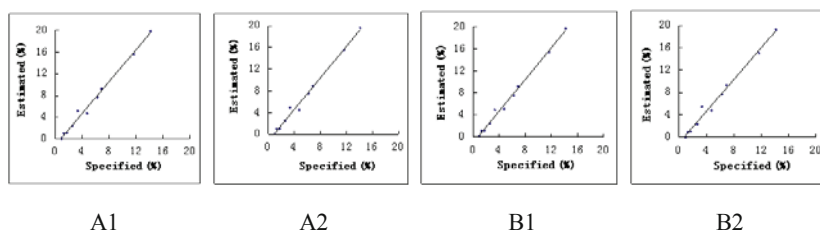
| A1 | A2 | B1 | B2 |

A1: the corrected peak height model of the raw spectra, A2: the corrected peak area model of the raw spectra, B1: the corrected peak height model of the 2nd derivative spectra, B2: the corrected peak area model of the 2nd derivative spectra

Fig. 5. The correlations between the estimated and specified values of the external validation set

It can be seen from figure 5 that a good correlation exists between the estimated and specified values of the external validation set for each model. The regression equations, correlation coefficients (r_p), standard errors of prediction (*SEP*), ratio performance deviations (*RPD*) and the *F-test* values (F_p) of the external validation set are showed in table 4.

It can be seen from table 4 that for all the regression equations of external validation set, the correlation coefficients between the estimated and specified values are no less than 0.99, standard errors of prediction are no more than 2.6, ratio performance deviation are no less than 1.7, *F-test* values are much larger than that looked up in *F-test* value table ($F_{(1,8)}$ = 11.26 when the significance level α = 0.01), which indicates that a very remarkable correlation exists between the estimated and specified values of the external validation set for each model, i.e. the CLS regression has a certain potential application in the quantitative prediction of lime in wheat flour by ATR-MIR spectra data.

Table 4. The regression equations, correlation coefficients (r_p), *SEP*, *RPD* and *F-test* values (F_p) of the external validation set

Models	Regression equation	r_p	SEP	RPD	F_p
Corrected peak height of raw spectra	y = 1.4658x -1.1709	0.9954	2.59	1.73	862
Corrected peak area of raw spectra	y = 1.4471x - 1.1580	0.9954	2.47	1.81	862
Corrected peak height of 2nd derivative spectra	y = 1.4425x - 1.0632	0.9964	2.48	1.81	1119
Corrected peak area of 2nd derivative spectra	y = 1.4224x - 1.0096	0.9935	2.42	1.85	607

Comparing the four models, it can be seen that the differences among the four models are not significant in regression equations, correlation coefficients and the prediction results of the external validation set, the reason of which is that the characteristic peak of lime in ATR-MIR spectrum (3642 cm^{-1}) is far away from the absorption peak of wheat flour in ATR-MIR spectrum (3300 cm^{-1}), therefore, the interference from the spectral background becomes weak. According to that the detection limit is three times the *RMSEC*[11], the detection limits of the corrected peak height / area models of the raw spectra and the corrected peak height / area models of the 2nd derivative spectra are 3.51 %, 3.21 %, 3.51 % and 3.69 % respectively, which are much less than that added in wheat flour for the purpose of weight increment. This method can fulfill the demand of rapid quality screening of wheat flour in the market.

4   Conclusion

Classical least square algorithm was applied in the development of the quantitative calibration models of lime in the wheat flour samples. The result indicated that the CLS regression has a certain potential application in the quantitative determination of lime in wheat flour by ATR-MIR spectra data. The detection limits of the corrected peak height / area models of the raw spectra and the corrected peak height / area models of the 2nd derivative spectra are 3.51 %, 3.21 %, 3.51 % and 3.69 % respectively, which are much less than that added in wheat four for the purpose of weight increment. This method can fulfill the demand of rapid quality screening of wheat flour in the market, which is of great and positive meaning to the monitoring of wheat flour in the market and the guarantee of the customers' health. This research, to some extent, will also provide some references to the quantitative determination of other banned additives and the design and manufacturing of the special MIR spectrometer for the quality screening of wheat flour in the market. The method to decrease the detection limit is still being researched.

References

1. Santos, C., Fraga, M.E., Kozakiewicz, Z., Lima, N.: Fourier transform infrared as a powerful technique for the identification and characterization of filamentous fungi and yeasts. Research in Microbiology 161, 168–175 (2010)
2. Al-Qadiri, H.M., Lin, M., Cavinato, A.G., Rasco, B.A.: Fourier transform infrared spectroscopy, detection and identification of Escherichia coli O157:H7 and Alicyclobacillus strains in apple juice. International Journal of Food Microbiology 111, 73–80 (2006)
3. Gurdeniz, G., Ozen, B.: Detection of adulteration of extra-virgin olive oil by chemometric analysis of mid-infrared spectral data. Food Chemistry 116, 519–525 (2009)
4. Maggio, R.M., Kaufman, T.S., Del Carlo, M., Cerretani, L., Bendini, A., Cichelli, A., Compagnone, D.: Monitoring of fatty acid composition in virgin olive oil by Fourier transformed infrared spectroscopy coupled with partial least squares. Food Chemistry 114, 1549–1554 (2009)
5. Pappas, C.S., Takidelli, C., Tsantili, E., Tarantilis, P.A., Polissiou, M.G.: Quantitative determination of anthocyanins in three sweet cherry varieties using diffuse reflectance infrared Fourier transform spectroscopy. Journal of Food Composition and Analysis 24, 17–21 (2011)
6. Jha, S.N., Gunasekaran, S.: Authentication of sweetness of mango juice using Fourier transform infrared-attenuated total reflection spectroscopy. Journal of Food Engineering 101, 337–342 (2010)
7. Iconomidou, V.A., Georgaka, M.E., Chryssikos, G.D., Gionis, V., Megalofonou, P., Hamodrakas, S.J.: Dogfish egg case structural studies by ATR FT-IR and FT-Raman spectroscopy. International Journal of Biological Macromolecules 41(1), 102–108 (2007)
8. Ebbesen, S.D., Mojet, B.L., Lefferts, L.: In situ ATR-IR study of nitrite hydrogenation over Pd/Al2O3. Journal of Catalysis 256(1), 15–23 (2008)
9. Li, J., Zhang, M., Li, Q.: Spectrophotometric determination of cefotaxime sodium with potassium ferricyanide. Chinese Journal of Applied Chemistry 28(1), 88–93 (2011) (in Chinese)
10. Yu, J., Hu, S., Lin, Y., Luo, X., Zhu, T.: Determination of tungsten in rare earth-magnesium master alloy by spectrophotometry with thiocyanate and acid digestion. Chinese Journal of Spectroscopy Laboratory 27(6), 2529–2532 (2010)
11. Lu, W., Yuan, H., Xu, G., Qiang, D.: Modern Near-infrared Spectroscopy Analysis Technology, pp. 124–182. China Petrochemical Press (2000) (in Chinese)
12. Shiroma, C., Rodriguez-Saona, L.: Application of NIR and MIR spectroscopy in quality control of potato chips. Journal of Food Composition and Analysis 22, 596–605 (2009)

Study of Quantitative Analysis for Moisture Content in Winter Wheat Leaves Using MSC-ANN Algorithm

Hao Ma[1], Haiyan Ji[1,*], Xue Liang[1], and Zhenhong Rao[2]

[1] Key Laboratory of Modern Precision Agriculture System Integration Research,
Ministry of Education, China Agricultural University, Beijing, China 100083
mah85@126.com, instru@cau.edu.cn
[2] College of Science, China Agricultural University, Beijing, China 100083

Abstract. Reflectance spectra of winter wheat leaves specimens was acquired with portable spectroradiometer and integral sphere, after pretreatment with the method of multiplicative scatter correction(MSC), the principal components calculated were used as the inputs of artificial neural networks to build the Back--Propagation artificial neural networks model(BP-ANN), which can be used to predict moisture content of winter wheat leaves very well. In the article we made a study of quantitative analysis for moisture content of winter wheat leaves in booting and milk stage. The correlation coefficient(r) of predicted set in booting stage was 0.918，the standard deviation(SD) was 0.995 and the relative standard deviation(RSD) was 1.35%. And in milk stage r = 0.922, SD = 2.24, RSD = 3.37%. The model can truly predict the content of water in winter wheat leaves. Compared with the classical method, the artificial neural networks can build much better predicted model.

Keywords: MSC, ANN, Moisture Content, Reflectance spectrum, Quantitative analysis.

1   Instruction

The moisture content is one of the most important factors to effect photosynthesis, respiration and biomass in plant[1-3]. Shortage of moisture would effect the physiological process, biochemical process and morphostucture of plant, and further effect the growing, output and quality of crops[4]. For the past few years there are many flood and drought disasters arising, which make the crop cutting into large area. Therefore, the scientific management of moisture content in crops increasingly becomes one of the most important measures in crop production.

Presently, the moisture content of large area crops has been researched by means of remote sensing, which indicates the deficit of moisture by measuring the temperature of plant canopy[5]. Another common method is to establish an inverse model about reflectance of leaves and moisture content, which bases on the characteristic of moisture

---

* Corresponding author.

absorbing infrared light. From the 80's, many famous scholars have proved that: the spectral reflectance of plant be used to detect status of moisture content is feasible[6-11]. In this paper we would investigate the method of processing spectrum to promote the result of modeling.

Multiplicative scatter correction(MSC) is a very effective spectrum preprocessing method of dealing with living leaves of plant. It is a multi-variable scatter correction technique[12]. The original spectrum processed with the method can effectively eliminate the scatting caused by the physical factors, enhance the spectral information related to the tested ingredients, reduce the impact of baseline translation and drift between samples caused by the scatting, and can greatly improve the signal-to-noise ratio of correlative spectra.

In this paper we gathered the reflectance spectra of a new collection of winter wheat leaves, then analyzed the spectral data with the method of principal component analysis after the spectra preprocessed with the method of MSC, and finally made study of the model with the training system of artificial neural network. From the result of modeling, we can get the better result by the model of MSC-ANN, especially dealing with the living leaves in the field.

2   The Algorithm

2.1   The Algorithm of MSC

The use of MSC firstly requires to create a "ideal spectrum" of the tested samples, which is used to correct the others as a standard spectrum[13]. Actually, we usually calculate the average spectrum of all the samples as the standard spectrum, operate each sample spectrum with the method of unary linear regression to get their regression constant and regression coefficients, and then correct the original spectra of each sample to improve the signal-to-noise ratio. The concrete process of arithmetic is as follow:

(1) The average spectrum:

$$\overline{A}_{i,j} = \frac{\sum_{i=1}^{n} A_{i,j}}{n}$$

(2) Unary linear regression:

$$A_i = \overline{A}_i m_i + b_i$$

(3) MSC:

$$A_{i(MSC)} = \frac{(A_i - b_i)}{m_i}$$

In the formula, i is the number of samples, j is the number of wavelength.

2.2   The Algorithm of ANN

The neural network developed in recent years is a very active interdiscipline, which refers to many disciplines such as biology, electronics, computers, mathematics, physics and so on, and it has a extensive future[14]. Artificial neural network is an effective means of researching spectroscopy, which is widely applied in the study of spectrum and NMR. Currently the most active model is the back propagation(BP) algorithm model, proposed by Rumelhart[15], which involves parallel distributed processing theory and multi-layer network. BP neural network model is a powerful learning system to achieve the nonlinear mapping between input and output in high degree[16], and it has proved the model can approximate to any continuous nonlinear curves[17]. The preprocessor spectra is operated with the method of principal component analysis to get appropriate principal components, which are used as the input of multi-layer feedforward neural network, and finally establishing the training model of BP neural network.

3   Experiment

3.1   Test Method and Instrument

we collected respectively 40 leaves as the samples in booting and milk stages of winter wheat, and then measured their reflectance spectra by using the FieldSpec@Pro FR portable spectrometer, which produced by the U.S. ASD company, and 1800 integrating sphere that produced by LI-COR company. The true values of moisture content were gathered by drying method[18]. We first surveyed the fresh weight of the leaves with AB104N precision electronic balance, and put the leaves in 105℃ to dry until their weight were constant, the moisture content is as follow:

Moisture content (%) = (fresh weight - dry weight) / fresh weight × 100%

3.2   The Information of Sample Spectra

Scanned range is from 400nm to 2100nm, the data are collected in each nanometer of the spectral region, so we can get 1700 data points from each spectrum. Actually, each sample was scanned 5 times, the average spectrum is calculated to be used as the true
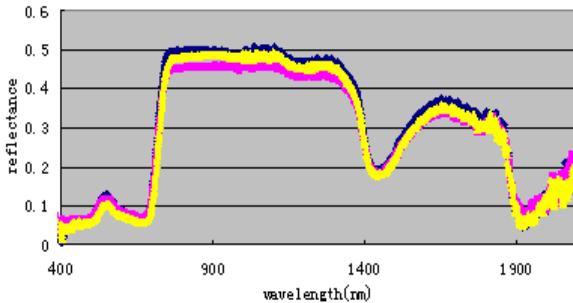
Fig. 1. Reflectance spectra of winter wheat leaves

spectrum of sample. That in figure 1 is three spectrograms choosed. From the figure we can see there are three characteristic peaks in each spectrum, 1160nm, 1440nm and 1930nm, especially the peak point of the 1440nm is the most notable.

4   The Results of Experiment and Analysis

4.1   Modeling for Winter Wheat Leaves in Booting Stage and Prediction Results

The reflectance spectrum is used for modeling ranging from 1431nm to 1470nm. Two distinct different samples are removed from the 40 samples collected and the remaining 38 samples are divided into two parts, 30 samples are put into calibration set and the others into test set. The number of principle component is setted to 5. The parameters of neural network model are as follows:

```
net.trainParam.lr=0.04; % the learning rate of training is setted to 0.04.
```

```
net.trainParam.mc=0.2;  % the momentum coefficient is setted to 0.2.
```

```
net.trainParam.epochs=20;  % the training epochs is setted to 20.
```

```
net.trainParam.goal=0.4;% the error of training objective is setted to 0.4.
```

After several rounds of training, we can get a better prediction network. Modeling for the calibration set, we can get the relationships between predicted and true values of moisture content of winter wheat leaves in booting stage from figure 2. The correlation coefficient r = 0.958, the standard deviation of forecast SD = 0.337, the relative standard deviation RSD = 0.45%. Detecting the data of test set, we can get the relationships between predicted and true values of moisture content in the model of test set from figure 3. r = 0.918, SD = 0.995, RSD = 1.35%.
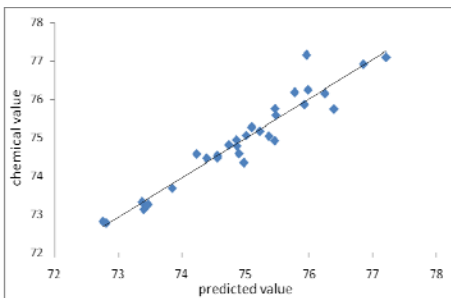
Fig. 2. Relationships between predicted and ture values of calibration set in booting stage
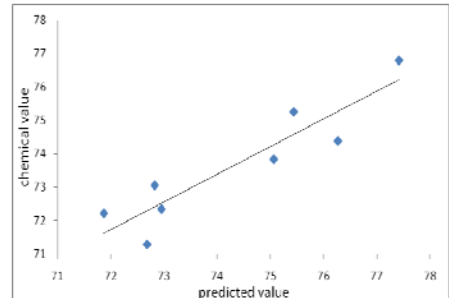
Fig. 3. Relationships between predicted and ture values of predicted set in booting stage

4.2   Modeling for Winter Wheat Leaves in Milk Stage and Prediction Results

The reflectance spectrum is used for modeling ranging from 1411nm to 1480nm. The 40 samples are divided into two parts, 30 samples in calibration set and the others in test set. The number of principal component is setted to 5. The parameters of neural network model are as follow:

```
net.trainParam.lr=0.04; % the learning rate of training is setted to 0.04.

net.trainParam.mc=0.2;  % the momentum coefficient is setted to 0.2.

net.trainParam.epochs=20; % the training epochs is setted to 20.

net.trainParam.goal=1.0; % the error of training objective is setted to 1.0.
```

After several rounds of training, we can get a better prediction network. Modeling for the calibration set, We can get the relationships between predicted and ture values of moisture content of winter wheat leaves in milk stage from figure 4, r = 0.963, SD = 0.949, RSD = 1.43%. Detecting the data of test set, we can get the relationships between predicted and ture values of moisture content in the model of test set from figure 5. r = 0.922, SD = 2.24, RSD = 3.37%.
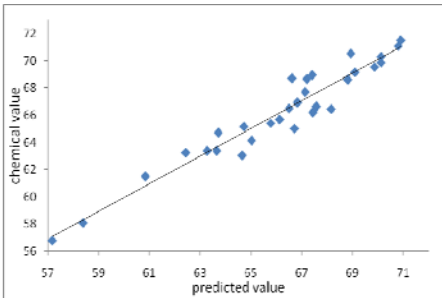
Fig. 4. Relationships between predicted and ture values of calibration set in milk stage
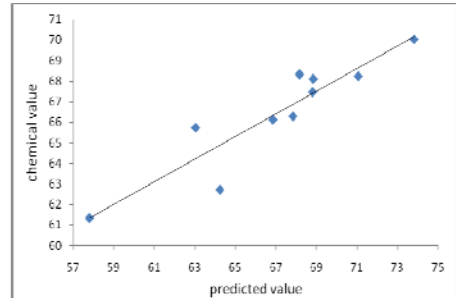
Fig. 5. Relationships between predicted and ture values of calibration set in milk stage

4.3   Error Analysis and Model Evaluation

There are several factors in source of the error. The first reason is that the smoothness, cleanliness of leaves and the external environment changing make a great influence on the reflectance spectra. And in addition, water lost in the process of measurement and hard to achieve constant weight in drying attribute to another reasons. These measurement errors accumulate eventually to cause differences in the results of the quantitative analysis. However, the artificial errors can be minimized by MSC-ANN algorithm. From the figure of prediction model, we can see that the prediction values

are very close to the chemical values measured by standard method. So it is feasible to quickly determine the moisture content of winter wheat leaves by NIR reflectance spectroscopy and the model of MSC-ANN algorithm.

5   Conclusion

In this paper, a new modeling method for quantitative analysis moisture content in winter wheat leaves is studied: Multiplicative scatter correction - artificial neural network(MSC-ANN). In the research, we establish the quantitative analysis model of moisture content in different growing seasons of winter wheat leaves, and then make an analysis of the model. By analyzing the experimental data, MSC-ANN algorithm in dealing with chemical and biological living components is better than the traditional algorithm, according to the source of errors, the measurement errors mainly due to human factors, and another factors are influenced by the noise of instrument and background noise caused by environment changing. These errors can be effectively decreased by the re-training of the artificial neural network, which is better than other algorithm, so the MSC-ANN algorithm proved that it is reliable as a method of modeling.

References

1. Dong, J., Niu, Z., Shen, Y.: Journal of Jiangxi Agricultural University 28(4), 587–592 (2006)
2. Elvidge, C.D.: Visible and near infared reflectance characteristics of dry plantma-terials. Int. J. Remote Sensing 11(11), 1775–1795 (1990)
3. Mooney, H.A., Vitousek, P.M., Mastson, P.A.: Exchange of materials between ter-restrial ecosystems and the atmosphere. Science 238, 926–932 (1987)
4. Xue, L., Luo, W., Cao, W., Tian, Y.: Journal of Sensing 7(1), 73–80 (2003)
5. Tanner, C.B.: Lant Temperature. Agronomy Journal 50, 210–211 (1963)
6. Penuelas, J., Pinol, J., Ogaya, R., et al.: Estimation of Plant Water Concentration by the Reflectance Water index WI (R900/R970). International Journal of Remote Sensing 18, 2869–2872 (1997)
7. Penuelas, J., Inoue, Y.: Reflectance Indices Indicative of Changes in Water and Pigment Contents of Peanut and Wheat Leaves. Photosynthetica 36(3), 355–360 (1999)
8. Tian, Q., Gong, P., Zhao, C.: Chinese Science Bulletin 45(24), 2645–2650 (2000)
9. Wang, J., Zhan, C., Guo, X.: Scientia Agricultura Sinica 34(1), 104–107 (2001)
10. Fan, Z., Simone, G., Bi, Y.: Transaction of the Chinese Society of Agricultural Engineering 21, 215 (2005)
11. Ji, H., Wang, P., Yan, T.: Spectroscopy and Spectral Analysis 3(27), 514–516 (2007)
12. Zhao, X., Liu, S., Wang, P., Wang, J., Tian, Z.: Geography and Geo-Information Science 20(3), 36–39 (2004)

13. Lu, Y., Qu, Y., Song, M.: Spectroscopy and Spectral Analysis 27(5), 877–880 (2007)
14. Pang, T., Yao, J., Du, L.: Spectroscopy and Spectral Analysis 7(27), 1337–1340 (2007)
15. Tang, C., Chen, M., Peng, Y.: Academic Periodical of Farm Products Processing 3(3), 52–54 (2007)
16. Bao, Y., Wu, Y., He, Y.: Joumal of Agricultural Mechanization Research (3), 162 (2004)
17. Qi, X., Zhang, L., Du., X.: Spectroscopy and spectral Analysis 23(5), 870 (2003)
18. Liu, L., Wang, J., Zhang, Y.: Journal of Sensing 11(3), 289–296 (2007)

FT-NIR and Confocal Microscope Raman Spectroscopic Studies of Sesame Oil Adulteration

Jun Luo[1,*], Tao Liu[2], and Yande Liu[2]

[1] School of Mechanical Engineering, Jingdezhen Ceramic Institute,
Jingdezhen Jiangxi 333403 China
[2] School of Mechanical and Electronical Engineering,
East China Jiaotong University, Nanchang 330013, P.R. China
`luolitian@qq.com, jxliuyd@163.com`

Abstract. Adulteration of edible vegetable oils is a very serious problem affecting its commercial value and customers' health. FT-NIR and Confocal Micro-Raman have been explored for discriminating adulteration of sesame oils. The soybean, corn, and peanut oil were mixed into sesame oils in the range of 5~20% (v/v),respectively,and their Confocal Micro-Raman and FT-NIR spectra were collected. Principal component analysis (PCA) and Partial least squares discriminant analysis (PLSDA) were used for the discrimination and classification of adulteration in sesame oil based on spectral data. In addition, the spectra were subjected to the pretreatements (eg, Savitzky-Golay smoothing, Multiplicatnive Scatter Correction, first derivatives and second derivatives) before developing PCA and PLS-DA models. FT-NIR was found to give better efficient in classification of adulteration oils by using PLS-DA with ca.100% classification accuracy, and Raman also gave 100% classification accuracy. The results demonstrated that FT-NIR and Confocal Micro-Raman both can be used to determine the authenticity of edible oils rapidlly.

Keywords: FT-NIR, Confocal Micro-Raman, Sesame Oil, Adulteration, PCA, PLSDA.

1   Introduction

As is well known, authenticity is a very important quality criterion for edible oils, and adulteration is a major problem for the specialist food markets. Many edible oils maybe become polluted through careless handling, or through unintentional or deliberate intervention. Many expensive oils are adulterated with cheaper alternatives to increase profit margins, or expensive ingredients are replaced with a cheaper alternatives. In some cases, this action simply means reducing the quality of the product, but such practices may even have moral and ethical implications. Numerous methods for assessing the authenticity of edible oils have been proposed, such as pyrolysis-mass spectrometry and gas chromatography[1], focused on the existence or

---

*Corresponding author.

absence of certain chemical compounds in the adulterated sample. However, the traditional methods for determination of authenticity based on titration are expensive, complex, laborious, time-consuming, require skilled personnel and sophisticated instrument capability. So, there is a need to develop a simple, cheap, and rapid alternative method to determine adulterants.

NIR spectra assesses chemical structures through the analysis of the molecular bands in the NIR spectrum (eg.C-H)[2], but the peaks in the near-infrared (NIR) region (1,100~2,500nm or 9,091~4,000 cm^{-1}) are broad and weak due to the overtone and combination vibrations originated from the chemical composition of samples. NIR spectroscopy was the first applied to determine the presence of adulterants (corn oil, sunflower oil, and raw olive residue oil) in olive oils[3].Kasemsumran, Sumaporn[4] made a new processing based on partial least squares (PLS) algorithm for the discrimination and determination of adulterants in pure olive oil using near-infrared (NIR) spectroscopy. Balabin,RomanM.[5] tried to evaluate the efficiency of different methods for motor oils classification by using near-infrared (NIR) spectroscopy and SVM and PNN chemometric techniques were found to be the most effective ones for classification of motor oil.

The potential of Raman spectroscopy for the analysis of oils has been recognized for some decades. Raman spectroscopy has already been successfully used to determine important composition or physical structure parameters in a wide range of lipid sample types. Ming-Qiang Zou[6]established a method based on Raman spectroscopy vibration bands to authenticate genuine/fake olive oil and this method can reliably distinguish the genuine olive oils from the olive oils containing 5% (volume percentage) or more of other edible oils, such as soybean oil. Ru-Hui Weng[7] determined the authenticity of Camellia oleifera Abel Oil by Near Infrared Fourier Transform Raman Spectroscopy and a linear relationship with a high correlation coefficient (R^2= 0.9938) was obtained. Rasha M. El-Abassy[8] used a visible Raman spectroscopic method for determining the free fatty acid (FFA) content of extra virgin olive oil with the aid of multivariate analysis.

This work shows the results of using Confocal Micro-Raman and FT-NIR spectroscopy combined with principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) to detect the classification of adulteration in sesame oil and compares the performance of two spectroscopic techniques.

2 Materials and Methods

2.1 Sample Description

The samples used in this investigation were sesame (Shenzhen, Jinlongyu brand), soybean (Shenzhen,jinlongyu brand), corn (Shanxi, Xiwang brand) and peanut oil (Shenzhen,jinlongyu brand) that obtained from a local supermarkets. The concentration of adulteration was in the range of 5~20% by weight respectively, and the same was repeated with each concentration. All weighing was done by use of a analytic balance AUX120 (SHIMADZU, Germany) with the accuracy of 0.0001g. Totally, 66 adulterations of sesame oil mixed with soybean oil(n=22), corn oil(n=22) or peanut oil(n=22) were prepared, and their spectra were collected.

2.2   FT-NIR Analysis

The NIR measurements were performed using transmittance mode in the region of 12,000~4,000 cm^{-1} by means of TENSOR37 FT-NIR spectrometer(Bruker Optics Inc Germany), equipped with the RT-InGaAs detector. All the spectral data were collected with a 4 cm^{-1} spectral resolution and a total of 32 scans were co-added to ensure a sufficient signal-to-noise ratio. The oil samples were scanned in a rectangular quartz cuvettes. The quartz cuvette was cleaned with pure ethyl alcohol after successive measurements and dried to ensure the best possible sample spectra. To avoid any spectral variation due to some factors, air background spectrum was collected before each sample spectrum.

2.3   Raman Analysis

The Confocal Micro-Raman Spectrometer (BRUKER, SENTERRA) combined with a spectrometer module and a microscope module was used in this experiment. The spectrometer module was equipped with a CCD detector (thermo-electrically cooled, 1025×256 pixel) and a semiconductor laser operating at of 785 nm excitation (50mW at sample). The total spectral range (90~3,500 cm^{-1}) was recorded within 15 s of total integration time at 9 cm^{-1} resolution and every measurement consisted of five averaged signal accumulations each with an actual exposure time of 3 s. The oil samples analyzed were poured into a silver spoon and placed on the sample stage (motorized stage 100mm×80mm).

2.4   Discriminant Analysis

The OPUS software (verion. 6.5; Bruker Optik GmbH, Germany) was employed for the spectral data collection. PCA and PLS-DA were performed by using the Unscrambler software (verion. 10.1; CAMO AS, Trondheim, Norway). In order to minimize the effect of light scattering, The spectra were subjected to Savitzky-Golay smoothing (7-point, 2orders), Multiplicative Scatter Correction, first derivatives (7-point Savitsky-Golay filter) and second derivatives (7-point Savitsky-Golay filter) before developing PCA and PLS-DA models for the discrimination. For the PLS-DA analysis, a set of values were given to the spectra of examined adulterant arbitrarily for the dummy variable "species", such as corn=1.0, soybean=2.0, peanut=3.0. All samples with an error around predicted value more than or equal to 0.5 were classified as not being members of the examined adulterant type. On the other hand, those with an error value lower than 0.5 were classified as being members of the adulterant type. The results are presented as percentage, the expression is given as correct discrimination rate (CDR) of models.

$$CDR = \frac{N\text{-}n}{N} \tag{1}$$

where N represents total number of calibration (validation)samples for each model, n is the number of erroneous judgement of samples in every model.

Principal component analysis (PCA) is a nondirected, nonsupervised method of data reduction, which determines the main sources of variation within the spectra,

allowing transformation of several related data points into a single number[9]. Discriminant analysis is a supervised classification technique where the number of groups and the samples that belong to each group are previously defined [10]. Partial least squares discriminant analysis (PLS-DA) differs from PCA in that each source of variation is not ranked in relation to the overall variation, but according to how it accounts for variation between data subsets(groups).

For chemometric analysis, five replications of each sample were used for calibration and validation models. A total of 110 spectra of NIR or Raman were acquired for each adulterant. The same sets of samples were used for enabling comparison of the spectral techniques. For each sample, four of 5 measurements were used for the development of calibration model and rest for validation. Hence a total of 264 individual samples were used for calibration and 66 for validation.

3   Results and Discussion

As we know, every oil differs in composition, length and unsaturated degree of the fatty acids as well as their positions in the chain. Raman spectra represent a combined fingerprint pattern unique to each oil and are used for discriminant analysis. Many analysis were conducted using regions specific to the C=C bond vibration. The double bond considered as an unsaturated bond and the intensities, areas or heights of its peak in this region might indicate the degree of unsaturation in fatty acids, sterols, and vitamins.

3.1   FT-NIR Spectra

Fig. 1 shows some FT-NIR spectra of sesame oil adulterated with corn oil、 soybean oil and peanut oil in the $12,000 \sim 4,500$ cm^{-1} region. The spectra shows three major band groups at around 8,263, 7,186, and 5,788 cm^{-1}. The region between 5,500 and $6,000$cm^{-1} is the first overtone of the C–H stretching from–CH_2, –CH_3 and –CH=CH– functional groups. The region between $7,000 \sim 7,400$ cm^{-1} contains bands assigned to the combination modes of the CH stretching and CH deformation vibration modes. The region between 8,000-8,500 cm^{-1} consists of bands due to the second overtone of CH stretching modes of the CH_3 and CH_2 groups stretching modes. It is noted in Fig. 1 that those mean NIR spectra are very similar to each other, and it is not easy to classify the spectra into each adulterant type. Therefore, the PCA and PLS-DA method are applied in the study to identify the unknown adulterations in sesame oil samples.

An overall PCA was carried out using all samples of each oil. Fig 2 shows the classification results of the PCA on data sets containing three kinds of the oil spectra in score plots. The PCA model was achieved using 1PCs and explaining 97% of variance in the spectral data. The plots shows that the peanut oil is clearly separated from the others. But corn oil has so many portion in common with soybean oil, which gives no boundaries. As with all non-supervised methods for classification, the plot built by PCA reveals and explains the patterns of the samples studied, but the assignation of new samples to a class may be a difficult task, as the method does not calculate a rule to generate boundaries or regions for the groups of samples obtained. Hence, supervised method (PLS-DA) was introduced in order to improve the results.
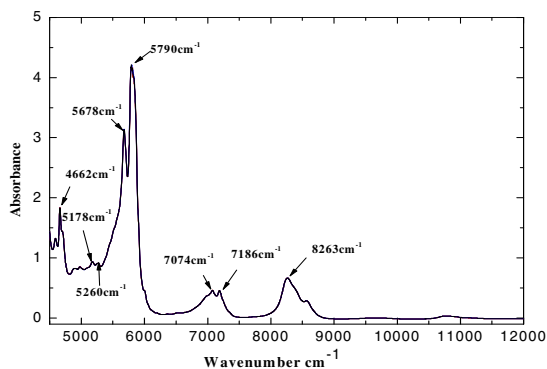
Fig. 1. FT-NIR spectra in the 12,000~4,500cm^{-1} region of sesame oil with adulteration
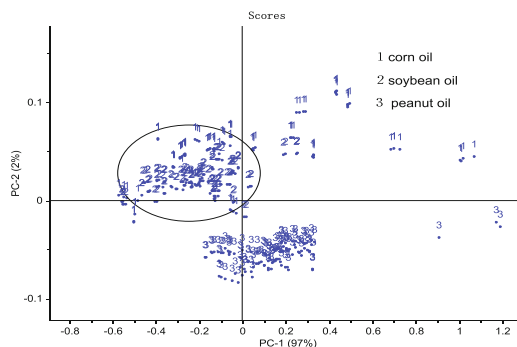
Fig. 2. PCA plots for classification of oils using FT-NIR spectra

The statistical results of PLS-DA based on FT-NIR spectra are summarized in Table 1. The discrimination models are developed by using two selected spectral ranges and five spectra types: original and four pretreatment spectra. The rate of correct classification obtained from discriminant analysis of the spectra between 6,410.4 and 8,917.4 cm^{-1} is found to be better than the data from the spectral region between 4,000 and 12,000 cm^{-1}. In the region 6,410.4~8,917.4 cm^{-1}, the models using any pretreatment yield 100% correct classification of the prediction set (%CP), especially the second derivative spectra with only the lowest number of PLS factor 2. As is well-known, a small number of PLS factor required for developing a model that provides acceptable results indicates the stability and accuracy of such model[11].However, when the wide spectral range(4,000~12,000 cm^{-1}) is used, the model developed by using the second derivative spectra shows the worst classification performance. This is because the wide spectral range includes a region that contains more noise than relevant information to the model development for the concentration of an analyte. Therefore, the optimal models could be selected from the models giving the best predictive accuracy with the smallest number of PLS factor

Table 1. Discrimination analysis of FT-NIR spectra of the adulterant types in sesame oil

	pretreatment	Factors	% CDRC[a]	% CDRV[b]
6,410.4- 8,917.4cm^{-1}	None	6	100	100
	smoothing	6	100	100
	MSC	4	100	100
	1st	3	100	100
	2nd	2	100	100
4,000- 12,000cm^{-1}	None	10	99.6	90.9
	smoothing	10	99.6	89.4
	MSC	10	99.6	95.5
	1st	4	86.4	63.6
	2nd	5	41.7	45.5

CDRC[a]=correct discrimination rate of calibration, CDRV[b]= correct discrimination rate of validation.

3.2   Confocal Micro-Raman Spectra

The Confocal Micro-Raman spectra of pure sesame oil are shown in Fig 3. The Fig 3 shows that many major bands of the oleic acid were detected in the Raman spectra of the oil samples, and are attributed to the main components in the oil, which are fatty acids. It should be noted that the region between 1,600 and 1,700 cm^{-1} contains the stretching vibration of the C=C bond which is very distinctive around 1,655 cm^{-1} in Raman spectra. It demonstrates that spectra can display non-polar groups of samples which can not be observed in the FT-NIR spectra.
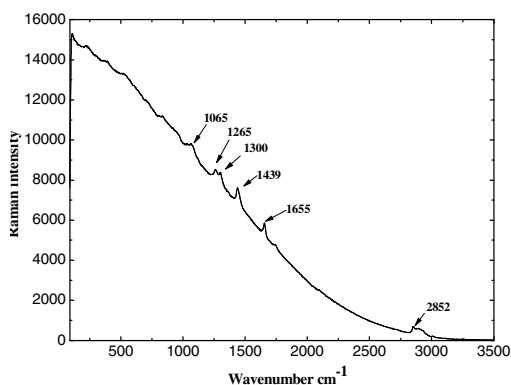
Fig. 3. Raman spectra of pure sesame oil in the 90~3,500cm^{-1} region

According to Fig 4, we can see that the discriminant models built with Raman spectra by PCA are the same weak to decide for samples as the models built with FT-NIR spectra. Although corn oil can be separated from others successfully，giving a value of good prediction 100%, Soybean and peanut oil are both distributed in the same circle. Therefore, we need better methods to make models perfect.
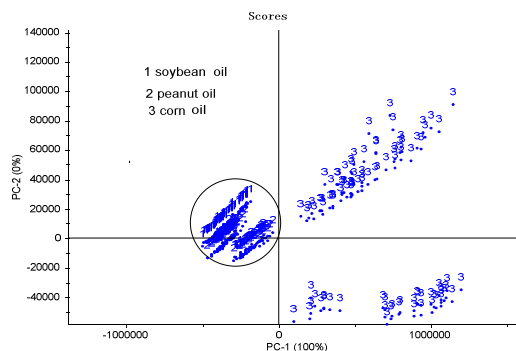
Fig. 4. PCA plots for classification of oils using Raman

The spectra regions (90~1,795, 2,818.5~3,159 and 90~3,500cm^{-1}) were used for discriminant analysis. The results obtained from PLS-DA for the adulteration discrimination in sesame oil are summarized in Table 2. The rate for correct classification with PLS-DA methods was found to be more than 95.5% for both the calibrationand validation samples due to the degree of unsaturation. It is noted in Table 4 that the excellent prediction results are obtained mostly from the models built with the regions of 90~1,795 and 2,818.5~3,159cm^{-1} at the same time. It also can be seen that when the spectral data are pretreated by MSC, the model gives the highest rate(100%) for correct classification. However, the results from Confocal Micro-Raman measurements were not as good as these from FT-NIR. Vibrations from polar groups, such as C=O and O–H, are very weak in Raman spectra, which result in the missing of some important vibrations and these vibrations maybe play very important roles in discriminant analysis.

Table 2. Discrimination analysis of Raman spectra of the adulterant types in sesame oil

	pretreatment	Factors	%CDRC[a]	%CDRV[b]
90~1,795 2,818.5~3,159cm^{-1}	None	10	100	96.9
	smoothing	9	99.2	97.0
	MSC	10	100	100
	1st	5	96.7	96.7
	2nd	7	100	97.0
90~3,500cm^{-1}	None	11	100	98.5
	smoothing	10	99.2	98.5
	MSC	9	99.6	100
	1st	6	96.2	95.5
	2nd	5	99.2	95.5

CDRC[a]=correct discrimination rate of calibration, CDRV[b]=correct discrimination rate of validation.

4   Conclusions

This study compared Confocal Micro-Raman and FT-NIR spectroscopy for detecting the classification of adulteration in sesame oil. The results revealed that either of the spectral methods is easier to perform and the FT-NIR method was found to be more efficient. However, Raman spectroscopy techniques could provide exquisite structural insights into functional groups of oils for discriminant analysis. In addition, it is clear from the results above section that the PLS-DA is a very powerful method to identify adulterant types in the adulteration of sesame oil samples.

References

1. Yang, Z., et al.: Oil fingerprinting analysis using commercial solid phase extraction (SPE) cartridge and gas chromatography-mass spectrometry (GC-MS). Anal. Methods 3, 628–635 (2011)
2. Cozzolino, D., Fassio, A., et al.: Verification of silage type using near-infrared spectroscopy combined with multivariate analysis. Journal of Agricultural and Food Chemistry 56, 9–83 (2008)
3. Wesley, I.J., Barnes, R.J., et al.: Measurement of Adulteration of Olive Oils by Near-Infrared Spectroscopy. J. Am. Oil Chem. Soc. 72, 289–292 (1995)
4. Kasemsumran, S., et al.: Partial Least Squares Processing of Near-InfraredSpectra for Discrimination and Quantification of Adulterated Olive Oils. Spectroscopy Letters 38(6), 839–851 (2005)
5. Balabin, R.M., Safieva, R.Z., et al.: Near-infrared (NIR) spectroscopy for motor oil classification: From discriminant analysis to support vector machines. Microchemical Journal 98(1), 121–128 (2011)
6. Zou, M.-Q., Zhang, X.-F., et al.: Rapid Authentication of Olive Oil Adulteration by Raman Spectrometry. J. Agric. Food Chem. 57(14), 6001–6006 (2009)
7. Weng, R.-H., Weng, Y.-M., Chen, W.-L.: Authentication of Camellia oleifera Abel Oil by Near Infrared Fourier Transform Raman Spectroscopy. Journal of the Chinese Chemical Society 53, 597–603 (2006)
8. El-Abassy, R.M., Donfack, P., Materny, A.: Rapid Determination of Free Fatty Acid in Extra Virgin Olive Oil by Raman Spectroscopy and Multivariate Analysis. J. Am. Oil Chem. Soc. 86, 507–511 (2009)
9. Renwick Beattie, J., Steven, E.J., et al.: Classification of Adipose Tissue Species using Raman Spectroscopy. Lipids 42, 679–685 (2007)
10. Cozzolino, D., Fassio, A., et al.: Verification of silage type using near-infrared spectroscopy combined with multivariate analysis. Journal of Agricultural and Food Chemistry 56, 79–83 (2008)
11. Martens, H., Næs, T.: Multivariate Calibration. John Wiley, Chichester (1989)

Non-Point Source Pollution Characteristics of Agriculture-Derived Nitrogen in Groundwater in Suburban Area of Shanghai Based on Models

Guangrong Shen[1,2], Xiumei Huang[1,2], Pei Zhou[1], Lumei Wang[1], and Yuee Zhi[1]

[1] Key Laboratory of Urban Agriculture (South) of Ministry of Agriculture,
Shanghai Jiao Tong University, Shanghai, 200240, China
[2] Research Center for Low Carbon Agriculture,
Shanghai Jiao Tong University, Shanghai 200240, China

Abstract. This paper analyzes the characteristics of the NPS nitrogen pollution of the groundwater in Guoyuan Village, Pudong District, Shanghai. And the associated effects on the surface and groundwater around the study area are discussed in detail based on the successive observed data, DNDC and L-THIA model. The results show that both of the surface and groundwater are polluted so seriously that they are not suitable to drink. The average content of total nitrogen in surface water is 6.3 mg/L and 16.85 mg/L in the groundwater, and both of them attribute to Grade Ⅴ surface water standard (≤2.0 mg/L) according to the national standard(GB 3838-2002). It is concluded that the nitrogen pollution comes mainly from the fertilizer of the peach orchard based on the further modeling analysis. Therefore, reasonable adjustment of fertilization measures and project may be an effective and practical approach to control the nitrogen pollution around the peach orchard.

Keywords: DNDC model, L-THIA model, nitrogen content, NPS pollution, peach orchard.

1   Introduction

Non-point source (NPS) pollution is the main factor of surface and groundwater pollution as point source pollution is becoming under control. Among all the non-point pollution, the agriculture-derived nitrate NPS pollution of groundwater has become an environmental issue which may cause algal bloom and eutrophication in aquifers, and even produce potential hazards to human health though the link is still disputable [1]. The heavy use of nitrogen (N) fertilization for intensive farming and cropping system with low N use efficiency is often responsible for nitrate overloading into groundwater. The role of agricultural NPS in water quality degradation is the   most prominent [2-3], especially in the suburbs of Shanghai where ecological agriculture is predominant. In 2001, the application of N in fields was up to 612.75 kg/hm^2, which is 59.6% higher than the average level in China, and with 1.3×10^6 kg N•a^{-1} flowing to the aquifers [4]. Characteristics of agriculture-derived NPS nitrogen pollution and the load estimation are the important part for regional water quality control. Concerning studies on the

pollution load include many aspects, such as estimation of the relationship between total precipitation and pollution load, estimating NPS pollution load of the long-term average output by using the model (L-THIA) [5], farmland runoff pollution load model [6], the use of NPS Watershed SWAT model system for different scenarios of watershed management [7] and agricultural NPS pollution load. There are some studies on the characteristics of agricultural NPS pollution focusing on analyzing the factors such as the land use types, topography and its associated impacts on the NPS pollution. The spatial and temporal variability of agriculture-derived NPS pollution is the subject of intensive research recently.

In this study, the characteristics of the agriculture-derived NPS nitrogen pollution of groundwater is analyzed based on the biogeochemical process model DNDC, the L-THIA model and the simultaneous observation data in soil, surface and groundwater. Guoyuan Village, Xinchang Town of Pudong District, located in the southern Shanghai, China, traditionally a high-yield area for growing peaches, is selected as the study area. We aim at exploring the optimum efficiencies of water and fertilizer usage in peach orchard by analyzing the relationship between agricultural management measures and agriculture-derived NPS nitrogen pollution. Results are discussed in terms of the multi-objective management for peach orchard in order to reduce or avoid the nitrogen pollution in groundwater.

2   Study Area

The research is conducted in the peach orchard of Guoyuan Village (31°03'N, 121°39'E), located in Xinchang Town of Pudong District on the east of Huangpu River in Shanghai, China (Fig.1(a)). The site situates in a subtropical monsoon zone with 4 m altitude. The average annual temperature is 15.7 ℃ and 1100 mm of the average annual rainfall, ranging from 300 to 1300 mm with most falling during the period from July to September[8]. There has been a long history of peach planting in this area since 1621. The planting area reached 5336 hm^2 in 2005 and peach became the symbol of this district [9]. The study area, which is in roughly rectangular shape with the area of about 33.35 hm^2, is adjacent to living area of Guoyuan Village (Fig.1(b)). Peach trees are planted at a density of an amount of 900 per hectare. The peach orchard is surrounded by river and some ditches distribute regularly inside. A main ditch from east to west divides the peach orchard into northern and southern parts. The north of the main ditch is segmented vertically by 11 ditches with equal space while several ditches distribute irregularly on the other side (Fig.1(b)). The sources of the water pollution are mainly from the peach orchard and residential area because there are no other land use types in the district. The compound planting mode that green vegetables are planted under the peach trees when the peach is ripe (from June to August) is employed in the peach orchard. 70% of the peach orchard is interplant one-season vegetables, others of two-season vegetables [10]. This compound planting and management mode in study area increases soil nitrogen content and associates with the surface and groundwater pollution.

(a)   (b)
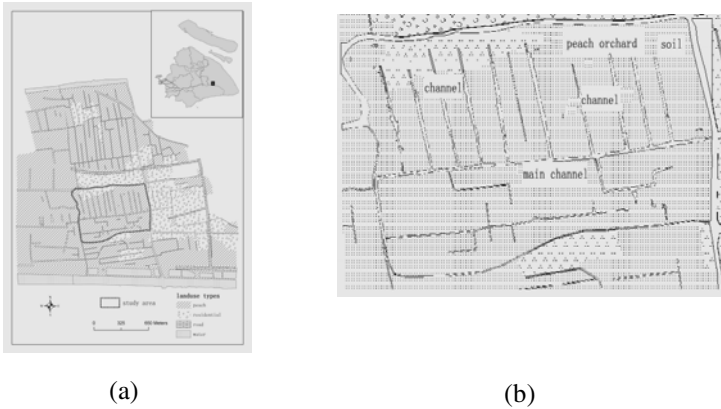
Fig. 1. The location of Guoyuan Village & ditches in peach orchard

3   Material and Method

3.1   Data Collection

Measurements of N in surface, groundwater and soil have been carried out since June, 2009. The monitoring locations in soil, surface and groundwater were chosen respectively according to their different features. GPS was used to fix the monitoring location. It is worthy of mentioning that the monitoring locations of surface and groundwater were collected precisely according to the topography and the flowing direction of the ditches to make sure that monitoring locations can present the spread of nitrogen in the soil runoff as much as possible (Fig.2).
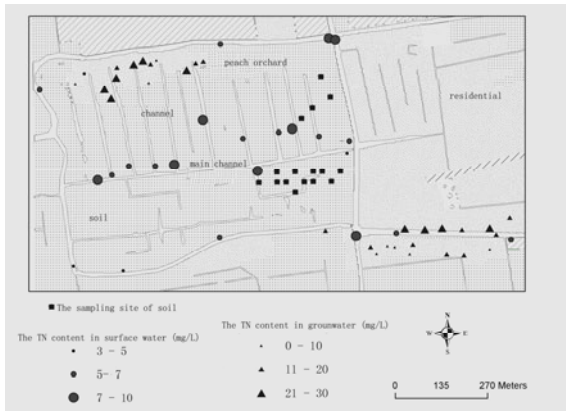
Fig. 2. The distribution map of monitoring locations

- Sixteen soil samples were collected from different depth (0 ~ 15 cm, 15 ~ 30 cm, 30 ~ 45 cm) in June, 2009. Total nitrogen in soil was analyzed by conventional method and nitrate (in terms of nitrogen) by Bremner method.

- Fifty-three water samples were collected from Oct.2009 to Dec.2009, Jan.2010 and from Mar.2010 to Aug.2010. Twenty-five samples were from surface water, and twenty-eight samples from groundwater. Total nitrogen in the water was analyzed by alkaline potassium persulfate ultraviolet spectro-photometric method and nitrate (in terms of nitrogen) by Zinc cadmium reduction method.

3.2   Model Parameters

Field management including fertilizer application and tillage system can influence soil fertility directly through fertilizer inputs and nitrogen circulation in agro-ecological process. Models can simulate the processes responsible for production, consumption and transport of N_2O in both long and short terms, and also allow spatial simulation [11]. Simulation mechanistic models consider all factors affecting N_2O circulation in the agriculture ecosystem. In this study, these factors include the planting mode, fertilizer application, irrigation strategy and the tillage measures of peach orchard in recent ten years (Tab.1), which were acquired through field survey, visiting local farmers and references [8].

Table 1. The management measures of the peach orchard

Month	Management
1	Fertilizer (compound fertilizer 750 kg/hm^2, organic fertilizer 1.5×10^4 kg/hm^2), deep tillage
3	Fertilizer (compound fertilizer 600 kg/hm^2)
5	Fertilizer (compound fertilizer 600 kg/hm^2), shallow tillage
9~11	Interplanting green vegetable
10	Irrigation, deep tillage
11	Fertilizer (compound fertilizer 545 kg/hm^2)

(The N concentration of the compound fertilizer is 15%, the organic material content in the organic fertilizer is 25%, and the N concentration is 0.45%.)

In addition, the spatial data such as the land use maps of different years, soil type maps of Guoyuan village are necessary, which are used to calculate the pollution load in agricultural runoff based on hydrologic impact assessment model. These data were provided by local village committee or obtained through remote sensing image interpretation. Meteorological data such as the temperature, rainfall data of recent thirty years were provided by the weather station of Pudong district.

3.3   Study Approach

The study area is a relatively independent agro-ecosystem. The compound management mode and agricultural management practices directly impact on the soil-water NPS nitrogen pollution in the area around peach orchard. Modeling allows the complex links

among soil physical, chemical and microbial processes that underpin nitrification, de-nitrification and decomposition to be examined. Therefore, this paper quantitatively calculates and analyzes the mechanism of nitrogen circulation in soil based on DNDC model, and estimates the nitrogen content in soil runoff by using NPS pollution load model L-THIA. The measure values of monitoring locations are compared with the simulating values in order to analyze in detail the mechanism of nitrogen dynamic change in agro-ecosystem and the relationship between the compound management model and the pollution characteristics of the surface and groundwater.

3.3.1   Biogeochemical Model

The DNDC model is a mechanism model, which is used to simulate and assess the emission of N_2O, NO, N_2 and CO_2 from agro-ecosystem [10-13]. Although this model was initially created according to the growing conditions of America, it has been verified in many agro-ecosystems such as grassland, farmland and forest ecosystem in many countries such as America, Canada, India, European, China and New Zealand and so on. DNDC model, which is considered as one of the best biogeochemical models [14], is able to simulate the emission of a fixed point and also can assess the emission of a region. The model has reasonable data requirement and is suitable for simulation at appropriate temporal and spatial scales. Tab.2 shows the parameters and associated values required by DNDC model in this study.

Table 2. The parameters and partial values in DNDC model

	Parameter	**Value**
Climate information	Temperature & precipitation	Daily data (omission)
	N content in rainfall (mg N/L)	2.53
	NH3 content in the air (ug N/m^3)	0.06
	CO_2 content (ppm)	350
Soil background information	Bulk density (g/cm^3)	1.33
	Texture	clay
	SOC (kg C/kg)	0.022
	PH	6.8
Crop information	Crop	Peach, green vegetable
	Rotate mechanism	Interplanting
	Growth parameter	Each organ (omission)
Management information	Times of tillage	3 times a year
	Fertilizer amount ($kg N/(hm^2 \cdot a)$)	Compound fertilizer:374 Organic fertilizer:150.62
	Times of irrigation	Once a year
	Portion of straw returning	Peach 100% Green vegetable 0%

DNDC model can simulate the biogeochemical process under a certain management of production and learn about soil nitrogen pollution and the nitrogen flux during the process. It also can analyze the effects of different management measures such as fertilization and tillage on soil organic carbon content [15]. This provides the basis for further quantitative analysis of nitrogen load in agricultural runoff.

3.3.2 NPS Pollution Load Model

Agricultural runoff is the key factor to quantify and simulate agricultural NPS pollution. The concentration of output pollutants in runoff and the pollution load per unit area are related with the regional climatic conditions [16]. Long-term hydrologic impact assessment model (L-THIA), which is integrated with GIS, can estimate the average annual runoff and NPS pollution load based on land use type, soil type and daily rainfall for at least thirty consecutive years [17].

The adjacent western and southern areas of the study area are also peach orchards, and the northern and the eastern areas are residential areas. Survey and experimental data show that the soil nitrogen is carried over by rainfall and irrigation water to enter the river via surface runoff, resulting in different levels of water eutrophication. The nitrogen content in the runoff is analyzed via simulating by L-THIA model. The migration of nitrogen is estimated comprehensively and quantitatively by using DNDC model. The results are tested and compared with the measured values to explore the mechanism of the nitrogen circulation in agro-ecosystem, and learn about the nitrogen pollution load in surface and groundwater under current field management practices and local climate conditions.

4   Results and Discussion

4.1   The Analysis of Water Samples

The total nitrogen and nitrate nitrogen contents measured in water of the peach orchard are showed in Fig.3. Obviously, the nitrogen content is widely different between surface and groundwater. The nitrogen content in groundwater is about 1 to 2 times higher than that in surface water. The standard error of nitrogen content in surface water is smaller than that in groundwater, which represents that the nitrogen content in surface water changed little among all the samples. In particular, as shown in Fig.3, total nitrogen in surface water measured in May 2010 (14.57 mg/L) is significantly higher than that in other times, which is probably due to large amount of fertilizer application in May. Besides, the rainfall of the day when we sampled reached 45.5 mm (22nd May). However, the variance of nitrogen content in groundwater is very distinct: the total nitrogen content ranges from 0.24 to 49.23 mg/L; the nitrate content from 0.18 to 45.41 mg/L.

Furthermore, we learn about the spatial variation features of nitrogen content in water body reference to Fig.2 and Fig.3. The nitrogen content in surface water samples around the orchard is relatively low, while those near the main ditch are higher. This distribution pattern of nitrogen content indicates that fertilization probably is the main
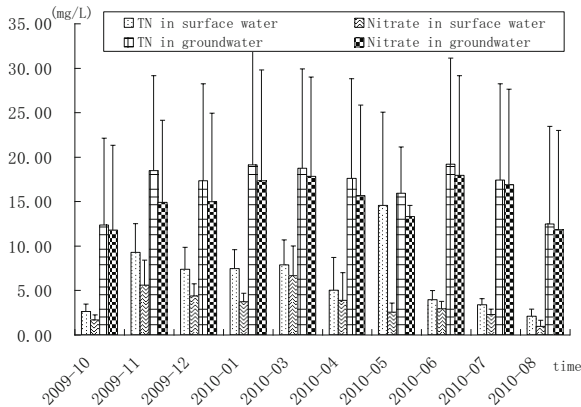
Fig. 3. Nitrogen content in water of the peach orchard

cause resulting in superfluous nitrogen in water. As for groundwater, the nitrogen content is lower in residential area and higher in the orchard. These characteristics are in accordance with that agricultural production activities contribute to NPS nitrogen pollution in water body.

The measured results state that monthly average total nitrogen content of surface water reaches 6.3 mg/L, higher than the value of surface water Grade Ⅴ (≤ 2.0 mg/L) according to Chinese surface water quality standard (GB 3838-2002) (in terms of total nitrogen), which means the surface water in study area is polluted seriously and belongs to surface water Grade Ⅴ based on the current water quality assessment method in China [18]. For groundwater, 16.85 mg/L of monthly average total nitrogen is also much higher than that of the surface water Grade Ⅴ (≤ 2.0 mg/L). The monthly average nitrate nitrogen in groundwater is 15.24 mg/L which falls in to groundwater Grade Ⅲ (≤ 20 mg/L) according to Chinese groundwater quality standard (GB/T14848-93) (in terms of nitrate). However, 20% of groundwater samples belong to water Grade Ⅴ (> 30 mg/L) in terms of nitrate nitrogen content. It is summarized that groundwater in the district belongs to groundwater Grade Ⅴ, far lower than surface water Grade Ⅴ. The water pollution in this area is serious, which is consistent with the conclusion by Huang Hongyan [8].

4.2   Characteristics of NPS Soil Nitrogen Pollution

To explore the characteristics of NPS nitrogen pollution in the peach orchard, DNDC and L-THIA model are employed to simulate C, N circulation in this agro-ecosystem. Climate, management measures such as fertilization application and soil input variables for DNDC are listed in Tab.2. Field data used for all of the variables are listed except for clay fraction and depth of the soil water retention layer. Here default values are used. In this study, the nitrogen input that come mainly from the organic fertilizer, fertilizer, nitrogen sedimentation in rainfall, totalizes to 548.53 kg N/(hm^2•a) according to DNDC. The output nitrogen is 339.48 kg N/(hm^2•a) in this agro-ecosystem including crop absorption, the volatilization of NH_3, N_2O, NO, N_2. So, it is calculated that the

yearly nitrogen flux change or the net increased nitrogen in soil is about 209.05 kg N/ $(hm^2 \cdot a)$ based on simulation.

In addition, the simulated results of DNDC model show that average total nitrogen content from different soil depths is 1887.2 kg/hm^2, while the field measured value in June 2010 is 1728.7 kg/hm^2. The fitness of the two parts is up to 91.6%, which suggests that the simulated results of the C, N circulation in peach orchard by DNDC is credible and valuable for quantificational C, N monitoring.

L-THIA model is used to simulate the runoff depth, total nitrogen & nitrate nitrogen load based mainly on the local climate and land use type data. Fig.4 displays the runoff depth map, total nitrogen & nitrate nitrogen load maps of the peach orchard derived from L-THIA.
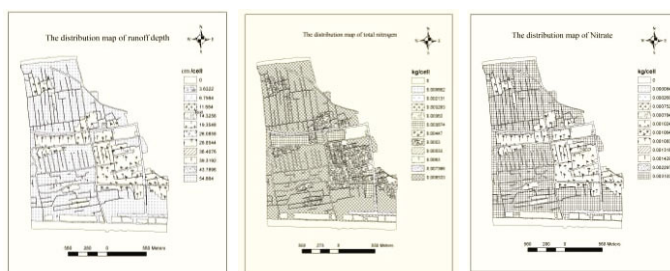
Fig. 4. The distribution maps of runoff depth, TN & NO$_3^-$N

Simulations show that the average annual runoff depth of the peach orchard is 190 mm. As it shows in Fig.4, the maximum nitrogen pollution load is in orchard when compared with other land use types, where the output of total nitrogen is 0.0085 kg/cell (each cell is 3 m * 3 m), or 9.44 kg/hm^2, and the output of nitrate nitrogen is 0.0031 kg/cell, or 3.44 kg/hm^2.

The estimated values from L-THIA model represent the average nitrogen output in a region for a long time, reflecting the average nitrogen pollution level. While the measured values of nitrogen in surface water depend on the crop growing stages, farm management practices and climatic conditions. Comparing the measured values with the simulation results from L-THIA model, the results indicate that simulated total nitrogen content in runoff accounts for 78.08% of measured values in surface water, and nitrate nitrogen covers 52.28% of measured values. Hence, it may be reasonable that there are other nitrogen sources affecting the surface water pollution besides the soil we have known. In addition, the compound management mode of interplanting vegetables in peach orchard promotes the soil nitrogen circulation and increases nitrogen content to some extent. However, the mode is not considered in L-THIA model, which may result in the simulated values littler than it is.

Shanghai, situated in the Yangtze River Delta region, owns a plentiful water source and its soil is with higher nitrogen content, all of which cause serious NPS nitrogen pollution [26]. The average annual rainfall is 1100 mm with most falling from July to September in Shanghai and nitrate could descend 1 cm in soil with about per 2 ~ 3 mm

precipitation [19]. According to this rule, 209.05 kg N/(hm^2•a) of the net increased nitrogen content in the peach orchard soil would go down 330 ~ 550 cm at differently speed. In this study, the samples in groundwater are from the wells whose depth is 4 m. Therefore, the descending nitrogen in soil from agricultural activities might be one of the main reasons resulting in the higher nitrogen concentration in groundwater. Furthermore, taking 11% as the soil seepage and runoff nitrogen loss occupying total nitrogen input [10], the detailed analysis based on L-THIA model and DNDC simulation shows that 1.7% of annual net increased nitrogen goes to the surface water via the soil runoff, 5.8% to the groundwater via soil leaching loss. Thus, 3.5% is cumulative in the soil leachate. In this case, here about 19.2 kg N/(hm^2•a) would enrich nitrogen in groundwater so much as resulting in pollution.

5   Conclusion and Discussion

The NPS nitrogen pollution in surface and groundwater is quantitatively analyzed in detail in this paper based on the monitoring location measured data, DNDC and L-THIA.The surface water falls into Grade Ⅴ in terms of nitrogen content and the groundwater quality belongs to Grade Ⅴ surface water standard, while the nitrate content in groundwater belongs to Grade Ⅴ groundwater standard according to national associated standard.

The fertilizer application in agricultural production is the main pollution source of surface and groundwater in the study area. The mechanism of migration and output of nitrogen in the process is directly related to NPS pollution of the groundwater where nitrogen content variation is wide. It cannot but mention that, there is the same agro-ecological environment around the study area with the same peach orchards and residential areas. The study area is a relatively independent region for the river separates it from the surrounding environment. This study analyzes the nitrogen characteristics of NPS pollution by taking the agro-ecosystem of peach orchard as the case. It is difficult to absolutely divide the boundary between the study area and the conjoint area because of the complexity of the environment of agricultural soil and water, which may affect the accuracy of results. Meanwhile, the same agro-ecosystems surrounding the peach orchard suggests that the results of this study can reflect the status of NPS nitrogen pollution for such agro-ecosystem. In particular, the simulated values, the measured values and the speculated or predicted values of soil nitrogen leakage should include or cover the NPS pollution from some possible fuzzy boundaries. In addition, soil drainage changing with rainfall intensity and duration is one of the main reasons to cause groundwater contamination although the critical cause of nitrogen pollution is heavy fertilizer application. Therefore, to improve fertilization strategy is an effective means to reduce nitrogen pollution. Some measures such as adjusting the fertilization structure, the ratio between compound fertilizer and organic fertilizer, the proportion of nitrogen, phosphorus and potassium in fertilizer and so on should be done to control and stable the amount of nitrogen fertilizer.

References

1. Liu, G.D., Wu, W.L., Zhang, J.: Regional differentiation of non-point source pollution of agriculture-derived nitrate nitrogen in groundwater in northern China. Agriculture, Ecosystems & Environment 107, 211–220 (2005)
2. Boers, P.C.M.: Nutrient emissions from a culture in Netherlands: causes and remedies. Water Sci. Technol. 33, 183–190 (1996)
3. Tim, U.S., Jolly, R.: Evaluating agriculture NPS pollution using integrated geographic information systems and hydrologic/water quality mode1. Environ. Qua1. 23, 25–35 (1994)
4. Gong, Z.-H.: Soil pollution countermeasure in Nanhui district. Shanghai Agricultural Science and Technology 2, 12 (2003)
5. Li, D.-Q., Wang, J.-Z., Wan, H.-F., et al.: Law of NPS pollutants losses in a typical small watershed of Dongjiang Drainage Basin of Guangdong Province. Journal of Soil Erosion and Soil and Water Conservation 4, 12–18 (1998)
6. Zheng, J.-Y., Zhou, N.-X.: Research on NPS nitrogen pollution model in farmland of Shanghai suburbs. Journal of East China Normal University (Natural Science) 6, 12–19 (2007)
7. Su, B.-L., Wang, J.-P., Jia, H.-F., Cheng, S.-T., Yang, Z.-S., Wu, D.-W., et al.: Identification of NPS pollution in the Miyun Reservoir watershed. Journal of Tsinghua University (Science and Technology) 46, 360–365 (2006)
8. Huang, H.-Y., Gao, Y., Cao, J.-J., Mao, L., Zhang, X., Huang, H.-B., et al.: Characteristics of phosphorus NPS pollution on groundwater in urban agricultural region of Shanghai. Journal of Soil and Waer Conservation 2, 101–105 (2010)
9. http://show.eastday.com/market/NANHUIGAME/
10. Huo, J.-F., Wang, J., Yu, M.-K.: Benefit analysis of fruit——crops compound management model——take peach-brassica chinensis compound model in Shanghai city Nanhui area as example. Agricultural Science and Technology Information 6, 12–13 (2009)
11. Willams, E.J., Hutchinson, G.L., Feshsenfeld, F.C.: NO_x and N_2O emissions from soil. Global Biogeochemical Cycles, 351–388 (1992)
12. Krobel, R., Sun, Q., Ingwersen, J., Chen, X., Zhang, F., Muller, T., et al.: Modeling water dynamics with DNDC and DAISY in a soil of the North China Plain: A comparative study. Environmental Modeling & Software 25, 583–601 (2010)
13. Smith, P., Smith, J.U., Powlson, D.S.: A comparison of performance of nine soil organic matter models using datasets from seven long term experiments. Geoderma 81, 153–225 (1997)
14. Wang, L.-G., Qiu, J.-J., Ma, Y.-L., Wang, Y.-C.: Apply DNDC model to analysis long-term effect of soil organic carbon content under different fertilization and plough mode. Journal of China Agricultural University 9, 15–19 (2004)
15. Bei, R.-T., Zhou, Y., He, M.: Potential relationship between soil nitrogen & phosphorus content and water pollution in Dianchi Lake. Journal of Northwest Forestry University 25, 30–34 (2010)

16. Tao, S., Liu, L.-Y., Ma, J.-F.: Spatial distribution characteristics of non-point pollution of Miyun reservoir areas based on L-TH IA model. Transactions of the CSAE 5, 62–68 (2007)
17. Lim, K.J., Engel, B.A., Kim, Y., Bhaduri, B., Harbor, J.: Development of the Long Term Hydrologic Impact Assessment (LTHIA) WWW Systems. In: The 10th International Soil Conservation Organization Meeting, pp. 1018–1023 (1999)
18. Tang, K.-W., Zhu, D.-S., Tang, Y., Wang, Y.: Groundwater quality assessment of urban drinking water sources in China. Water Resources Protection 25(1), 1–4 (2009)
19. Hou, Y.-L., Li, H.-Y., Zhao, H.-M.: China N NPS pollution of farmland and empirical estimation: various types of pollution levels and trends. Journal of Agro-Environment Science 28, 1341–1345 (2009)

Nondestructive Estimation of Total Free Amino Acid in Green Tea by Near Infrared Spectroscopy and Artificial Neural Networks[*]

Zhiming Guo[1], Liping Chen[1,**], Chunjiang Zhao[1],
Wenqian Huang[1], and Quansheng Chen[2]

[1] National Engineering Research Center of Intelligent Equipment for Agriculture,
Beijing 100097, China
[2] School of Food & Biological Engineering,
Jiangsu University, Zhenjiang 212013, China
chenlp@nercita.org.cn

Abstract. The total free amino acid of green tea was nondestructive estimated by near-infrared (NIR) spectroscopy combined with multivariate calibration, compared the performance of back propagation neural networks (BP-NN) and partial least squares (PLS) regression analysis. The original spectra of tea samples in wavelength range of 10000-4000cm^{-1} were acquired. Spectral pretreatment methods were applied to reduce the systematic noise, and enhance the contribution of the chemical composition. The model was optimized by cross validation, and its performance was evaluated according to root mean square error of prediction (RMSEP) and correlation coefficient (R) in prediction set. Experimental results showed that the performance of BP-NN model was superior to the performances of PLS model, from the point of view of the predictive ability. The optimal results of the BP-NN model with multiplicative scatter correction spectral pretreatment were achieved as follow: RMSEP=0.246 and R$_p$=0.958 in the prediction set, respectively. It can be concluded that NIR spectroscopy combined with BP-NN has significant potential in quantitative analysis and monitor of free amino acid content in green tea.

Keywords: Near infrared spectroscopy, back propagation neural network, nondestructive estimation, free amino acid, green tea.

1 Introduction

Tea is considered one of the most popular beverages consumed worldwide. It is made from the leaves of the plant *Camellia sinensis* (L.). There are different tea varieties with different external qualities and inner qualities due to morphologic and chemical

[**] Corresponding author.

diversities [1, 2]. Tea has important physiological properties and potential health benefits due to the presence of compounds such as polyphenols, amino acids, vitamins, carbohydrates, caffeine, and purine alkaloids, and those chemical constituents influence the taste and flavor of tea infusions [3, 4]. Tea flavor is enhanced by the umami taste of amino acids. The predominant amino acid in tea infusion is a unique amino acid, theanine, which exists only in the free form and accounts for about 50% of the total free amino acids (TFAA). It has been reported that theanine can decrease the level of norepinephrine and serotonin in brain [5], and intake of theanine by hypertensive rats results in decreased blood pressure [6]. The total content of amino acids in green tea is the greatest in all kinds of tea. It has been demonstrated that there is a relationship between the quality of green tea and the amino acid contents [7-9]. However, amino acid levels in tea vary during each manufacturing stage. Young two-leaf flushes are typically chosen for manufacturing the finest tea because theanine accumulates in growing shoots [10]. Therefore, a rapid assay for TFAA in tea is essential for quality control, evaluation and classification in tea processing.

There many descriptions of procedures for analysis of amino acids in green tea based on high-performance liquid chromatography (HPLC) [11,12], capillary electrophoresis (CE) [13], fluorometric flow-injection analytical [14] and ninhydrin colorimetric assay. Although these assays mentioned above are accuarate to quantify individual compounds, they are labor intensive, time-consuming and requires large quantities of regents for quality control purposes.

Near infrared reflectance (NIR) spectroscopy has been proved to be a powerful analytical tool that can be acted as a replacement of time-consuming chemical assay and there is no need for sample preparation or trained staff to perform the analysis routinely. NIR spectroscopy can be used online as well as in the field using the appropriate instrument. It has been introduced as a rapid and non-destructive method for evaluating tea quality parameters (such as polyphenols, catechins, gallic acid, caffeine and theobromine) [15-18]. The few published studies on green tea focused mainly on the prediction of bioactive components. In the literature no studies were found testing NIR spectroscopy for the prediction of free amino acid composition of green tea. Although the linear partial least squares (PLS) analyses is the most common techniques for spectral calibration and prediction [19], with PLSR being the most accurate, other nonlinear techniques e.g. artificial neural network (ANN) have got much less attention and were rarely explored for the spectral analysis in tea sciences. There has been no report on using artificial neural networks for studying the TFAA dependence on the main components of the tea samples.

The objective of this work was to evaluate the ability of NIR spectra to predict the development of TFAA in whole green tea leaves. In addition, the prediction capabilities of different calibration models, including PLS and ANN were investigated and compared for their use in modeling the TFAA. The results of this study may open a wide range of applications in process control or in laboratory routine for the quantification of TFAA in tea.

2   Materials and Methods

2.1   Preparation of Samples

A total of 90 green tea samples were collected from different geographical regions of PR China (i.e. Jiangsu, Henan, Zhejiang, Hunan, Fujian, Anhui, Jiangxi, Sichuan, Guangdong and Yunnan) for experiments. Based on the results of the pre-test, the moisture content of each commercial green tea was 4.0 ± 0.5%. The chemical compositions of these tea samples varied with their species, cultivation conditions such as weather, temperature, moisture, latitude, and season, and process of manufacturing [20]. Samples were stored under controlled circumstance (20°C and 45% relative humidity) before measurements.

2.2   Reference Measurement

The content of TFAA in tealeaves was measured by a ninhydrin colorimetric assay, which can refer to GB/T 8314-2002, a national standard method used to determine free amino acids content of tea infusions in China. The calibration standard is L-theanine, which was obtained from Sigma (St. Louis, MO, USA). In a 25-ml volumetric flask, 1.0 ml of each tea infusions was added. This was followed by addition of 0.5 ml of 1/15 mol/L phosphate buffer solution (pH 8.0) and 0.5 ml of 2% ninhydrin solution containing 0.8 mg/ml of $SnCl_2·2H_2O$. The mixtures in the volumetric flasks were then placed on a boiling water bath for 15 min. The probes were quickly cooled with cold water, and adjusted to 25 ml with water. After they were left to stand still for 10 min, a UV-VIS spectrophotometer (UV-1601, Beijing Rayleigh Analytical Instrument Co., Ltd, China) used to detect the absorbance of the reaction solution in a 10mm light-path cell at 570 nm.

2.3   NIR Spectra Acquisition

An AntarisTM II Method Development Sampling (MDS) system (Thermo Fisher Scientific Inc., Madison, WI, USA) spectrometer was used to measure diffuse reflectance spectra from 10,000 to 4000 cm^{-1} at 4 cm^{-1} resolution by co-adding 32 scans. Spectra were using an integrating sphere and an empty cell as a reference. FT-NIR measurements were performed with a high sensitivity InGaAs detector with a tungsten lamp as the NIR source. The reflectance values (R_v) were converted into absorbance (A) values using the formula $A=log (1/R_v)$ for data analysis.

The standard sample accessory holder was used for performing the tea spectra collection. The sample accessory holder is sample cup specifically designed by Thermo Fisher Scientific. For each tea sample, 10±0.1g of dry tealeaves was filled into the sample cup in the standard procedure depending upon the bulk density of materials. When spectra collecting, tea sample was collected one times every rotating the cup 120 degree angle, thus, each sample was collected three times. The average of these three spectra was used in the next analysis.

2.4   Spectral Data Processing

NIR spectra could be affected by the physical properties of the samples and other environmental noises. Thus, it is necessary to perform mathematical processing to reduce the systematic noise, and enhance the contribution of the chemical composition [21]. Pretreatment is usually performed in order to furnish more parsimonious PLS models, by avoiding the presence of certain spectral trends before regression. In the present work, it was only applied to the experimental NIR data set, where scattering corrections are frequently applied. Spectral derivatives may also be employed to improve resolution and to highlight the selectivity towards a particular analyte when strong multicollinearity is present. Fig.1a presents the raw spectra profile of green tea sample and the raw spectra require to be preprocessed.

   In this study, several spectral preprocessing methods were applied comparatively. These methods are standard normal variate transformation (SNV), multiplicative scatter correction (MSC), first derivative and second derivative. SNV is a mathematical transformation method used to remove slope variation and correct scatter effects in spectra [22]. Each spectrum is corrected individually by first centering the spectral values, and then the centered spectrum is scaled by the standard deviation calculated from the individual spectral values. MSC is another important procedure for the correction of scatter light [23]. It is used to modify the additive and multiplicative effects in the spectra on the basis of different particle sizes. First and second derivatives focus on eliminating baseline drifts and enhancing small spectral differences. Comparing with these preprocessing methods, MSC method was as good as SNV, better than first and second derivatives. The reason was that dry green tea particle solid was easy to bring in scatter light. MSC and SNV spectral preprocessing methods have better ability in correcting light scatter, and can also remove slope variation. MSC pretreated spectra was shown in Fig. 1b.
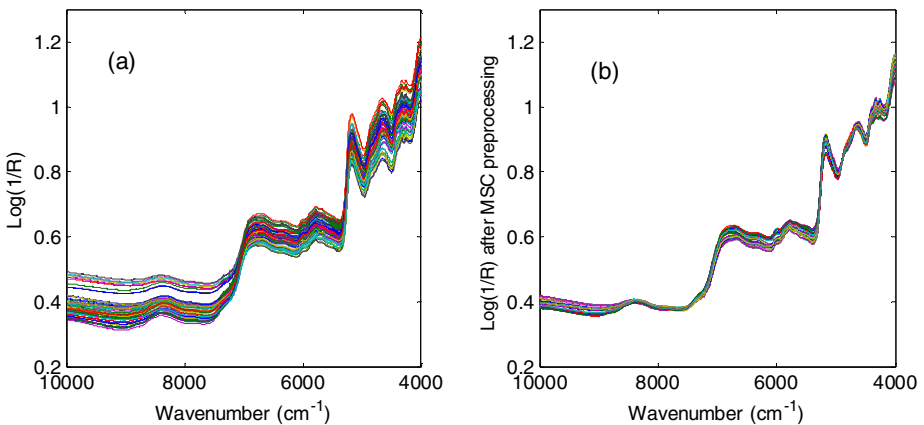
Fig. 1. Spectra of green tea obtained from (a) raw data and (b) MSC preprocessing data

2.5   ANN Description

ANN is multivariate calibration tools, which are especially suited for nonlinear systems. ANN is a powerful analytical tool because of the flexible learning algorithm, diverse network topology, fast learning algorithm and high error tolerance [24]. These advantages will lead to more sophisticated applications of ANN in food science, particularly in sensory analysis and analytical chemistry. Back propagation (BP) is the most commonly used algorithm to train feedforward neural networks for spectral data interpretation [25]. In this study, a multi-layer perception (MLP) based feedforward ANN which uses back propagation learning algorithm, was applied for modeling TFAA concentration.

The network consists of an input layer, hidden layers and an output layer. Fig. 2 illustrates the architecture of the network that is most commonly used with the back propagation algorithm-the multilayer feed-forward network. The determination of number of neurons in hidden layers is very important as it affects the training time and generalization property of neural networks. A higher value of neurons in hidden layer may force the network to memorize the patterns which it has seen during training whereas a lower value of neurons in hidden layer would waste a great deal of training time in finding its optimal representation [26]. There is no general rule for selecting the number of neurons in a hidden layer. The most popular approach to finding the optimal number of neurons in hidden layer is by trial and error [27]. In this study, trial and error approach was used to determine the optimum neurons in hidden layer of the network.
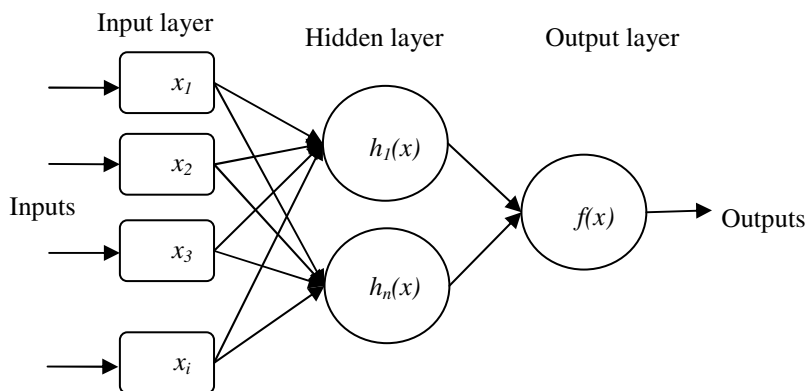
Fig. 2. Illustration of a feed-forward back propagation neural network containing input, hidden and output layers

2.6   Software

All computations were performed using a Matlab (version 7.9) Toolbox (Mathworks Inc., USA) under Windows XP. For the spectral acquisition Result Integration 3.0 (Antaris II System, Thermo Electron Co., USA) was used.

3   Results and Discussion

3.1   Reference Data Description

Green tea samples were divided randomly into calibration sets of 60 samples and prediction sets of 30 samples. Table 1 shows the descriptive statistics (mean, standard deviation and range) for calibration and prediction sets. As seen from Table 1, the range of reference measurements results of TFAA in the calibration set almost covers the range in the prediction set, and their standards deviations between the calibration and prediction sets are no significant differences. The reference values of these four properties had a broad range of variation, which was helpful for the calibrations. Therefore, the distribution of the samples is appropriate both in the calibration and the prediction sets.

Table 1. Reference measurement of free amino acid content (g/g, %) in calibration and prediction sets

Subset	No.[a]	Range	Mean	SD[b]
Calibration set	60	0.976~5.127	2.986	0.905
Prediction set	30	1.236~4.898	2.939	0.883

[a]No: Number of samples; [b]SD: Standard deviation.

3.2   Models Evaluation

The performance of the calibration model was evaluated in terms of the root mean square error of cross-validation (*RMSECV*). For *RMSECV*, a leave-one-sample-out cross-validation was performed: the spectrum of one sample of the calibration set was deleted from this set, and a model was built with the remaining spectra of the calibration set. The left-out sample was predicted with this model, and the procedure was repeated with leaving out each sample of the calibration set. Root mean square error of prediction (*RMSEP*) was used to evaluate the performance of the prediction set in the prediction process. Correlation coefficient (*R*) between the predicted and the measured value were calculated for both the calibration and the prediction sets. Generally, a good model should have lower *RMSECV*, *RMSEP* and higher *R* but small differences between *RMSEP* and *RMSECV*.

3.3   PLS Models

A PLS model performance evaluation was performed using leave-one-out cross validation for the prediction of TFAA in tea samples. It is generally known that the number of PLS factors and the spectra preprocessing are two critical factors in calibrating model for the application of PLS algorithm [18]. In the experiment, the number of PLS factors and the spectral preprocessing were optimized by cross validation, and, they were determined by the lowest value of *RMSECV*. MSC and

SNV spectral pretreatments were used to produce reasonable results for the spectra, respectively. Different PLS factors were applied to build the calibration models, and no outliers were detected in the calibration set during the development of PLS models.

The results of calibration and prediction sets for each model through different preprocessing methods are presented in Table 2. With a comparison of results for calibration and prediction by the aforementioned evaluation standards, the PLS models based on MSC spectral pretreatment turned out to be the optimal results which were achieved as follows: $Rc = 0.927$ and $RMSECV = 0.337$ in the calibration set; $Rp= 0.925$ and $RMSEP = 0.323$ in the prediction set. Fig. 3 is the scatter plot of reference measurements and NIR predicted results by the optimal PLS model.

Table 2. Results of calibration and prediction for PLS and BP-NN models with MSC and SNV spectral pretreatments

Method	Pretreatment	PCs[a]	Calibration		Prediction	
			R_c	$RMSECV(\%)$	R_p	$RMSEP(\%)$
PLS	Raw	8	0.898	0.424	0.890	0.401
PLS	SNV	7	0.925	0.344	0.910	0.356
PLS	MSC	7	0.927	0.337	0.925	0.323
BP-NN	SNV	3	0.972	0.207	0.950	0.265
BP-NN	MSC	4	0.979	0.182	0.958	0.246

[a] principal component factors
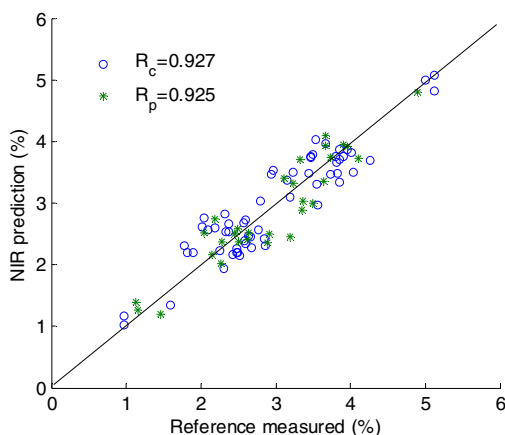
Fig. 3. PLS results: predicted vs. measured TFAA content for calibration (o) and prediction (∗) set

3.4   BP-NN Models

In order to improve the training speed and reduce the training error, after spectra pretreatment, principal components (PCs) issued from principal component analysis (PCA) for dimensionality reduction were applied as input of BP-NN models because the training time increased with the square of the number of training samples and linearly with the number of variables (dimension of spectra). The PCs from the NIR spectra was used as new eigenvectors to TFAA. Four and three PCs were obtained by the criterion for an increment of explained variance lower than 0.25% [28], and the cumulative reliability was 99.39% and 99.28% for MSC and SNV spectral data, respectively.

In the model development by using BP-NN method, three layers of BP-NN were arranged. The number of PCs was also optimized according to cross validation. The output of BP-NN was TFAA content in tea. The hidden nodes were optimized by 'trial and error', and optimal of hidden nodes is evaluated by the minimal mean square error (MSE) value [29]. Several network architectures were tested by varying the number of neurons in the hidden layer with different initial weights. Adjusting by many times, learning rate factor and momentum factor were all set 0.1; all initial weights were set 0.3; scale function was set the 'tanh' function. The permitted regression error was set 0.01, and the maximal iterative epochs were set 1000 times. The optimal performance could be achieved, when 4 PCs and 8 nodes in the hidden layers were included for MSC pretreatment, and when 3 PCs and 8 nodes in the hidden layers were included for MSC pretreatment. With these optimal parameters, the BP-NN model, based on MSC spectral pretreatment, could obtain better prediction results with $Rc = 0.979$ and $RMSECV = 0.182$ in the calibration set, and $Rp = 0.958$ and $RMSEP = 0.246$ in the prediction set. Fig. 4 is the scatter plot of reference measurements and NIR predicted results by the optimal BP-NN model with MSC spectral pretreatment.
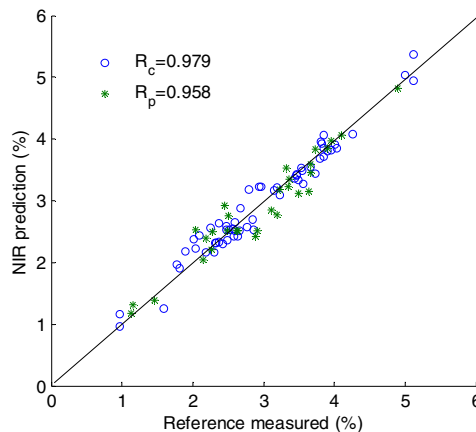
Fig. 4. Predicted values of calibration (o) and prediction (∗) set against reference values using BP-NN method

3.5   Discussion of the Results

The scope of the paper is to compare the performances of linear PLS and nonlinear BP-NN analyses for the prediction accuracy of TFAA in green tea. The PLS models achieved an acceptable results for the prediction of TFAA content of green tea, and these results indicated that it was feasible to utilize spectroscopic technique for the detection and monitoring of free amino acid composition in tea leaves. The BP-NN models performed a better performance than that of PLS models, which indicated that the latent nonlinear information was helpful to improve the prediction performance. The performance of BP-NN models was a little better than that of PLS models. ANNs are suitable for solving nonlinear problems. This suggests that there are nonlinear relationships between the TFAA content and absorption of NIR spectra, more work would be done to discover the useful information or effective wavelength or wavelength bands for the nondestructive determination of TFAA content in green tea leaves.

4   Conclusion

Near Infrared spectroscopy combined with a suitable preprocessing method and with BP-NN regression is shown to be a promising method for predicting free amino acid composition in whole tea samples. The BP-NN models were developed for the prediction of properties of interest, and BP-NN models outperformed PLS models. The correlation coefficient R_p and *RMSEP* in prediction set by BP-NN models were 0.958 and 0.246 for TFAA, respectively. The NIR technique can estimate free amino acid compositions in whole green tea, nondestructively with good accuracy and has enormous potential in improving the efficiency for quality control, evaluation and classification in tea processing.

Based on results achieved in this study, it is recommended to adopt BP-NN analysis as the proper calibration method to predict TFAA in green tea. However, further studies are needed to provide in depth interpretation of the successful measurement of tea properties that do not possess direct spectral response in NIR region.

References

1. Moreda-Pineiro, A., Fisher, A., Hill, S.J.: The classification of tea according to region of origin using pattern recognition techniques and trace metal data. Journal of Food Composition and Analysis 16(2), 195–211 (2003)
2. Chen, Q., Guo, Z., Zhao, J.: Identification of green tea's (Camellia sinensis (L.)) quality level according to measurement of main catechins and caffeine contents by HPLC and support vector classification pattern recognition. Journal of Pharmaceutical and Biomedical Analysis 48(5), 1321–1325 (2008)

3. Naithani, V., Nair, S., Kakkar, P.: Decline in antioxidant capacity of Indian herbal teas during storage and its relation to phenolic content. Food Research International 39(2), 176–181 (2006)
4. Hung, Y.-T., Chen, P.-C., Chen, R.L.C., Cheng, T.-J.: Sequential determination of tannin and total amino acid contents in tea for taste assessment by a fluorescent flow-injection analytical system. Food Chemistry 118(3), 876–881 (2010)
5. Juneja, L.R., Chu, D.-C., Okubo, T., Nagato, Y., Yokogoshi, H.: L-theanine: a unique amino acid of green tea and its relaxation effect in humans. Trends in Food Science & Technology 10(6), 199–204 (1999)
6. Kuroda, Y., Hara, Y.: Antimutagenic and anticarcinogenic activity of tea polyphenols. Mutation Research/Reviews in Mutation Research 436(1), 69–97 (1999)
7. Sakanaka, S., Juneja, L.R., Taniguchi, M.: Antimicrobial effects of green tea polyphenols on thermophilic spore-forming bacteria. Journal of Bioscience and Bioengineering 90(1), 81–85 (2000)
8. Ding, Y., Yu, H., Mou, S.: Direct determination of free amino acids and sugars in green tea by anion-exchange chromatography with integrated pulsed amperometric detection. Journal of Chromatography A 982(2), 237–244 (2002)
9. Alcazar, A., Ballesteros, O., Jurado, J.M., Pablos, F., Martin, M.J., Vilches, J.L.: Differentiation of green, white, black, oolong, and Puerh teas according to their free amino acids content. Journal of Agricultural and Food Chemistry 55, 5960–5965 (2007)
10. Kim, T., Lee, Y.K., Park, S.G., Choi, I.S., Ban, J.O., Park, H.K.: L-theanine, an amino acid in green tea, attenuates β-amyloid-induced cognitive dysfunction and neurotoxicity: Reduction in oxidative damage and inactivation of ERK/p38 kinase and NF-κB pathways. Free Radical Biology and Medicine 47(11), 1601–1610 (2009)
11. Reich, E., Schibli, A., Widmer, V., Jorns, R., Wolfram, E., DeBatt, A.: HPLC methods for identification of green tea and green tea extract. Journal of Liquid Chromatography and Related Technologies 29, 2141–2151 (2006)
12. Wang, L., Xu, R., Hu, B., Li, W., Sun, Y., Tu, Y.: Analysis of free amino acids in Chinese teas and flower of tea plant by high performance liquid chromatography combined with solid-phase extraction. Food Chemistry 123(4), 1259–1266 (2010)
13. Aucamp, J.P., Hara, Y., Apostolides, Z.: Simultaneous analysis of tea catechins, caffeine, gallic acid, theanine and ascorbic acid by micellar electrokinetic capillary chromatography. Journal of Chromatography A 876, 235–242 (2000)
14. Kato, M., Gyoten, Y., Sakai-Kato, K., Toyooka, T.: Rapid analysis of amino acids in Japanese green tea by microchip electrophoresis using plastic microchip and fluorescence detection. Journal of Chromatography A 1013(1), 183–189 (2003)
15. Luypaert, J., Zhang, M.H., Massart, D.L.: Feasibility study for the use of near infrared spectroscopy in the qualitative and quantitative analysis of green tea, Camellia sinensis (L). Analytica Chimica Acta 478(2), 303–312 (2003)
16. Chen, Q., Zhao, J., Zhang, H., Wang, X.: Feasibility study on qualitative and quantitative analysis in tea by near infrared spectroscopy with multivariate calibration. Analytica Chimica Acta 572(1), 77–84 (2006)
17. Sinija, V.R., Mishra, H.N.: FT-NIR spectroscopy for caffeine estimation in instant green tea powder and granules. LWT - Food Science and Technology 42(5), 998–1002 (2009)
18. Chen, Q., Zhao, J., Chaitep, S., Guo, Z.: Simultaneous analysis of main catechins contents in green tea (Camellia sinensis (L.)) by Fourier transform near infrared reflectance (FT-NIR) spectroscopy. Food Chemistry 113(4), 1272–1277 (2009)

19. Kasemsumran, S., Du, Y.P., Maruo, K., Ozaki, Y.: Improvement of partial least squares models for in vitro and in vivo glucose quantifications by using near-infrared spectroscopy and searching combination moving window partial least squares. Chemometrics and Intelligent Laboratory Systems 82(2), 97–103 (2006)
20. Ngetich, W.K., Stephens, W.: Responses of tea to environment in kenya. 1. Genotype environment interactions for total dry matter production and yield. Experimental Agriculture 37(3), 333–342 (2001)
21. Lin, H., Chen, Q., Zhao, J., Zhou, P.: Determination of free amino acid content in Radix Pseudostellariae using near infrared (NIR) spectroscopy and different multivariate calibrations. Journal of Pharmaceutical and Biomedical Analysis 50(5), 803–808 (2009)
22. Fearn, T., Riccioli, C., Garrido-Varo, A., Guerrero-Ginel, J.E.: On the geometry of SNV and MSC. Chemometrics and Intelligent Laboratory Systems 96(1), 22–26 (2009)
23. Ely, D.R., Thommes, M., Teresa Carvajal, M.: Analysis of the effects of particle size and densification on NIR spectra. Colloids and Surfaces A: Physicochemical and Engineering Aspects 331(1), 63–67 (2008)
24. Huang, Y., Kangas, L.J., Rasco, B.A.: Applications of artificial neural networks (ANNs) in food science. Critical reviews in food science and nutrition 47(2), 113–126 (2007)
25. Liu, Y., Sun, X., Ouyang, A.: Nondestructive measurement of soluble solid content of navel orange fruit by visible–NIR spectrometric technique with PLSR and PCA-BPNN. LWT - Food Science and Technology 43(4), 602–607 (2010)
26. Benardos, P.G., Vosniakos, G.-C.: Optimizing feedforward artificial neural network architecture. Engineering Applications of Artificial Intelligence 20(3), 365–382 (2007)
27. Abutaleb, A.S.: A neural network for the estimation of forces acting on radar targets. Neural Networks 4(5), 667–678 (1991)
28. Urbano-Cuadrado, M., de Castro, M.D.L., Pérez-Juan, P.M., García-Olmo, J., Gómez-Nieto, M.A.: Near infrared reflectance spectroscopy and multivariate analysis in enology: Determination or screening of fifteen parameters in different types of wines. Analytica Chimica Acta 527(1), 81–88 (2004)
29. Svozil, D., Kvasnicka, V., Pospichal, J.: Introduction to multi-layer feed-forward neural networks. Chemometrics and Intelligent Laboratory Systems 39(1), 43–62 (1997)

Daily Reference Evapotranspiration Estimation Based on Least Squares Support Vector Machines

Dachun Chen[*]

Hehai University, Nanjing 210098, P.R. China,
Water Conservancy and Civil Engineering College,
Xinjiang Agricultural University, Urumqi 830052, P.R. China
vision_studio@sina.com

Abstract. As a key hydrological parameter, daily reference evapotranspiration (ETo) determines the accuracy of the hydrological number of the crop, and, consequently, the regional optimization disposition of water resources. At present, the main methods for ETo estimation are the Penman-Monteith (PM) equation and its modified formula, both of which are based on climatic factors such as temperature, radiation, humidity, and wind velocity, among others. Unfortunately, these required data are not always available in Xinjiang Uighur Autonomous Region, China, which is a semiarid area. Hence, this paper puts forward, for the first time, a least squares support vector machine (LSSVM) model for estimating ETo. The LSSVM model used in this study considers climatic factors as input variables and the ETo calculated by the Penman-Monteith equation as an output variable. Compared with the artificial neural network (ANN) model, which was developed with the same data, LSSVM prediction shows higher accuracy, efficiency, and generalization performance. Therefore, it can be used as a complementary ETo estimation method.

Keywords: Daily reference evapotranspiration, LSSVM, ANN, Semiarid area, Xinjiang.

1 Introduction

Accurate evapotranspiration (ETo) estimates are necessary to determine the water requirement of crops for irrigation scheduling. The only recommended method for ETo estimation is the physically-based complex Penman-Monteith (PM) equation using complete meteorological data [1]. The ANN model has been successfully applied to evapotranspiration estimation. Kumar [2] built an ANN model for ETo estimation and concluded that it could predict ETo better than the conventional method, that is, the PM equation. Trajkovic [3] applied a sequentially adaptive radial basis function (RBF) network for ETo forecasting. As well, many studies have used neural networks to model ETo as a function of climatic variables [4–12].

---

[*] Corresponding author.

The support vector machine (SVM) approach was recently recognized for its ability to capture nonlinear regression relationships between variables [13]. Eslamian and Kisi [13,14] investigated the accuracy of SVM in modeling ETo, and their comparison results revealed that the SVM could be successfully used in modeling the ETo process.

This paper discusses ETo estimation by using a least squares support vector machine (LSSVM). In comparison with ANN and SVM, it shows higher accuracy, efficiency, and generalization performance.

2 Materials and Methods

2.1 Investigation Areas

The 145th Regiment of Shihez Reclaimation Area, located south of the XinJiang Uighur Autonomous Region, was explored in the present study. The data sets were gathered from meteorological stations located at east longitude 86.05°, north latitude 44.32°, and altitude of 443.7 m. Data for the following weather variables used for this study were gathered from 1998 to 2009: precipitation (mm), relative humidity (%), wind speed (m/s), and maximum and minimum temperature (^{o}C).

2.2 Estimation of Reference Evapotranspiration

The Penman-Monteith equation for calculation of ETo proposed by Allen [1] is expressed as:

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma\dfrac{900}{T+273}u_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \tag{1}$$

where ETo is the reference evapotranspiration [mm day^{-1}], R_n is the net radiation at the crop surface [MJ m^{-2} day^{-1}], G is the soil heat flux density [MJ m^{-2} day^{-1}], T is the mean daily air temperature at 2 m height [°C], u_2 is the wind speed at 2 m height [m s^{-1}], e_s is the saturation vapor pressure [kPa], e_a is the actual vapor pressure [kPa], e_s - e_a is the saturation vapor pressure deficit [kPa], Δ is the slope vapour pressure curve [kPa °C^{-1}], and γ is the psychrometric constant [kPa °C^{-1}].

2.3 LSSVM and Models Evaluation [15, 16]

Suppose that we are given a training data set of n data points $\{x_i, y_i\}_{i=1}^{n}$, where $x_i \in R^n$ is the i-th input vector and $y_i \in R$ is the corresponding i-th target. For

binary classification problems, yi takes only two possible values {-1, +1}, whereas yi takes any real value for regression problems. The goal is to find a function $V : x \rightarrow y$ that minimizes the residuals for the given data and generalizes well with unseen data. We transform the input patterns into the reproducing kernel Hilbert space (RKHS) by a set of mapping functions $\phi(x)$ [16]. An inner product in the feature space has an equivalent kernel in input space, $K(x, x') = \{\phi(x), \phi(x')\}$, provided certain conditions hold. K is a symmetric positive definite function, which satisfies Mercer's conditions. To learn the unknown function, we solve a Tikhonov functional of the special form:

$$\min_{w,b,e} P(w,e) = \frac{1}{2} w^T w + C \sum_{i=1}^{n} e_i^2$$
$$s.t. \quad f(x) = w^T \phi(x_i) + b + e_i$$
$$i = 1,2,\cdots n \tag{2}$$

where $C > 0$ is the regularization factor, e_i is the residual between the output y_i and $f(x_i)$, w is the weight vector, and b is called the bias term.

Using the Lagrangian multiplier method, the Lagrangian for (2) is:

$$L(w,b,e;\alpha) = P(w,e) - \sum_{i=1}^{n} \alpha_i \{w^T \phi(x_i) + b + e_i - y_i\} \tag{3}$$

where $\alpha_i, i = 1,2,\cdots n$ are the Largrangian multipliers corresponding to (2). The Karush-Kuhn-Tucker (KKT) conditions (3) are:

$$\begin{cases} \dfrac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{n} \alpha_i \phi(x_i), \\ \dfrac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{n} \alpha_i = 0, \\ \dfrac{\partial L}{\partial e_i} = 0 \rightarrow \alpha = Ce_i, \\ \dfrac{\partial L}{\partial \alpha_i} = 0 \rightarrow w^T \phi(x_i) + b + e_i - y_i = 0, \\ \qquad i = 1,2,\cdots n \end{cases} \tag{4}$$

In the numerical solution proposed by Suykens [12], the KKT conditions of (3) are reduced to a linear system by eliminating w and e, resulting in:

$$\begin{bmatrix} 0 & \vec{1}^T \\ \vec{1} & K(x_i, x_j) + C^{-1}I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{5}$$

where $y = [y_1, y_2, \cdots, y_n]^T, \alpha = [\alpha_1, \alpha_2, \cdots, \alpha_n]^T, \vec{1} = [1,1,\cdots,1]^T$. Note that $K(x_i, x_j) + C^{-1}I$ is symmetric and positive definite since the matrix K is semi-positive definite and the diagonal term C^{-1} is positive. Solving (5) for α and b, the discriminate function can be obtained from $f(x) = \sum_{i=1}^{n} \alpha K(x, x_i) + b$.

Four common choices of kernels are: $K(x, z) = x^T z$ (linear kernel), $K(x, z) = (r + x^T z)^d$ (polynomial kernel of degree d), $K(x, z) = e^{\frac{-\|x-z\|_2^2}{\sigma^2}}$ (RBF kernel), and $K(x, z) = \tanh(k_1 x^T z + k_2)$ (multilayer perceptron kernel).

The performance of the different models is evaluated based on the criteria of the root mean square error (RMSE) and square value of coefficient of correlation r. These two statistical parameters used for the performance evaluation are given as follows [11]:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_p - y_e)^2}{n}} \tag{6}$$

$$r = \frac{\sum_{i=1}^{n}(y_p - \bar{y}_p)(y_e - \bar{y}_e)}{\sqrt{\sum_{i=1}^{n}(y_p - \bar{y}_p)^2 \sum_{i=1}^{n}(y_e - \bar{y}_e)^2}} \tag{7}$$

where y_p and y_e represent PM method and temperature-based model ETo values estimated for the i-th values, $\bar{y}_p$ and $\bar{y}_e$ represent the average values of the corresponding variables, and n represents the number of observations.

2.4 Data Normalization

In this study, limited climatic data gathered from 1998 to 2009 were used. Daily average data of these 12 years in the 145th Regiment, making up 2568 data points,

were divided into three parts for the purposes of training (60%), validation (20%), and testing (20%). Note that only those data in the crop growth period (March–September) were used in this study. In order to overcome the negative result associated with extreme values, input and output data sets were scaled in the range [0, 1] using the following equation:

$$y_{norm} = \frac{y_i - y_{min}}{y_{max} - y_{min}} \tag{8}$$

where y_{norm} is the normalized dimensionless variable, y_i is the observed value of variable, and y_{min} and y_{max} are the minimum and maximum values of the observed variable.

3   Discussion of Results

For similar bases of comparison, the same training and verification sets were used for the ANN and LSSVM models. RMSE and r^2 were employed to evaluate the performances of the models developed. In this study, the ANN model architecture was 4–10–1, the Levenberg-Marquardt algorithm parameter was 0.002, and the LSSVM model kernel function used was RBF with parameters $(C, \varepsilon, \sigma) = (50, 0.0001, 1.5)$.

A summary of the statistical performances of the different models is presented in Table 1. Compared with the ANN model's consumption time (119.78 s), the LSSVM model shows better performance (35.265 s). Results from the statistical criteria indicate nearly the same fitting degree r^2 (0.986–0.987), but the LSSVM model performs better than ANN model in terms of RMSE.

To illustrate the model estimation performance, Figure 1 shows ANN and LSSVM ETo estimates with 642 nodes (March 2006–April 2009) configuration and PM ETo. The figure shows that both ANN- and LSSVM-estimated ETo values agree closely with the PM-estimated value and follow the same trend. LSSVM also has less deviation around the ETo peak compared with ANN, and shows better generalization capacity due to its Vapnik-Chervonenkis (VC) theory base.

Table 1. ANN and LSSVM models and their coefficients of correlation for the training and validation periods in the 145th regiment

Model	r2			RMSE			Consumption Time(s)
	Training	Validation	Testing	Training	Validation	Testing	
ANN	0.996	0.994	0.986	4.212	4.317	3.189	119.78
LSSVM	0.996	0.995	0.987	0.112	0.115	0.117	35.265
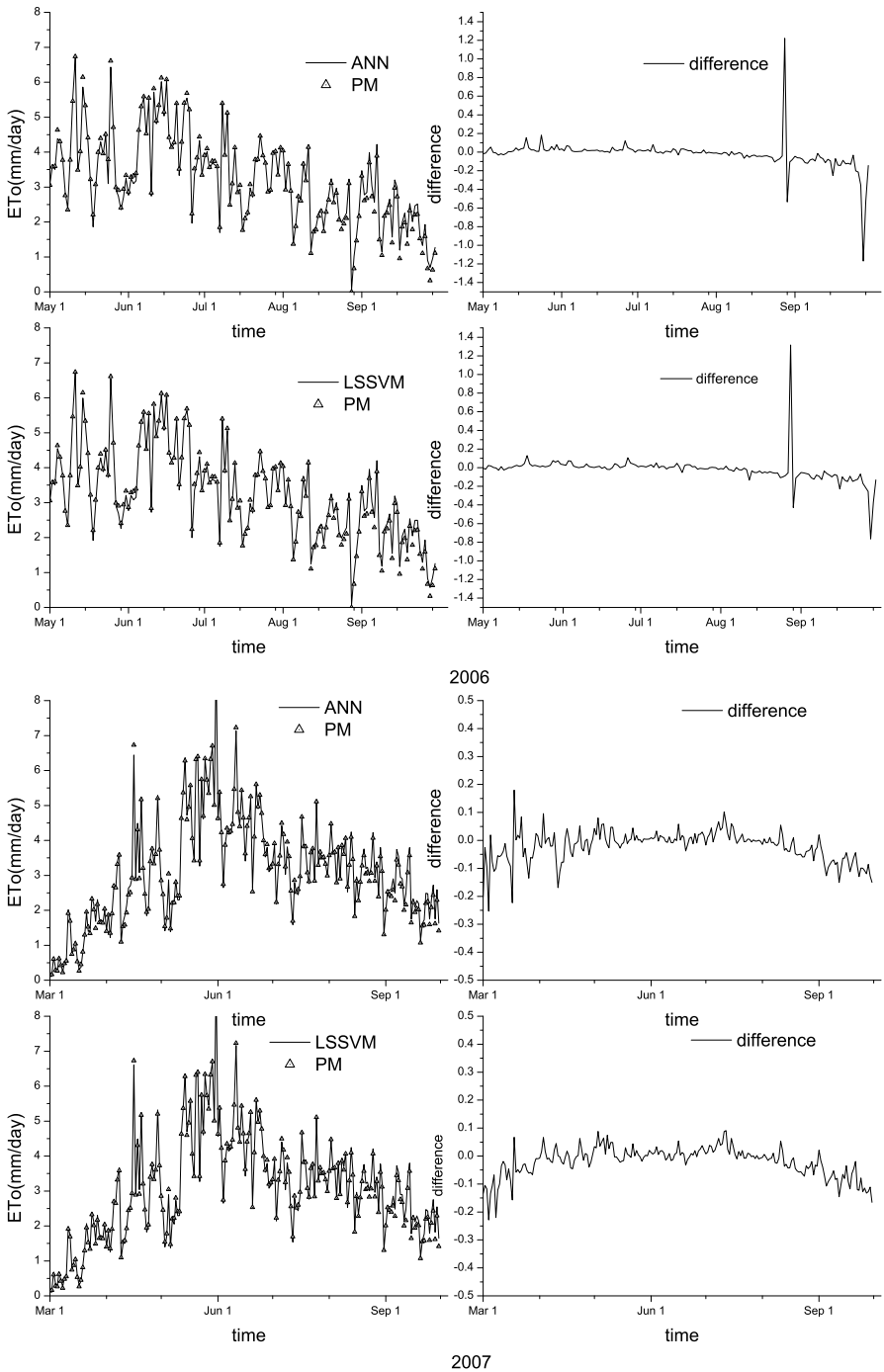
Fig. 1. Comparison of ETo values calculated by the proposed ANN, LSSVM, and PM methods

2008
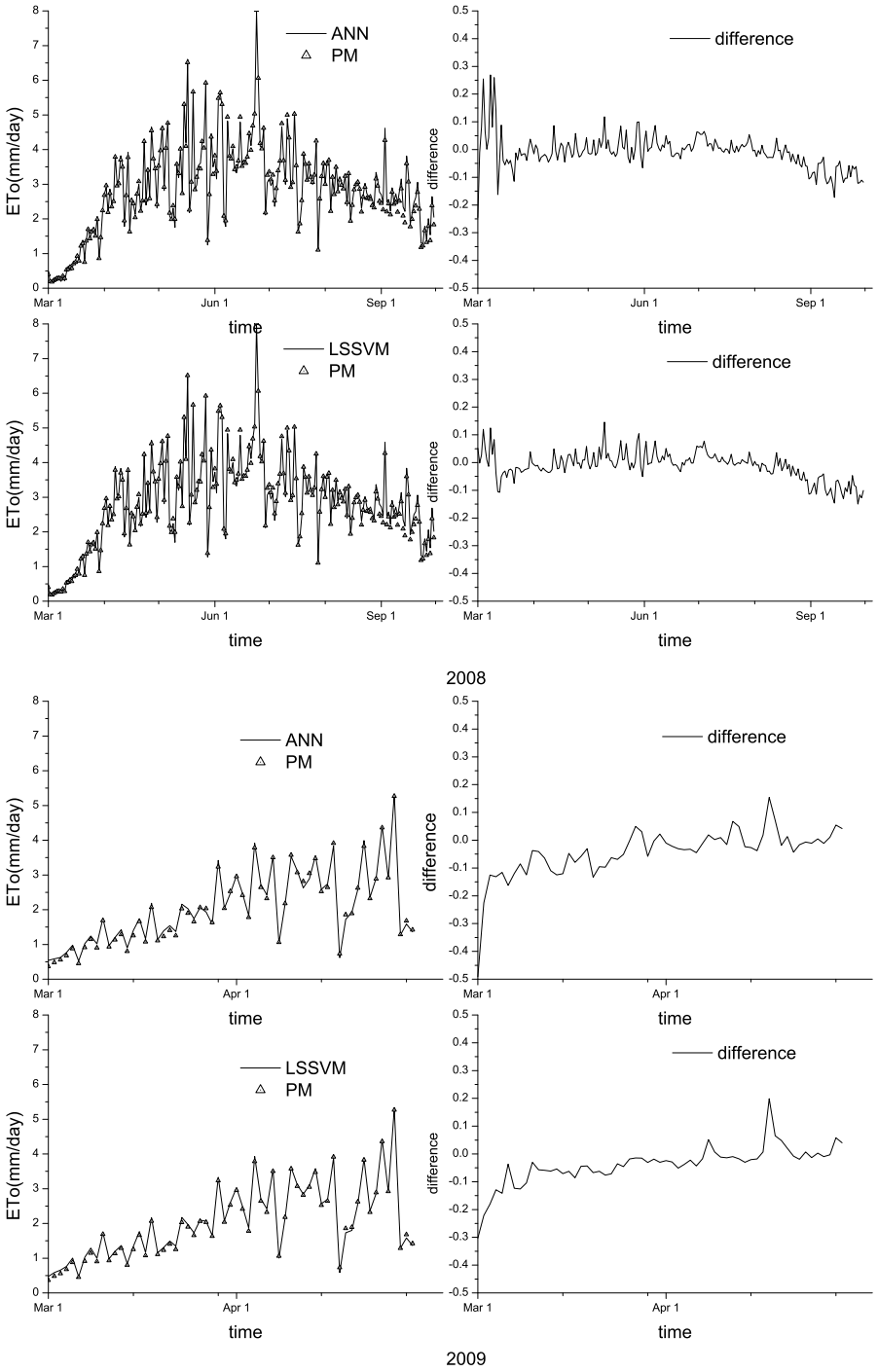
2009

Fig. 1. (*Continued*)

Table 2. Comparison data in April 2009

Date	PM	ANN		LSSVM	
		Testing	Error	Testing	Error
2009-4-1	2.954	2.985	-0.010	3.025	-0.024
2009-4-2	2.413	2.465	-0.022	2.482	-0.029
2009-4-3	1.779	1.833	-0.030	1.870	-0.051
2009-4-4	3.783	3.912	-0.034	3.926	-0.038
2009-4-5	2.645	2.730	-0.032	2.704	-0.022
2009-4-6	2.320	2.425	-0.045	2.420	-0.043
2009-4-7	3.499	3.542	-0.012	3.560	-0.017
2009-4-8	1.058	1.038	0.019	1.004	0.051
2009-4-9	2.179	2.174	0.002	2.163	0.007
2009-4-10	3.572	3.537	0.010	3.612	-0.011
2009-4-11	3.070	3.118	-0.016	3.112	-0.014
2009-4-12	2.816	2.626	0.067	2.843	-0.009
2009-4-13	3.044	2.894	0.049	3.097	-0.017
2009-4-14	3.473	3.554	-0.023	3.578	-0.030
2009-4-15	2.520	2.586	-0.026	2.571	-0.020
2009-4-16	2.643	2.743	-0.038	2.690	-0.018
2009-4-17	3.913	3.840	0.019	3.884	0.007
2009-4-18	0.730	0.617	0.154	0.584	0.199
2009-4-19	1.854	1.722	0.071	1.733	0.065
2009-4-20	1.888	1.923	-0.019	1.796	0.048
2009-4-21	2.621	2.586	0.013	2.568	0.021
2009-4-22	3.826	3.991	-0.043	3.856	-0.008
2009-4-23	2.321	2.358	-0.016	2.366	-0.019
2009-4-24	2.885	2.904	-0.007	2.864	0.007
2009-4-25	4.359	4.407	-0.011	4.417	-0.013
2009-4-26	2.922	2.908	0.005	2.913	0.003
2009-4-27	5.267	5.327	-0.011	5.319	-0.010
2009-4-28	1.284	1.271	0.011	1.288	-0.003
2009-4-29	1.678	1.586	0.055	1.580	0.058
2009-4-30	1.420	1.360	0.042	1.363	0.040

4   Conclusions

In this work, LSSVM is proposed to be a novel technique for nonlinear function approximation. It is a very specific type of learning algorithm characterized by the capacity control of the decision function, the use of the kernel functions, and the sparseness of its solution. Established on the unique theory of structural risk minimization to estimate a function by minimizing the upper boundary of the generalization error, LSSVM is very resistant to the over-fitting problem, eventually achieving high generalization performance in solving various nonlinear function approximation problems. Another key property of LSSVM is that training it is equivalent to solving linear systems; its solution is always unique and globally optimal, unlike the training of other networks, which requires non-linear optimization with the danger of getting stuck with local minima.

An LSSVM model was developed to estimate ETo based on an extensive number of data sets with limited climatic information. Its superiority was demonstrated by comparison with the ANN model. All the simulation results showed that the LSSVM model is more effective and efficient in estimating ETo than the ANN model.

Although this paper showed the effectiveness of the LSSVM model, several issues require further investigation. Firstly, how to determine the hyperparameters of LSSVM is an important issue that needs to be addressed. Secondly, in this study, only the RBF kernel function was investigated. Additional research is necessary to explore more useful kernel functions to improve the performance of the LSSVM model.

References

1. Allen, R.G., Pereira, L.S., Raes, D., Smith, M.: Crop evapotranspiration: guidelines for computing crop water requirements. In: Proceedings of the Irrigation and Drainage Paper No. 56. Food and Agricultural Organization, United Nations, Rome (1998)
2. Kumar, M., Raghuwanshi, N.S., Singh, R., et al.: Estimating evapotranspiration using artificial neural network. Journal of Irrigation and Drainage Engineering 128, 224–233 (2002)
3. Trajkovic, S., Todorovic, B., Stankovic, M.: Forecasting of reference evapotranspiration by artificial neural networks. Journal of Irrigation and Drainage Engineering 129, 454–457 (2003)
4. Odhiambo, L.O., Yoder, R.E., Yoder, D.C., et al.: Optimization of fuzzy evapotranspiration model through neural training with input-output examples. Trans. ASAE American Society of Agricultural and Biological Engineering 44, 1625–1633 (2001)
5. Aksoy, H., Guven, A., Aytek, A., et al.: Discussion of Generalized regression neural networks for evapotranspiration modeling. Hydrological Sciences Journal 52, 825–831 (2007)
6. Kişi, O.: Generalized regression neural networks for evapotranspiration modeling. Hydrological Sciences Journal 51, 1092–1104 (2006)
7. Kişi, O., Ozturk, O.: Adaptive Neurofuzzy Computing Technique for Evapotranspiration Estimation. Journal of Irrigation and Drainage Engineering 133, 368–379 (2007)

8. Gorka, L., Amaia, O.B., Jose, J.L.: Comparison of artificial neural network models and empirical and semi-empirical equations for daily reference evapotranspiration estimation in the Basque Country (Northern Spain). Agricultural Water Management 95, 553–565 (2008)
9. Ali, R.K.: Comparative study of Hargreaves's and artificial neural network's methodologies in estimating reference evapotranspiration in a semiarid environment. Irrigation Science 26, 253–259 (2008)
10. Sungwon, K., Hung, S.K.: Neural networks and genetic algorithm approach for nonlinear evaporation and evapotranspiration modeling. Journal of Hydrology 351, 299–317 (2008)
11. Wang, Y.-M., Seydou, T., Tienfuan, K.: Neural network approach for estimating reference evapotranspiration from limited climatic data in Burkina Faso. WSEAS Transactions on Computers 7, 704–713 (2008)
12. Seydou, T., Wang, Y.-M., Tienfuan, K.: Artificial neural network for modeling reference evapotranspiration complex process in Sudano-Sahelian zone. Agricultural Water Management 97, 707–714 (2010)
13. Eslamian, S.S., Abedi-Koupai, J., Amiri, M.J., et al.: Estimation of daily reference evapotranspiration using support vector machines and artificial neural networks in greenhouse. Res. J. Env. Sci. 3, 439–447 (2009)
14. Kisi, O., Cimen, M.: Evapotranspiration modelling using support vector machines. Hydrological Sciences Journal 54, 918–928 (2009)
15. Vapnik, V.N.: The Nature of Statistical Learning. Springer, Berlin (1995)
16. Suykens, J.A.K., Vandewale, J.: Least squares support vector machine classifiers. Neural Processing Letters 9, 293–300 (1999)

Data Analysis of Cold Rice Blast
Based on Near Infrared Spectroscopy

Feng Tan[1], Xiaodan Ma[1], Chun Wang[2,*], and Tingyi Shang[1]

[1] College of Information Technology, Heilongjiang Bayi Agricultural University
DaQing, China 163319
[2] College of Engineering, Heilongjiang Bayi Agricultural University
DaQing, China 163319

Abstract. In order to overcome the low diagnostic accuracy of traditional blast visual defects, set cold rice blast for example, four data groups has been acquainted through near infrared spectroscopy, the first group is healthy and diseased planting stock; the second group is three diseased levels of leaf blast; the third group is five diseased levels of grain blast; the forth group is four diseased levels of panicle blast, according to the analysis, we can know that different circumstances of plants had their own near-infrared spectral bands which made the preliminary basis for real-time detection of cold Rice blast.

Keywords: Cold Rice, Blast, Near Infrared Spectroscopy, Disease Detection.

1 Introduction

The rice planting area in China accounts for 1/4 of the food crops area, While the production close to 1/2 of the national's total grain output, and over 1/2 of the commodity, So rice plays a very important role in China's grain production. Rice blast is one of the most important diseases in rice production, Widely distributed in the world rice area in the whole growth period of rice causing widespread damage. Blast is a major disease in the production of rice production in cold land, Year of the disease cuts 23% of the rice production in cold, severe up to 45%, or even never produced. Among them, the rice leaf blast and the rice ear and neck blast caused the largest loss. The annual incidence rate of leaf blast is 40-50%, Up to 80% when it is very serious; Up to 50% of neck blast. Traditional method of classification detection of rice disease resistance is based on visual concept of the disease by plant protection experts, it has drawbacks of subjective factors, and time-consuming. Therefore, there is an urgent need for a new technology to solve the technical problems in cold blast resistance classification, In modern agricultural production, there's an increasingly high demand for the frequency, speediness and accuracy of work in the forecast on cold blast, classification and the evaluation of prevention work. This claims a new research of cold rice disease diagnosis technology.

In recent years, for the advantages of non-destructive, fast and accurate , multi-spectral, The technology [1-5] is more and more used in the field of plant diseases detection. Adams et al [6] used 550nm, 650 and 750nm spectral image to detect trace elements in vegetables; Bravo [7] detected the wheat yellow rust in the visible and near infrared bands ;Yang Z et al [8] ,with the help of the 16 band spectral analyzer studied the wheat canopy invasion which was Infringed by Schizaphis germanium .In 1994,Qi-Ying Lin [9] and other Scientists imposed a laser Raman spectrometer to take a Raman scattering experiment towards the clusters of rice dwarf virus in three samples which were purified ,The results showed that the intensity of different viruses is different. mechanism of action which the Small molecule cysteine stay in the silver surface, Hao Yaqiong 2002 [10] using surface-enhanced Raman scattering discussed and studied the adsorption method, the use of molecular spectroscopy lays the foundation that explores the mechanism and further improve the enzyme activity. (In)2002, Cheng Hongyan [11] found that C60 derivatives which its Raman spectra is significantly different from the original in Raman spectroscopy study ,Analysis of the emergence of organic functional groups makes the C60 changes the molecular structure.

In 2003 Yu duowei [12] utilized Raman spectra analysis whether the acid induce the part of the DNA purine, pyramiding fall off,And whether the reasons for the protonated pyramiding purine could be stronger than the relevant. In 2003, Exotic [13] and others used the Raman scattering method to study the quality of the SiC single crystal ,which can improve Lely.they found that the structure of the sample found is 6H-SiC,also pointed out that more defects exist in the sample, gave the Raman spectra of SiC within 100-4000cm-1 for the first time, and detected the crystal structure successfully with the Raman spectroscopy. In 2010, Kang Yi Pu [14] and other people used the portable Raman spectrometer to measure the silver membrane Xoo seven races of the surface enhanced Raman spectroscopy (SERS),The results showed that the seven races of the SERS spectra in the peak position and relative intensity of the peak which has many differences, They Proved that this method can quickly and easily test the different Xoo races.Wu Di et al [15] technology achieve the early detection of Botrytis cindered eggplant leaves with the near infrared spectroscopy. Infected with the blast of cold rice, The changes of their Agencies within the physiological and external shape of each band is bound to cause the differences in spectral reflectance characteristics of change. Based on a combination of these changes，we can create a model for disease diagnosis. In this paper, We combine the advantages of the spectral analysis and image processing technology，adopt multi-spectral vision technology to explore identified classification test on disease resistance of rice varieties, The work provide technical support for automated identification .

2   Materials and Methods

2.1   Sample Selection

Experimental samples were taken from experimental plots of rice in cold region in Heilongjiang Reclamation, during the experiment, four groups of experimental

samples were selected, healthy and infected plants in different parts of plants; infected leaves of the three disease rating; infected grain level of the five diseases; infected stems of four disease rating.

2.2   Experimental Apparatus

This experiment use WQF-600N FT-NIR spectrometer, the instrument is a measurement of material absorption of infrared radiation (or transmittance) of the analytical instruments. Since each substance has a characteristic absorption spectrum - there are only certain wavelengths to absob while other wavelengths is not absorbed, so the characteristics of the absorption spectra can be used to carry out qualitative analysis of substances. In addition, the total amount of material is proportional with the absorption, the absorption spectra can also make use of quantitative analysis of substances. Wave number range: 3300cm-1 ~ 10000cm-1; resolution: better than 4cm-1; wave number accuracy: better than the resolution set by the 1 / 2; transmission repeatability: 0.5% T; 100% T line of the letter noise ratio: 4300cm-1 ~ 4400cm-1, 32 scans (equivalent to 1 minute measurements), S / N better than 10000:1 (RMS value); splitter: CaF2; detector: DTGS (Pubs, InGaAs can beelection); source: air-cooled tungsten light source; Data Systems: General computer, connect inkjet printer or laser printer can output high-quality spectra; Software: Chinese Windows operating system under the common operating software systems, including spectrum acquisition offer, spectral transform, spectral processing, spectral analysis, construction the conventional mode of operation, and all other functions, which can provide both chemo metric software. Working with a lot of third party software in WINDOWS run together under the same time, users can also develop new operational procedures of spectral data in their needs .

Data Acquisition and Analysis

3   Healthy Plants and Infected Plants in Different Parts of Data Collection

In order to study cold Blast spectral characteristics, In the Experiment, We first study the different parts between healthy plants and infected plants, stems, leaves and grains. The spectral characteristics is Obtained by near infrared spectroscopy instruments projection. The result is shown in Figure 1 and Figure 2 below.

Figure 1 shows what we select is rice plant stems, leaves and grains were compared with the spectrum, from which we can see that the transmittance of the spectrum of three parts are not in the same position; Figure 2 shows the spectrum of plant stems, leaves and seeds with a serious blast of damage. Figure 1 and Figure 2 shows the comparison which can be seen that stems transmission are relatively low, while the leaves and transmittance spectra of grain almost the same intensity, Wave number from the variation of transmittance on the stem and the transmittance spectrum of grain is in good agreement.
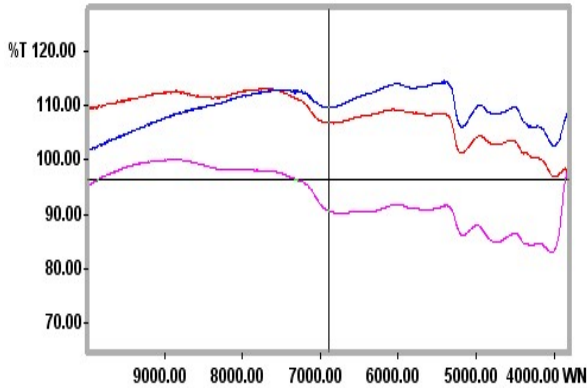
Fig. 1. spectrum contrast between healthy stem, leaf and grain

3.1   Blast Leaves of Three Levels of Data Collection Disease

According to different diseases of rice plant of infrared spectrum acquisition, the collected results are shown in Figure 3.Three lines from top to bottom in the figure correspond to the three diseases of rice plants in order to reduce the transmittance spectrum. It can be seen from the figure, leaves transmittance in the race's serious
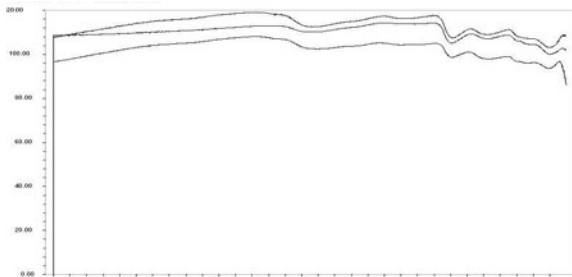
Fig. 2. spectrum contrast between diseased stem, leaf and grain
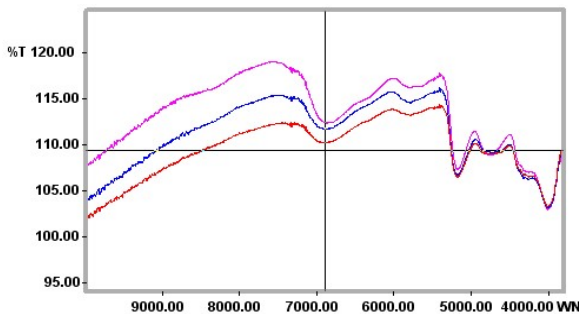
Fig. 3. Degree Comparison of three leaf blast disease

incidence is the strongest, Bottom curve is the transmittance spectra(of) healthy plants, and its transmission strength is the weakest. And from the general trend, as the wave number decreases, the difference become smaller between transmission intensity, and local transmission in a large wave number is relatively large different In other words, in the near infrared band, as the wavelength increases gradually disease transmission in different degrees of intensity become not obvious. when it is close to 4000nm, transmission rate of diseased leaves and no disease is very close.

3.2   Five Diseased Levels of Infected Seed of Data Collection

Figure 4 shows the near infrared transmittance spectroscopy of the five grains , Contrast can be seen in Figure 3, The absorption peak in Grains is very high consist with rice leaves. The reason is that leaves organic components and grain shell organic components is very similar so the absorption peak location will be the same. Although part of the near-infrared light into the leaves or seeds inside, the grain ,is also the seed inside part of the internal components and the internal components inside leaves are not still change, so the absorption spectrum of the intensity at each wavelength is different. It can be seen from Figure 4, five different grains in case of blast injury of the transmission spectrum do not show very good consistency in parallel. according to analysis, it is because in the acquisition of spectrum, There are too many grains entering into the loading room, and with random arrangement grains, also the near infrared spectrometer is smaller than the grain size of the spot, so every time in the collection process, light pass through the outer surface of grain and incident angle of the interface is different, They lead to the light penetration depth in the grain interior change, so the curve characteristics in the Characterization of grain display inconsistency.

From the analysis view, although the five kinds grains of spectral curves are irregular, due to different levels of disease, in some bands they have relatively strong discrimination, and also five curves do not cross too many points. If the differential sum rule act on the entire point ,it can eliminate part of the   interference-free spectrum.
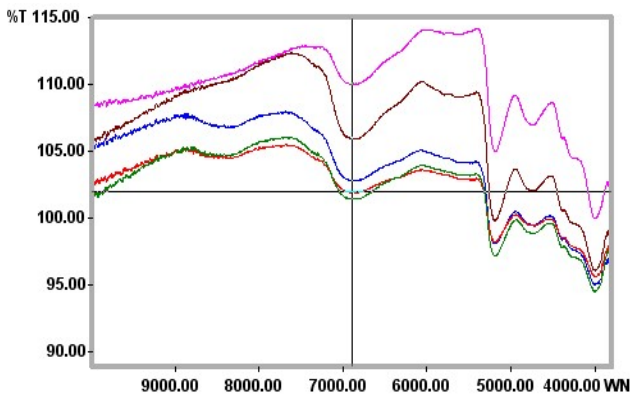
Fig. 4. Blast Grain contrast between five infected level

3.3   Four Diseased Levels of Infected Stems of Data Collection

It can be seen from figure 5,like the leaves and seeds, the stem has a feature that the transmission of serious rice blast is higher than that of normal plants in the range of 4000WN ~ 10000WN.Besides,the regularity of the stem and seed is better than that of leaf, differences of stems in 40000WN ~ 70000WN are more obvious while differences in grains are more obvious in the 7000WN ~ 10000WN.In race's growing process, the changes of stem and seed are slight, only change with some reasons such as illness or under the influence of external transformation. Therefore, it is more reliable and more regularity among detection of the disease. The leaves and the light, moisture and other factors related larger, such as leaves in absence of sufficient water performed significantly, so the regularity is relatively poor. Although the parallel spectrum and regularity are not strong, the same conclusion still can be drawn from the analysis. Integrated to analyze, It will be more effective to diagnose rice blast, and more easier to analyze quantitatively of the rice blast, if it is possible to analyze the rice blast combine with the strongest differences of stem, leaves and seeds.
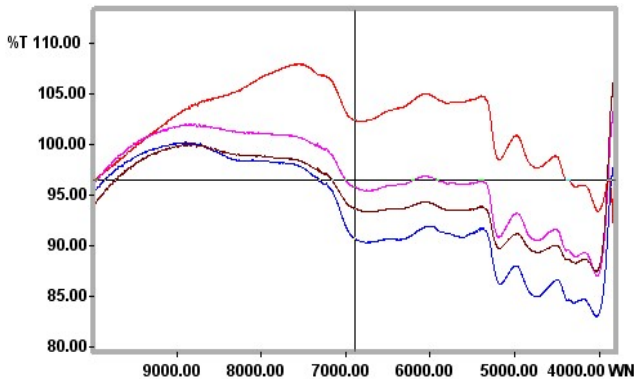
Fig. 5. Panicle blast contrast between three Panicle level

The transmission of spectra can be collected to the appropriate map data. Table 1 shows the spectral data comparison of leaves between serious leaf disease and disease-free. It can be seen that larger wave number, the two large difference appears transmission intensity, and where is the small wave numbers becomes smaller. The difference reaches the maximum value in the 7899WN from the difference of transmission intensity view. When we subtract the two spectra ,we find that it has the largest difference of 6.106 in the 7565WN. It also indicates that the largestly different wavelength must concern the wave number 7565.

4   Error Sources

The spectroscopy which is collected by near infrared spectrometer will have a slight error compared with the actual spectroscopy, mainly from the following:

1) The requirement which the moving mirror shifts from negative infinity to positive infinity leads to the ends of the interferogram data suddenly cutting off and then gets the incomplete data when the interference spectrum is acquired,due to the the actual moving distance of the moving mirror spectrometer.

2) when the interferogram is changed to the Spectrogram,the compensation of incomplete data (apodization) also make the peak distortion, resulting in error;

3) The transform used for a fast algorithm of discrete generalized fourier transform, will lead to the error between the data and continuous fourier transform.

5   Summary

This study conducted qualitative analysis of rice in cold land. there are four groups of experimental samples, namely, different parts in healthy plants and infected plants; three disease levels of infected leaves; five disease levels of infected grains; four disease levels of infected stems; The results showed that the plants have their own near-infrared spectral bands under different circumstances, the conclusion lays an early foundation for real-time detection of rice blast cold disease using the near infrared spectroscopy technology.

Table 1. The spectrum constract between healthy and diseased lamina

	Wave number	Transmittance	Wave number	Transmittance
Diseased leaves	9898	108.255	9403	111.691
	8898	114.439	8403	116.055
	7899	118.233	7404	118.661
Heathy leaves	9898	103.375	9403	106.256
	8898	108.708	8403	110.387
	7899	112.245	7404	113.045
	6900	110.271	6404	112.369

References

[1]   Zhang, G.-M., Aihua, X.: Study of Rice Blast Retrospect and Prospect. Agricultural Sciences (6), 156–158 (2008)
[2]   Wenbo, P.: Potential for rice production in Northeast China and its development strategy. Shenyang Agricultural University (2009)
[3]   Gold, S.: Occurrence of Rice Blast Causes and countermeasures. Crops Journal (1), 69–70 (2006)
[4]   Yuan, X., Wang, X.: Off into the macro. 2005 blast in Heilongjiang Province of the outbreak investigation and control. Modernization of Agriculture 9, 7–8 (2006)

[5] Wu, X., Xu, S., Jing, Z., et al.: Blast in the north and integrated control. RICE (1), 46–47 (2006)

[6] Adams, M.L., Philpot, W.D., Norvell, W.A., et al.: Yellowness index: an application of spectral second derivatives to estimate choruses of leaves in stressed vegetation. International Journal of Remote Sensing 20(18), 3663–3675 (1999)

[7] Bravo, C., Moshou, D.: Early disease detection in wheat fields using spectral reflectance. Biosy systems Engineering 84(22), 137–145 (2003)

[8] Yang, Z., Rao, M.N., Elliott, N.C., et al.: Using ground-based multispectral radiometry to detect stress in wheat caused by geenbug (Homoptera: Aphididae) infestation. Computers and electronics in agriculture 47, 121–135 (2005)

[9] Lin, Q.-Y., Xie, L.-H., Shi, X.L., et al.: Dwarf Disease of Rice cluster spectral properties of the virus VII. Plant Pathology 24(1), 5–9 (1994)

[10] Yaqiong, H., Yuqing, W., Junqiu, L., et al.: Cysteine substrate in the adsorption mechanism of silver Raman spectra. Light Scattering 14(3), 172–175 (2002)

[11] Cheng, H., Zhong, T., Yong, D., et al.: C60 derivative molecules Raman spectroscopy. Light Scattering 14(3), 127–130 (2002)

[12] Yu, Z., Ke, W.-Z., et al.: DNA molecules China and apyrimidinic fiber Raman spectroscopy. Spectrum Analysis Report 23(4), 734–738 (2003)

[13] Feng, M., Wang, Y.-F., Min, H.J., et al.: Raman spectroscopy of crystalline SiC crystals. Light Scattering 15(3), 158–161 (2003)

[14] Yi, K.P., Yu, L., Qiong, H., Liu, R., et al.: Xanthomonas oryzae races in the nano-silver membrane of surface-enhanced Raman spectroscopy. Spectroscopy and Spectral Analysis 30(2), 372–375 (2010)

[15] Di, W., Feng, L., et al.: Based on visible / near-infrared spectroscopy of the early detection of Botrytis cinerea eggplant leaves. Infrared and Millimeter Waves 26(4), 269–273 (2007)

Study on Dynamic Variation Regularities of Regional Groundwater Depth Based on Complexity Diagnosis

Dong Liu[*], Miao Yu, Nan Sun, and Ying Qi

School of Water Conservancy & Civil Engineering,
Northeast Agricultural University,
Harbin Heilongjiang 150030, China
liu72dong@126.com

Abstract. In the course of analyzing and forecasting regional groundwater depth, its variation process complexity was ignored in the past, so that it was difficult to realize scientific management of groundwater resources. Aiming at the aforesaid defect, taking Jiansanjiang branch bureau as research platform, this paper diagnosed the complexity of each monthly groundwater depth series through using complexity measurement methods which include wavelet theory, fractal theory, Approximate Entropy and so on. The comprehensive measurement analysis results show that the complexity order of monthly groundwater depth series in the eighth district of Farm Nongjiang is highest in four farms of northern region. On that basis, the most complexity groundwater depth series combination stochastic forecasting model based on wavelet transform in northern region of Jiansanjiang branch bureau was set up. The model forecasting results show that local annual mean groundwater depth will be in a state of fluctuant lifting, and have considerable gap in comparison to groundwater ideal depth in the future. So, local government should reinforce the scientific management of groundwater. The research achievements break through the traditional study paradigm of hydrological forecasting, provide a research mode for studying regional hydrologic process, and provide scientific evidence for sustainable utilizing regional groundwater resources.

Keywords: northern region of Jiansanjiang branch bureau, groundwater depth, complexity diagnosis, wavelet transform, combination stochastic forecasting model.

1 Introduction

Groundwater is an important basic strategy resource to support the economic sustainable development of the regional society. Groundwater resources is a complex system [1], containing system factor such as water level, water quality, water quantity and water temperature etc., among which, groundwater level is the most sensitive and active one. Under conditions of unreasonable exploitation of groundwater, excessive

---

[*] Corresponding author.

use of fertilizers and pesticides and some nature factors, fall of groundwater level, water quality deterioration and intensification of agricultural non-point source pollution etc. serious problem are induced, making the complexity of groundwater more and more clear which is general concerned by all circles. But the complexity of changing process was ignored by the domestic scholars when they studied the groundwater depth dynamic change regularity of the groundwater, resulting it imposable to fully excavate information of groundwater buried deeply change process and lack of practicality. Under this background, it is necessary to analyze predict the regional groundwater buried deeply series dynamic change to achieve the sustainable use of local groundwater, premised the complexity diagnoses.

Generally speaking, the higher complexity of the hydrological time series, the lower predictability of it[2]. So, to increase the prediction accuracy of the higher hydrological time series complexity, appropriate method should be chosen. Traditional prediction method of groundwater groundwater depth contains time series analysis[3], gray theory[4], artificial neural networks[5], multiple linear regression, seasonal decomposition model[6], self-memory model[7], etc., some of them base on linear theory, considered questions simply, some has ineffective extension, some has complex theories, these all resulting lower prediction accuracy or difficult to use. Wavelet analysis, developed recently, can demonstrate the fine structure series fully and extract change regularity easily of the hydrological time. The authors study the case of north district of Jiansanjiang branch bureau, to measure the complexity of the groundwater groundwater depth series to found the random predicting model of complex groundwater depth series of north district of Jiansanjiang branch bureau.

2   Research Area and Methods

2.1   Research Area

Heilongjiang Land Reclamation Jiansanjiang branch bureau is rice production base of China, locating at the junction of two cities and two countries of Fujin, Tongjiang, Raohe in Sanjiang Plain with the agricultural acreage of 682 thousand hm2[8] of total area of 12.3 thousand km2[9]. The north district of Jiansanjiang branch bureau includes farms of Qinglongshan, Qindeli, Nongjian and Yalu River with total land area of 2900.8km2, and in the year of 2009, the agricultural acreage is 147 hm2. The district has provincial nature reserve of Qindeli sturgeon fish, Qinglonghe and Yalu River branch level nature reserve, mainly planted crops of rice, maize, soybean and wheat with higher mechanization and commodity grain. From 1920s, the district tied rice planting, and the acreage increased year by year, by the year of 2009, it was 105 thousand, taking up 71.4% of the total agricultural acreage. Because of the lack of control projects, agriculture irrigation is mainly used groundwater, resulting continuous decline of the groundwater level. Meanwhile, under the influence of precipitation, evaporation and hydrological conditions, the complexity of the groundwater dynamic change become more and more clear of the district. Therefore, the authors try to

analysis and predict the complex groundwater depth series dynamic change of the north district of Jiansanjiang branch bureau under the premise of complexity diagnose, to realize the sustainable use of the local groundwater resource.

2.2   Data Sources

From Heilongjiang Land Reclamation Jiansanjiang branch bureau Water Authority collected the monthly groundwater level series monitoring data including 15 farms practice of Jiansanjiang branch bureau from 1997 to 2007(n=132).   Draw the groundwater level series changing curves of every areas in fig. 1.As can be seen from fig. 1, depth of the groundwater monitoring sites in Jiansanjiang branch bureau showed a gradual growth trend overall.
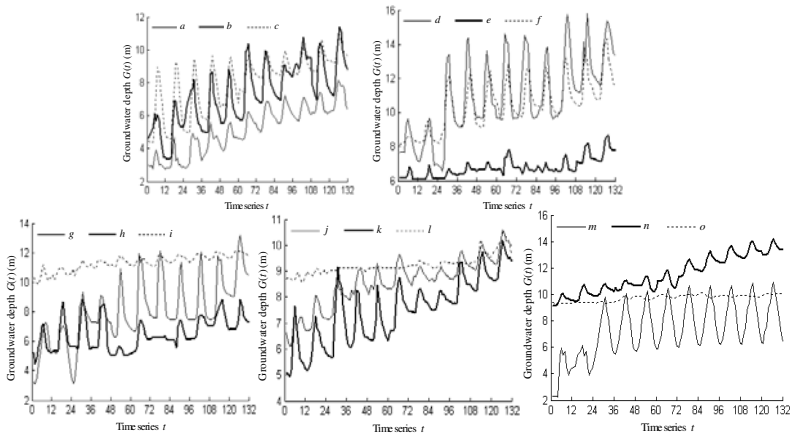
a. 1st team of Farm 859; b.69th team of Farm Qixing; c.22nd team of Farm Qianjin; d.24th team of Farm Chuanye; e.5th team of Farm Yalu River; f.28th team of Farm Hongwei; g.8th district of Farm Nongjiang; h.17th team of Farm Qinglongshan; i.36th team of Farm Qindeli; j.main field of Farm Qianfeng; k.6th district of Farm Honghe; l.5th district of Farm Erdaohe; m.11th team of Farm Daxing; n.31st team of Farm Shengli; o.12th team of Farm Qianshao

Fig. 1. Monthly groundwater level series changing curves of Jiansanjiang branch bureau (1997~2007)

2.3   Diagnose Measurements of Hydrological Series Complexity

Frequently used diagnose measurements of hydrological series contain information cost function based on wavelet transform(abbr. *WT-ICF*), approximate Entropy(abbr. *ApEn*), Lempel-Ziv Complexity(abbr. *LZC*), Resealed Range Analysis, (abbr. *R/S*), discrete wavelet transform(abbr. *DWT*), continue wavelet transform(abbr. *CWT*), etc., specific references as [10] and [11], not repeated here.

2.4   Modeling Principles of Hydrological Series Random Model Based on Wavelet Transform

2.4.1   Fast Wavelet Transform Method

Computational complexity is big for information redundancy[12] of the obtained wavelet transform coefficient when the hydrological time series are wavelet transformed. So, fast wavelet transform (abbr. *FWT*) is usually used in practical application to calculate the wavelet transform coefficient. The authors use *A Trous* method [10] which is simple, fast and small computational complexity.

To wavelet transform the hydrological series $f(t)$ ($t=1, 2, …, N$), letting $C^0(t) = f(t)$, the factorization and reconstruction process of *A Trous* method is as following:

$$\begin{cases} C^j(t) = \sum_{k=-\infty}^{+\infty} h(k)C^{j-1}(t+2^j k) \\ W^j(t) = C^{j-1}(t) - C^j(t) \end{cases} \tag{1}$$

$$C^0(t) = C^J(t) + \sum_{j=1}^{J} W^j(t) \tag{2}$$

Among which, $C^j(t)$, $W^j(t)$ are the scale coefficient (background signal) and wavelet coefficient (detail signal) under j scale; $h(k)$ is discrete lowpass filter, generally be chosen the compactly supported symmetric third-order *B*-spline one, as $h(k)=(1/16,1/4,3/8,1/4,1/16)$. $\{W^1(t), W^2(t), …, W^j(t), C^j(t)\}$ is called the wavelet transform series under *J* scale.

2.4.2   Modeling Procedures

Wavelet transform method (*e.g. A Trous* Method) is used to transform the hydrological time series $f(t)$ to obtain wavelet transform series $\{W^1(t), W^2(t), …, W^J(t), C^J(t)\}$ under some *J* scale firstly, then recognize the main components (random or certain components) of the wavelet transform series and found suitable mathematic models of each wavelet transform series, finally, to get the random predicting model[10] of the studying hydrological time series $f(t)$ by the method of wavelet transform and reconstruction.

3   Results and Analyses

3.1   Complexity Measures of Groundwater Level Series

Estimated the complexity of each monthly groundwater depth series through using six kinds of complexity measurement methods mentioned above in Jiansanjiang branch bureau. The complexity diagnosis results of northern region of Jiansanjiang branch bureau shown in table 1.

Table 1. Calculated Results of the complexity index about monthly groundwater depth series in the northern region of Jiansanjiang branch bureau

Long view well location	WT-ICF (0.16)		ApEn (0.16)		LZC (0.16)		D						comprehensive complexity index	complexity sort
							R/S Analysis (0.12)		Wavelet estimation					
									DWT (0.16)		CWT (0.24)			
	sort	score	sort	score	sort	score	sort	score	sort	score	sort	score		
17th team of Farm Qinglongshan	⑧	8	⑧	8	③	13	①	15	⑩	6	⑥	10	9.80	③
36th team of Farm Qindeli	⑤	11	⑩	6	②	14	⑨	7	⑬	3	⑮	1	6.52	⑪
8th district of Farm Nongjiang	⑦	9	⑫	4	⑦	9	⑤	11	②	14	①	15	10.68	②
5th team of Farm Yalu River	④	12	⑨	7	⑤	11	④	12	⑤	11	⑩	6	9.44	④

Note: ① The digital in parentheses is the weight of groundwater depth sequence complexity measure1;

②D is fractal dimension.

3.2   Synthetical Complexity Measures of Groundwater Level Series

In the above complexity measurement methods, Range Analysis is sensitive to the sequence length which is belongs to biased estimation and poor stability[13] ,while the method Continuous Wavelet Transform Fractal Theory has good stability[10] ,and others' be placed in the middle.In order to give full play to the advantages of various complexity measurement methods, according to the above analysis, determined the weight of that six complexity measures w_i ($i= 1,2,…, 6$) (shown in table 1). Assigned the corresponding score $s_i = 15 \sim 1$ to the sort result （①~⑮） of each monthly groundwater depth sequence complexity in Jiansanjiang branch bureau. Then get the groundwater depth sequence comprehensive complexity calculation formula:

$$C_{Ij} = \sum_{i=1}^{6} s_i w_i \tag{3}$$

C_{Ij} is the comprehensive complexity index of which in the sort of j in groundwater depth sequence , $j= 1, 2, …, 15$.

Calculate the comprehensive complexity index of groundwater depth sequence in northern region of Jiansanjiang branch bureau followed calculation formula (3) ,and the result seen in table 1.From table 1 we known that the synthetical complexity sort of

northern region of Jiansanjiang branch bureau monthly groundwater level series is the eighth district of Farm Nongjiang >the 17th team of Farm Qinglongshan>the fifth district of Farm Yalu River>the 36th team of Farm Qindeli>the 12th team of Farm Qianshao.

3.3 The Model of Complexity Groundwater Depth Series Combination Stochastic Forecasting Model Which Is Based on Wavelet Transform in Northern Region of Jiansanjiang Branch Bureau

By the calculation analysis shows that the complexity of the eight district of Farm Nongjiang is the highest in the sort of five farms in northern region.Therefore, chose the monthly groundwater depth series of the eighth district of Farm Nongjiang as a representative, describe northern region of Jiansanjiang branch bureau groundwater depth dynamic change general characteristics based on analyzes dynamic change regulation above.Set up groundwater depth series combination stochastic forecasting model based on wavelet transform according to the monthly groundwater depth data surveied in the eighth district of Farm Nongjiang from 1997 to 2006. Reserved the data of 2007 as verify results.

(1) Wavelet decomposition and reconstruction of monthly groundwater depth series

Adopted *A Trous* algorithm mentioned above and boundary extension method which include 35 percent symmetric extension and the same proportion of linear extension and 30 percent orthogonal polynomials fitting extension to decompose the monthly groundwater depth series $G(t)$ (t = 1 ~ 120) of the eighth district of Farm Nongjiang from 1997 to 2006 through comparative analysis. The scale number *p=2* here .Obtained wavelet decomposition sequence $\{W^1(t), W^2(t), C^2(t)\}$, shown in fig. 2 (a),(b),(c).Then we can got reconstruction sequences by superposed every wavelet decomposition sequence.Shown in fig. 2(d).As can be seen from fig. 2(d), reconstruction complete agreement with the monthly groundwater depth series change precess results in fig.1.So, used *ATrous* algorithm to decompose monthly groundwater depth data series in the eighth district of Farm Nongjiang is feasible.
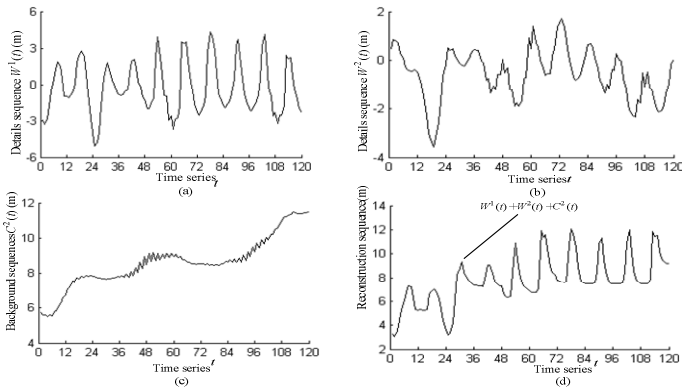
Fig. 2. Wavelet decomposition and reconstruction of monthly groundwater depth series in the eighth district of Farm Nongjiang

(2) Component identification about wavelet transform sequence

Draw the self-relevant chart and variance spectral density fig. [14] of the wavelet transform sequence. Seen fig. 3 and fig. 4.
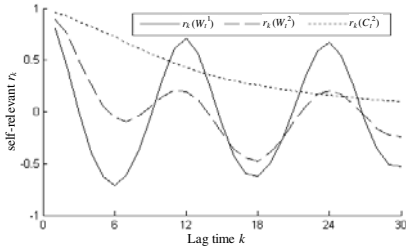
Fig. 3. The self-relevant charts of each wavelet transform series
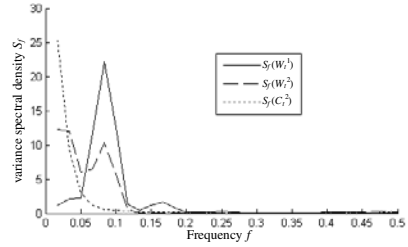
Fig. 4. The variance spectral density of each wavelet transform series

It can be seen from fig. 2, fig.3, fig. 4 that the wavelet decomposition sequence has a periodical for 12 months,and it has changing regularity powerful.So $W^1(t)$ which belongs to deterministic component can represent periodic term of $G(t)$ .Although the wavelet transform sequence $W^2(t)$ also has a certain of periodicity that inferior to $W^1(t)$, it mainly reflect randomness. So we used $W^2(t)$ of which belongs to random component stand for randomness of $G(t)$.The sequence $C^2(t)$ not has periodic changing regularity but clear trend.For this reason, $C^2(t)$ which belongs to deterministic component [14]can be approximated as trend term of $G(t)$.

(3) Cross-correlation analysis about wavelet transform sequence

It can be seen each cross-correlation analysis curve [14] about wavelet transform sequence from fig. 5 .In this fig. $r_k(W^1, W^2)$, $r_k(W1, C^2)$,and $r_k(W, {}^2C^2)$ showed the cross correlation coefficients among sequences $W^1(t)$, $W^2(t)$,and $C^2(t)$ respectively. In addition, as the complex characteristics of original sequence is obvious, data sequences $W^1(t)$ and $W^2(t)$ are non stationary sequence with periodic or approximate periodic components, both part of the information contained overlap lead to many points in the cross correlation coefficient $r_k(W^1, W^2)$ beyond 95% allows limited scope.So, Need to smooth with sequence $W^1(t)$ and $W^2(t)$.
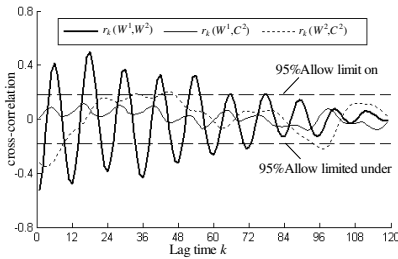
Fig. 5. The cross-correlation charts of each wavelet transform series
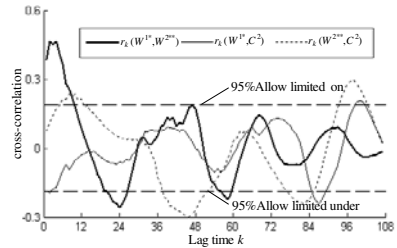
Fig. 6. The cross-correlation charts of each wavelet transform series after smooth processing

(4) Smoothly with the wavelet transform $W^1(t)$ and $W^2(t)$

Take seasonal difference [15] to the series $W^1(t)$ who has 12 months (1 year) cycle of change rule obviously, will get the wavelet transform difference sequence $W^{1*}(t)$.For the sequences $W^2(t)$ also has 12 months (1 year) cycle of change rule obviously,and has nonstationarity remarkable.So,deal it with logarithmic transformation[16] method first of all ,obtain the logarithmic transformation series $W^{2*}(t)$.Then, take seasonal difference to $W^{2*}(t)$ to get the wavelet logarithmic transformation difference sequences $W^{2**}(t)$.

(5)Cross-correlation analysis about wavelet transform sequence after smooth processing

Draw the cross-correlation curve about $W^{1*}(t)$, $W^{2**}(t)$ and $C^2(t)$ in fig. 6.It can be seen from the fig. that after smooth processing,the cross correlation coefficient are almost followed in 95% allows limited scope, and close to zero. Therefore, the mutual correlationof sequence is lesser.We can approximate think that they are independent of each other.The above cross-correlation analysis result showed that the components of $W^{1*}(t)$, $W^{2**}(t)$and $C^2(t)$ nearly single ,more simple than $G(t)$.At the same time series $W^{1*}(t)$and $W^{2**}(t)$ can be reduced to the wavelet sequences $W^1(t)$ and $W^2(t)$ after smooth processing.So analysis and deal with G(t) is means to manage $W^{1*}(t)$, $W^{2**}(t)$and $C^2(t)$

(6) Set up the combination stochastic model

From those analysis mentioned above, we known that sequence $W^{1*}(t)$ is got from periodic term $W^1(t)$ of original sequence after seasonal difference,and sequence $C^2(t)$ which belongs to deterministic component is approximate trend term of original sequence. At the same time ,sequence $W^{2**}(t)$ is got from stochastic $W^2(t)$ of original sequence after logarithmic transformation and seasonal difference,it belongs to stochastic component and it is approximate random component of original sequence. Therefore, set up auto-regression model[14] of $W^{1*}(t)$ and $C^2(t)$ (cut out random variables ε_t) and auto-regression model of $W^{2**}(t)$by using point of view of stochastic hydrology. Reduction and superposition each series' mathematical model then will got the groundwater depth series combination stochastic forecasting model which is based on wavelet transform in the eighth district of Farm Nongjiang

$$
\begin{cases}
\hat{W}^{1*}(t) = 0.7216W^{1*}_{t-1} - 0.2698W^{1*}_{t-12} \\
\hat{W}^1(t) = \hat{W}^{1*}(t) + W^1(t-12) \\
\hat{W}^{2**}(t) = 0.0595 + 0.8300(W^{2**}_{t-1} - 0.0595) + 0.6934(W^{2**}_{t-2} - 0.0595) \\
\qquad\qquad - 0.8232(W^{2**}_{t-3} - 0.0595) + 0.3840(W^{2**}_{t-5} - 0.0595) + \varepsilon_t \\
\hat{W}^{2*}(t) = \hat{W}^{2**}(t) + W^{2*}(t-12) \\
\hat{W}^2(t) = 2 - e^{\hat{W}^{2*}(t)} \\
\hat{C}^2(t) = 8.5940 + 0.8889(C^2_{t-1} - 8.5940) \\
\qquad\qquad + 0.2438(C^2_{t-2} - 8.5940) - 0.2247(C^2_{t-3} - 8.5940) \\
\hat{G}(t) = \hat{W}^1(t) + \hat{W}^2(t) + \hat{C}^2(t)
\end{cases}
\tag{4}
$$

Type with "^" said the corresponding sequence of predictive value

(7) Model fitting

Adopted the most complexity groundwater depth series combination stochastic forecasting model which is based on wavelet transform to fit the monthly groundwater depth series in eighth district of Farm Nongjiang from 1997 to 2006. Shown in fig. 7. As can be seen from fig. 7, the model fitting effect is good.
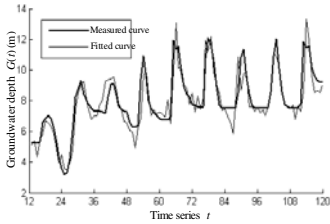
Fig. 7. The fitting curve of monthly groundwater depth series combination stochastic forecasting model in the eighth district of Farm Nongjiang (1997~2006)
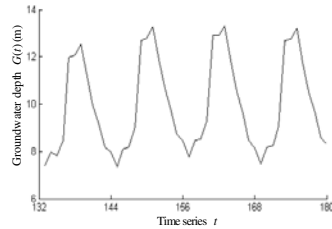
Fig. 8. The forecasting curve of monthly groundwater depth series combination stochastic forecasting model in the eighth district of Farm Nongjian (2008~2011)

(8) Model Accuracy Test

Used the data of most complexity groundwater depth series combination stochastic forecasting model which is based on wavelet transform to test the fitting effect, and used the left monthly modeling groundwater depth measurement data of 2007 which is not adopted in the model to test trial reported result. The specific results in Table 2.

Table 2. Accuracy test result of combination stochastic forecasting model based on wavelet transform

Fitting Effect index				Simulated forecast index
C	p	E1 （％）	E2	E3 （％）
0.3208	0.9444	14.13	0.9259	83.33

According to relevant literature we can see that posterior error ratio C, fitting accuracy rate E2 and trial reported effect indexes E3 were reached I level standard , minimum error frequency p is II level, and the relative mean square error E1 reached III level standard [17]. Therefore, the model based on wavelet transform mentioned above have good fitting effect and highly forecast precision, can be used to predict future groundwater depth in the eighth district of Farm Nongjiang.

(9) Groundwater depth prediction

Now adopted the most complexity groundwater depth series combination stochastic forecasting model which is based on wavelet transform to predict monthly groundwater level series of the eighth district of Farm Nongjiang from 2008 to 2011.The prediction data and prediction curve of groundwater table can be seen respectively in table 3 and fig. 8.From table 3 and fig. 8 shown that the groundwater level of the eighth district of Farm Nongjiang will fluctuate lift, if still using the past groundwater exploitation mode. The inter-annual luffing of groundwater level from 2008 to 2011 were -1.28 m, 0.52 m, 0.08 m and - 0.12 m respectively in the fluctuation change down-up-up-down state. And the groundwater level maintained at about 10m.This result have considerable gap in comparison to groundwater ideal depth which is about 4m to 6m in the future [18]. Therefore, the eighth district of Farm Nongjiang should increase groundwater management efforts, to protect local limited of groundwater resources.

Table 3. The prediction data of monthly groundwater depth in the eighth district of Farm Nongjiang（2008~2011）(m)

Year \ Month	1	2	3	4	5	6	7	8	9	10	11	12	Average
2008	7.4118	7.9645	7.8243	8.4863	11.9801	12.0779	12.5336	11.1655	9.9586	9.1503	8.1838	7.9454	9.56
2009	7.3383	8.0817	8.1816	9.0356	12.6974	12.7893	13.2663	11.8944	10.6546	9.8109	8.7365	8.4515	10.08
2010	7.7792	8.4760	8.5244	9.3032	12.9034	12.9148	13.3161	11.8589	10.5339	9.6153	8.4798	8.1595	10.16
2011	7.4709	8.1710	8.2395	9.0524	12.6953	12.7523	13.2019	11.7968	10.5272	9.6653	8.5838	8.3112	10.04

4   Conclusions

It is more obvious, driven in a variety of factors, the complexity of characteristics during the course of the regional groundwater table changed has became. So, it has broken the traditional hydrological forecasting patterns and has more practical significance for the regional groundwater resource management that analysis of groundwater depth development trend based on identify regional groundwater depth sequence complex. This paper diagnosed the complexity of 15 monthly groundwater level series through many kinds of data mining methods in Heilongjiang Agricultural Reclamation Jiansanjiang branch bureau. The complexity diagnosis results show that the complexity order of monthly groundwater depth series in the eighth district of Farm Nongjiang is highest in five farms of northern region. On that basis, the most complexity groundwater depth series combination stochastic forecasting model based on wavelet transform in northern region of Jiansanjiang branch bureau in the eighth district of Farm Nongjiang was set up. This model concepts are clear, structured and high reliability. The model forecasting results show that local annual mean groundwater depth will be keep 10 m, and have considerable gap in comparison to groundwater ideal depth in the future. So, local government should reinforce the scientific management of groundwater. Need to point out is that The stability of the

complexity measure results by certain effect as the groundwater depth monitoring data insufficient length in Jiansanjiang branch bureau. In the future, with the monitoring data accumulation, we should update this paper research.

References

[1] Wang, G., Liang, G.: Application of Multi-period, Multi-level and Multi-objective Fuzzy Optimization Model in Groundwater System Management Decision-making. Journal of China Hydrological 29(6), 24–27 (2009)

[2] Peng, T., Chen, X., Zhuang, C.: Analysis on complexity of monthly runoff series based on sample entropy in the Dongjiang river. Ecology and Environmental Science 18(4), 1379–1382 (2009)

[3] Liu, D., Ma, Y.-S., Fu, Q.: Study on Dynamic Variation Regularities of Groundwater in Area of Well Irrigation Paddy in Sanjiang Plain. Journal of Irrigation and Drainage 25(5), 42–46 (2006)

[4] Zhang, W.-T., Sun, J.-Y., Wu, C.-Y.: Research on the application of groundwater dynamic change regulation based on the gray-theory in pingdu. Shandong Water Resources (3-4), 45–48 (2009)

[5] Banerjee, P., Prasad, R.K., Singh, V.S.: Forecasting of groundwater level in hard rock region using artificial neural network. Environmental Geolog 58(6), 1239–1246 (2009)

[6] Xu, Q., Shu, L., Yang, D., et al.: Model on Groundwater Level Dynamic Statistical in Pinggu Plain of Beijing. Water Resources and Power 27(5), 58–61 (2009)

[7] Li, R., Shen, B., Zhang, J.: Self-memory model for predicting groundwater depth series with periodical fluctuation. Transaction of The Chinese Society of Agricultural Engineering 21(7), 34–37 (2005)

[8] Zhao, Q.: Research on the Trend of Underground Water Change in Jiansanjiang Area Based on Gray Prediction. Journal of Water Resources and Water Engineering 20(5), 128–130, 134 (2009)

[9] Guo, L., Ma, K.-M., Zhang, Y.: Landscape assessment on wetland degradation during thirty years in Jiansanjiang region of Sanjiang Plain, Northeast China. Acta Ecologica Sinica 29(6), 3126–3135 (2009)

[10] Wang, W.-S., Ding, J., Li, Y.-Q.: Hydrological wavelet analysis. Chemical Industry Press, Beijing (2005)

[11] Tong, C.-S., Huang, Q., Liu, H., et al.: Study on runoff series complexity based on approximate entropy. Journal of Northwest Sci-Tech University of Agriculture and Forestry 33(6), 121–126 (2005)

[12] Ding, S.-Y.: Laplacian Pyramid image data compression based on wavelet transformation. Journal of Weifang University 9(4), 34–36 (2009)

[13] Rakhshandehroo, G.R., Shaghaghian, M.R., Keshavarzi, A.R., et al.: Temporal variation of velocity components in a turbulent open channel flow: Identification of fractal dimensions. Applied Mathematical Modelling 33(10), 3815–3824 (2009)
[14] Ding, J., Liu, Q.-S.: Stochastic hydrology. China Water Power Press, Beijing (1997)
[15] Yu, Y.-H., Wang, J., Song, J.-D.: Research on traffic prediction and result analysis of using multiplicative seasonal autoregressive integrated moving average. Computer Engineering and Application 45(20), 99–102 (2009)
[16] Chen, D.-G., Hu, N.-L., Li, G.-Q.: Stationarity for the non-normal distribution of regionalized variables. Journal of University of Science and Technology Beijing 31(4), 412–417 (2009)
[17] Chen, N.-X.: Evaluating the Precision of Forecasting Model for Groundwater Dynamic. Geotichnical Investigation and Surveying 3, 35–38 (1999)
[18] Yuan, B., Du, S.-M., Yang, W.-Q., et al.: Status Overdraft of Groundwater on Sanjiang Plain(2). Journal of Heilongjiang Hydraulic Engineering College 35(1), 18–21 (2008)

Comparison of Spectral Indices and Principal Component Analysis for Differentiating Lodged Rice Crop from Normal Ones

Zhanyu Liu[1,2,*], Cunjun Li[3], Yitao Wang[4], Wenjiang Huang[3], Xiaodong Ding[1], Bin Zhou[1], Hongfeng Wu[4], Dacheng Wang[3], and Jingjing Shi[5]

[1] Institute of Remote Sensing and Earth Sciences,
Hangzhou Normal University, Hangzhou 311121, China
[2] Key Laboratory of Urban Wetland and Region Change in Zhejiang Province,
Hangzhou 311121, China
[3] Beijing Research Center for Information Technology in Agriculture, Beijing 100097, China
[4] Institute of Scientific and Technological Informatics,
Heilongjiang Academy of Land Reclamation Sciences, Harbin 150036, China
[5] Institute of Agricultural Remote Sensing & Information Technology,
Zhejiang University, Hangzhou 310029, China
liuzhanyu@zju.edu.cn

Abstract. Hyperspectral reflectance of normal and lodged rice caused by rice brown planthopper and rice panicle blast was measured at the canopy level. Over one decade broad- and narrow-band vegetation indices (VIs) were calculated to simulate Landsat ETM+ with *in situ* hyperspectral reflectance. Principal component analysis (PCA) was utilized to obtain the front two principal components (PCs). Probabilistic neural network (PNN) was employed to classify the lodged and normal rice with VIs and PCs as the input vectors. PCs had 100% of overall accuracy and 1 of Kappa coefficient for the training dataset. While PCs had the greatest average overall accuracy (97.8%) and Kappa coefficient (0.955) for the two testing datasets than VIs consisting of broad- and narrow-bands. The results indicated that hyperspectral remote sensing with PCA and artificial neural networks could potentially be applied to discriminate lodged crops from normal ones at regional and large spatial scales.

Keywords: Hyperspectral remote sensing, Lodged rice, Principal component analysis (PCA), Vegetation indices (VIs), Artificial neural networks (ANN).

1   Introduction

Detecting plant health condition plays an important role in controlling disease and insect stress in precise pest management (PPM), which always results in yield loss and poor quality [1]. Although plant pest stress is predominantly concentrated in patches around stress centers, it is still widespread to spray agrochemicals indiscriminately over the entire field in practice [2]. It is necessary to accurately assess the pest stress distribution and damage caused by disease and insect pests [3]

---

* Corresponding author.

so that agrochemicals can be applied to reduce environmental pollution due to agrochemicals in the stressed patches.

Lodging at the late growth stage is caused by insect and fungal disease in paddy rice field, which often leads to yield loss and quality decrease [4]. Accurate distribution information of lodging rice is helpful to spray agrochemicals and harvest

Table 1. Broadband and narrowband spectral indices used in this study

Abbreviation	Formula	Reference
DVI	$NIR-R$	Richardson and Everitt [11]
RVI	NIR/R	Jordan[12]
NDVI	$(NIR-R)/(NIR+R)$	Rouse et al. [13]
SAVI$_{L=0.5}$	$1.5(NIR-R)/(NIR+R+0.5)$	Huete[14]
OSAVI	$1.16(NIR-R)/(NIR+R+0.16)$	Rondeaux et al. [15]
MSAVI$_2$	$\dfrac{2NIR+1-\sqrt{(2NIR+1)^2-8(NIR-R)}}{2}$	Qi et al. [16]
TVI	$60(NIR-G)-100(R-G)$	Broge et al. [17]
IPVI	$NIR(NIR+R)$	Crippen [18]
TDVI	$\dfrac{1.5(NIR-R)}{\sqrt{NIR+R+0.5}}$	Bannari et al. [19]
RDVI	$\dfrac{NIR-R}{\sqrt{NIR+R}}$	Roujean and Breon [20]
EVI	$2.5\dfrac{NIR-R}{NIR+6.5R-7.5B+1}$	Liu and Huete [21]
GEMI	$\eta(1-0.25\eta)-(R-0.125)/(1-R)$ $\eta=\dfrac{2(NIR^2-R^2)+1.5NIR-0.5R}{NIR+R+0.5}$	Pinty and Verstraete[22]
BI	0.2909B+0.2493G+0.4806R+0.5568NIR +0.4438$SWIR_1$+0.1706$SWIR_2$	Crist et al. [23]
GVI	-0.2728B-0.2174G- 0.5508R+0.7721NIR+0.0733$SWIR_1$ -0.1648 $SWIR_2$	Crist et al. [23]
WI	0.1446B+0.1761G+0.3322R+0.3396NIR- 0.621$SWIR_1$ -0.4186 $SWIR_2$	Crist et al. [23]
NDWI	$(\lambda_{860nm}-\lambda_{1240nm})/(\lambda_{860nm}+\lambda_{1240nm})$	Gao [24]
LSWI	$(\lambda_{860nm}-\lambda_{1640nm})/(\lambda_{860nm}+\lambda_{1640nm})$	Xiao et al. [25]
NMDI	$\dfrac{\lambda_{860nm}-(\lambda_{1640nm}-\lambda_{2130nm})}{\lambda_{860nm}+(\lambda_{1640nm}-\lambda_{2130nm})}$	Wang et al. [26]

*$B,G,R,NIR,SWIR_1$ and $SWIR_2$ represent surface spectral reflectance averaged over ranges of wavelengths in Landsat ETM+, viz.TM1 (450-515nm), TM2 (525-605nm), TM3 (630-690nm), TM4 (775-900nm), TM5 (1.55-1.75μm) and TM7 (2.09-2.38μm), respectively. $\lambda_{i\,nm}$ denotes surface spectral reflectance at i nanometer.

in agricultural field management. Remote sensing has been proved to be more effective than ground surveys for detecting plant health condition in numerous studies [5]. Perhaps vegetation indices (VIs) is the most popular technique of remote sensing to predict agricultural crop biophysical variables such as leaf area index, chlorophyll content/concentration, above-ground biomass, and percent vegetation cover and so forth [6]. In addition to VIs, principal component analysis (PCA) has been extensively studied as a data compression technique [7].

Furthermore, the application of artificial intelligent methods including artificial neural networks (ANN) and support vector machines (SVM) in plant stress detection has also been reported recently[8-9]. For example, SVM technology was used to detect weed and nitrogen stress in corn by Karimi *et al* [8]. Shi *et al*. [9] applied SVM model to discriminate the health status in rice and great classification was obtained. However, few studies have been conducted to resolve classification problems by VIs and PCA [10].

The main goal of this research is to evaluate and compare the performances of broad- and narrow-band VIs (Table 1) and PCA in differentiating the normal rice crop from the lodged ones stressed by rice brown planthopper or rice panicle blast.

2   Methodology

2.1   Study Sites and Materials

There were two sampling sites in this research. One site was located in Friendship Farm (120°43′E, 46°39′N), Heilongjiang Province. The rice crop cultivar was Suijing 4, which was naturally infected with rice panicle blast. The other site was in Jiuxian Village (119°37′E, 29°48′N), Tonglu County, Zhejiang Province. And the rice crop cultivar was Hybrid rice 718, which was stressed by rice brown planthopper.

2.2   Hyperspectral Measurement

Canopy hyperspectral reflectance were measured using an ASD FieldSpec Pro FRTM Spectroradiometer (Analytical Spectral Devices Inc., Boulder, CO, USA) throughout the whole spectral range from 350 nm to 2 500 nm. The spectral resolution of the instrument varied from 3 nm (<1 000 nm) to 10 nm (>1 000 nm), while the spectra were interpolated by the spectrometer software in 1nm interval. Therefore, each measurement generated a spectrum ranging between 350 nm and 2500 nm at 1nm increment.

The fiber optic sensor with a 25° instantaneous field of view (IFOV) was pointed to the paddy rice canopy about 1 m height at the nadir position. Reference panel measurements were made at the beginning of each set of canopy measurements. The reference panel painted by $BaSO_4$ is a Lambert surface with a reflectance of no less than 99%. These radiance units were in turn converted to reflectance using scans of the $BaSO_4$ white reference panel.

The hyperspectral measurement date was taken on 24th August and 28th September of 2007 in Heilongjiang and Zhejiang Province, respectively. The number of reflectance spectra of normal paddy rice were 12 and 35 in the two sites, respectively, and the number of reflectance spectra of lodged paddy rice were 10 and 35 in the two sites, respectively.

2.3   Hyperspectral Data Preprocessing

Hyperspectral reflectance was smoothed with a five-step moving average to suppress instrumental and environmental noise [27]. The hyperspectral reflectance represented the domains of 400 ~ 2 350 nm, and the missing segments corresponding to strong instrument and environment noise were not considered for further analysis [10].

The total datasets collected from the two paddy rice fields were divided into three different subsets, namely, one training dataset and two testing subsets. The training dataset (n=52) was used to calibrate the artificial neural network (ANN) classification models with three quarters of the canopy hyperspectral reflectance data collecting on 28th September of 2007, and one quarter was applied to validate the ANN classification models as the first testing dataset (n=18). While the hyperspectral reflectance data collecting on 24th August of 2007 was applied as the second testing dataset (n=22).

2.4   Analytical Techniques

Principal component analysis (PCA) has been used as a data compression technique for preserving total variance in transformation and minimizing mean square approximate errors [7]. This technique is very suitable for hyperspectra data for high dependence and autocorrelation in adjacent wavebands [10]. In this study, the front two principal component spectra (PCs) were derived from the canopy hyperspectral reflectance, which could explain over 98% of information variation.

Probabilistic neural networks (PNN) are a novel and composite neural network with radial basis function and competitive neural network. The neural architecture of PNN (Fig.1) consists of a radial basis layer and competitive layer with the exception of input layer [28]. In this study, PNN has been used to identify the lodged rice crop from normal ones at the canopy level. The neural nodes in the input layer were the broad- and narrow-band VIs and PCs. The competitive layer is also called as the
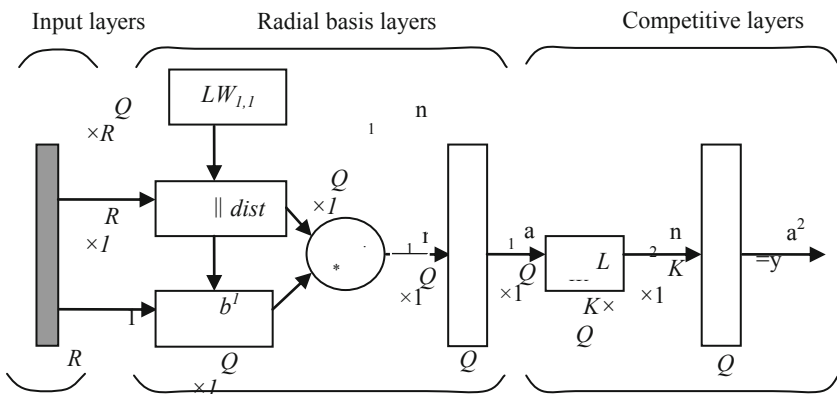
Fig. 1. Neural architecture for probabilistic neural networks (PNN)

output layer in practice. The neural nodes in the output layer were the classification of rice crops, viz. lodged and normal. The objective of the training process is examined to find the probabilistic mode between neural node of input layer and different classification of output layer. Additional detailed information on PNN can be found in the literature by Sivandm *et al.* [28].

The performance of VIs and PCA for classifying the lodged and normal paddy rice was evaluated with two statistical parameters, namely overall accuracy and Kappa coefficient [29].

PNN and PCA were conducted with Matlab functions in Matlab 7.0 (R14). Hyperspectral reflectance data preprocessing was made in ViewSpec Pro (V5.6.10) and VIs were derived from canopy hyperspectral reflectance in Microsoft Office Excel 2003.

3   Results and Discussions

3.1   Canopy Hyperspectral Reflectance

The spectral response properties of rice crop to disease and insect stresses were very significant for identifying lodged rice from normal ones in PPM using hyperspectral remote sensing technique. Fig.2 showed the canopy hyperspectral reflectance curves of normal rice and lodged rice caused by rice brown planthopper. Six wavelength intervals indicated the spectral differences (Table 2) because the photon transferring process was different between normal and lodged rice crop. Maybe it's because after the rice plant was damaged and lodged due to disease and insect stresses, which resulted in evident changes in the arrangement structure of rice panicles, leaves and stems.
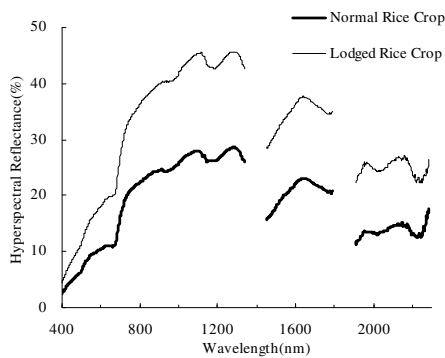
Fig. 2. Canopy hyperspectral reflectance of normal rice crop and lodged rice crop stressed by rice brown planthopper

As shown in Fig.2, the spectral curve shape of lodged rice crop was similar to that of the normal ones through the entire range (400~2350 nm), but the spectra amplitude was different. The hyperspectral reflectance of lodged rice crops increased about 75.7%, 70.9%, 79.2%, 62.3%, 64.7% and 76% in the blue-green (450~515 nm), green (525~605 nm), red (630~690 nm), near-infrared (775~900 nm), shortwave infrared (1550~1750 nm and 2090~2350 nm), respectively (Table 2).

Table 2. Average canopy hyperspectral reflectance of rice crops at six wavelength intervals representing the Enhanced Thematic Mapper Plus (ETM+) of Landsat-7 (Unit: %)

Spectrum range (nm)	Site 1 (rice brown planthopper)		Site 2 (rice panicle blast)	
	Healthy (n=35)	Lodged (n=35)	Healthy (n=12)	Lodged (n=35)
Blue-Green (450~515)	5.9c	10.4a	3.3d	7.6b
Green (525~605)	9.6d	16.4b	10.9c	17.2a
Red (630~690)	11.4b	20.5a	8.2c	18.1a
NIR (775~900)	23.3c	37.7b	45.7a	46.7a
SWIR$_1$ (1550~1750)	22.0b	36.2a	20.5c	41.6a
SWIR$_2$ (2090~2350)	14.2c	25.0b	13.4c	35.3a

*: Reflectance values within a row followed by the same letter are not significantly different from each other by Duncan test at $\alpha = 0.05$.

3.2   Results of Probabilistic Neural Network (PNN) Classifier

The overall accuracy and Kappa coefficient of broad-band and narrow-band spectral indices and PCA for discriminating lodged rice from normal ones based on PNN were

Table 3. Comparison of the two accuracy measures for VIs and PCA based on PNN (Unit: %)

Dataset types	Training dataset		Testing dataset 1		Testing dataset 2	
Input vectors	OA	KC	OA	KC	OA	KC
DVI	96.2%	0.923	88.9%	0.778	90.9%	0.814
RVI	100%	1	61.1%	0.222	63.6%	0.214
NDVI	88.5%	0.769	72.2%	0.444	63.6%	0.214
SAVI$_{L=0.5}$	94.2%	0.885	55.6%	0.111	68.2%	0.319
OSAVI	94.2%	0.885	66.7%	0.333	72.7%	0.421
MSAVI$_2$	96.2%	0.923	61.1%	0.222	63.6%	0.214
TVI	90.4%	0.808	66.7%	0.333	63.6%	0.214
IPVI	94.2%	0.885	66.7%	0.333	63.6%	0.214
TDVI	100%	1	77.8%	0.556	63.6%	0.214
RDVI	94.2%	0.885	66.7%	0.333	59.1%	0.108
EVI	94.2%	0.885	66.7%	0.333	72.7%	0.421
GEMI	94.2%	0.885	44.4%	-0.111	72.7%	0.421
BI	100%	1	100%	1	77.3%	0.56
GVI	86.5%	0.731	88.9%	0.778	81.8%	0.621
WI	94.2%	0.885	88.9%	0.778	81.8%	0.621
NDWI	92.3%	0.846	44.4%	-0.111	63.6%	0.214
LSWI	94.2%	0.885	33.3%	-0.333	68.2%	0.319
NMDI	84.6%	0.692	72.2%	0.444	27.3%	-0.467
2 PCs	100%	1	100%	1	95.5%	0.909

* OA and KC denoted the overall accuracy and Kappa coefficient, respectively.

shown in Table 3. PCA had the highest average overall accuracy (97.8%) and Kappa coefficient (95.5%) than the broad-band and narrow-band spectral indices for the two testing datasets, DVI and BI followed. Among the spectral indices, DVI had relative higher average overall accuracy (90.9%) and Kappa coefficient (0.796). Maybe it was because that DVI was more sensitive to changes of photosynthesis pigment (i.e. chlorophyll and carotenoid) of rice organs due to lodging caused by rice brown planthopper or rice panicle blast at the late growth stage.

4   Conclusions

The results of our study showed that the spectral indices based on distance from isoline to soil line such as DVI, BI, GVI and WI had higher overall accuracy and Kappa coefficient than those spectral indices based on ratios of spectral bands such as RVI, NDVI, SAVI series, and so on. Ratio-based spectral indices suppressed the spectral difference between different spectral bands, which made it more difficult to discriminate lodged rice from normal ones than distance-based spectral indices that amplified the spectral difference between different spectral bands [30]. The front several principal components (PCs) always accounted for most proportion of the variance of the original hyperspectral reflectance dataset [10], and retained more information than the broad-band and narrow-band spectral indices, and then PCs had the higher overall accuracy and Kappa coefficient than spectral indices.

The results indicated that hyperspectral remote sensing with PCA and ANNs has considerable potential in discriminating the lodging rice from normal ones as a supplementary and even alternative technique against spectral indices at the regional and even large spatial scales. However, there was only two kinds of stress status (i.e., healthy and lodging) in our research. Different lodging angles bring about different damage severity in practice. Future study is needed to include more stress status and extrapolate from ground to airborne and spaceborne platforms.

References

1. Everitt, J.H., Escobar, D.E., Summary, K.R., Davis, M.R.: Using airborne video, global positioning system, and geographical information system technologies for detecting and mapping citrus blackfly infestations. Southwest. Entomol. 19(2), 129–138 (1994)
2. McCartney, H.A., Fitt, B.D.L.: Dispersal of foliar fungal plant pathogens: mechanisms, gradients and spatial patterns. In: Gareth Jones, D. (ed.) Plant Disease Epidemiology, pp. 138–160. Kluwer Publishers, London (1998)

3. Pedigo, L.P.: Closing the gap between IPM theory and practice. J. Agri. Entomol. 12, 171–181 (1995)
4. Sōgawa, K.: The rice brown planthopper: feeding physiology and host plant interactions. Ann. Rev. Entomol. 27, 49–73 (1982)
5. West, J.S., Bravo, C., Oberit, R., Lemaire, D., Moshou, D., McCartney, H.A.: The potential of optical canopy measurement for targeted control of field crop diseases. Ann. Rev. Phytopathol. 41, 593–661 (2003)
6. Elvidge, C.D., Chen, Z.K.: Comparison of broad-band and narrow-band red and near-infrared vegetation indices. Remot. Sens. Environ. 54(1), 38–48 (1995)
7. Holden, H., LeDrew, E.: Spectral discrimination of healthy and non-healthy corals based on cluster analysis, principal component analysis, and derivative spectroscopy. Remot. Sens. Environ. 65(2), 217–224 (1998)
8. Karimi, Y., Prasher, S.O., Patel, R.M., Kim, S.H.: Application of support vector machine technology for weed and nitrogen stress detection in corn. Comput. Electron. Agr. 51(1-2), 99–109 (2006)
9. Shi, J.J., Liu, Z.Y., Zhang, L.L., Zhou, W., Huang, J.F.: Hyperspectral recognition of rice damaged by rice leaf roller based on support vector machine. Ric. Sci. 23(3), 331–334 (2009)
10. Liu, Z.Y., Wang, D.C., Li, B., Huang, J.F.: Discrimination of lodged rice based on visible/near-infrared spectroscopy. J. Infra. Milli. Wav. 28(5), 321–324 (2009)
11. Richardson, A.J., Everitt, J.H.: Using spectral vegetation indices to estimate rangeland productivity. Geocarto Int. 1, 63–77 (1992)
12. Jordan, C.F.: Derivation of leaf area index from quality of light on the forest floor. Ecol. 50(4), 663–666 (1969)
13. Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W.: Monitoring vegetation systems in the great plains with ERTS. In: 3rd ERTS Symposium, NASA, Washington, USA, pp. 48–62 (1973)
14. Huete, A.R.: A soil-adjusted vegetation index (SAVI). Remot. Sens. Environ. 25(3), 295–309 (1988)
15. Rondeaux, G., Steven, M., Baret, F.: Optimization of soil-adjusted vegetation indices. Remot. Sens. Environ. 55(2), 95–107 (1996)
16. Qi, J., Chehbouni, A., Huete, A., Kerr, Y., Sorooshian, S.: A modified soil-adjusted vegetation index (MSAVI). Remot. Sens. Environ. 48(2), 119–126 (1994)
17. Broge, N.H., Leblanc, E.: Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. Remot. Sens. Environ. 76(2), 156–172 (2000)
18. Crippen, R.E.: Calculating the Vegetation Index Faster. Remot. Sens. Environ. 34(1), 71–73 (1990)
19. Bannari, A., Asalhi, H., Teillet, P.M.: Transformed difference vegetation index (TDVI) for vegetation cover mapping. In: Proceedings on CD-Rom, paper I2A35, International Geoscience and Remote Sensing Symposium, Toronto, Ontario, Canada, p. 1508 (2002)
20. Roujean, J.L., Boreon, F.M.: Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. Remot. Sens. Environ. 51(3), 375–384 (1995)
21. Liu, H.Q., Huete, A.R.: A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. IEEE Transaction on Geosci. Remot. Sens. 33, 457–465 (1995)
22. Pinty, B., Verstraete, M.M.: GEMI: A non-linear index to monitor global vegetation from satellites. Vegetatio. 101(1), 15–20 (1992)

23. Crist, E.P., Laurin, R., Cicone, R.C.: Vegetation and soils information contained in transformed thematic mapper data. In: International Geoscience and Remote Sensing Symposium (IGARSS 1986), Zurich, Switzerland, pp. 1465–1470 (1986)
24. Gao, B.C.: NDWI: A normalized difference water index for remote sensing of vegetation liquid water from space. Remot. Sens. Environ. 58(3), 257–266 (1996)
25. Xiao, X., Boles, S., Frolking, S., Salas, W., Moore III, B., Li, C.: Observation of fl ooding and rice transplanting of paddy rice fields at the site to landscape scales in China using VEGETATION sensor data. Int. J. Remot. Sens. 23(15), 3009–3022 (2002)
26. Wang, L., Qu, J.: NMDI: A Normalized Multi-band Drought Index for Monitoring Soiland Vegetation Moisture with Satellite Remote Sensing. Geophy. Research Lett. 34, 204–208 (2007)
27. Kobayashi, T., Kanda, E., Kitada, K., Ishiguro, K., Torigoe, Y.: Detection of rice panicle blast with multispectral radiometer and the potential of using airborne multispectral scanners. Phytopathol. 91(3), 316–323 (2001)
28. Sivandm, S.N., Sumathi, S., Deepa, S.N.: Introduction to neural networks using Matlab 6.0. Tata Mcgraw-Hill Publishing Company Limited (2006)
29. Congalton, R.G.: A review of assessing the accuracy of classification of remotely sensed data. Remot. Sens. Environ. 37(1), 35–46 (1991)
30. Asrar, G.: Theory and applications of optical remote sensing, pp. 119–125. John Wiley & Sons, Inc., New York (1989)

Sustainable Fertilizer Level for Winter Wheat in Different Rainfall Regions on the Loess Plateau of China

Xuechun Wang[1], Shishun Tao[1], Mingde Hao[2], and Wei Li[3]

[1] School of Life Science and Technology,
Southwest University of Science and Technology,
Mianyang, Sichuan 712100, China
[2] Institute of Soil and Water Conservation, CAS & MWR,
Yangling, Shaanxi 712100, China
[3] Fujian Academy of Agricultural Sciences Central Laboratory Fuzhou,
Fujian 350003, China

Abstract. Higher fertilization on winter wheat increased the fluctuation of winter wheat yield in different rainfall years and impacted the sustainable development of winter wheat production on the Loess Plateau. Based on the long term field experimental data at Chagnwu Agricultural Station, this paper evaluated the EPIC model. And this paper also suggested a sustainable fertilizer level for winter wheat, based on the analysis of simulation results in different rainfall regions. Results of this study indicated that: 1) The EPIC model simulated both winter wheat yields and soil water among different fertilizer levels well, with the mean R value of 0.91 and 0.89 respectively. 2) With the increasing of fertilizer, the value of IRFG (Increase Rate of Grain yield by Fertilizer) and WUEG (Water Using Rfficiency for Grain yield) became higher, when soil water in deep soil was not be used excessively; however, the value of IRFG became lower, when soil water in deep soil was used excessively. 3) In the semi-humid region, fertilizer for winter wheat should be from N_4 to N_5; in the semi-humid and drought-prone region and in the semi-arid region, it should be from N_3 to N_4; in the semi-arid and drought-prone region, it should be lower than N_3.

Keywords: The loess plateau, Winter wheat, Fertilizer, EPIC model.

0   Introduction

Winter wheat is a major food and feed grain crop in the world. It occupies a large area (82%) of the Loess Plateau rain-fed region of china [1]. Wheat is mostly grown under dry land conditions, therefore, its growth, development and yield depended mainly on available water and fertilizer. An increase in fertilization can stimulate deeper rooting of winter wheat, making a greater quantity of stored soil-water available to the plant, thereby reducing potential water stress and harvesting more yields [2, 3]. With the spreading and applying of fertilizer, winter wheat yield on the Loess Plateau had increased from an average of 1696 kg/hm^2 for the period of 1980–1985 to 3438 kg/hm^2 for 1986–2010. However, larger above ground biomass and transpirational leaf area,

stimulated by increased fertilizer, results in greater transpiration demands and amount of water loss from the crop canopy [4]. Therefore, this increased productivity had increase soil water depletion and reduced available soil water in deep soil [5]. Excessive consumption of soil water has become the key reason for soil desiccation and yield fluctuation in high-yield land farm [6]. It is urgently need to determine a sustainable fertilizer level for this region as well as the similar region in the world.

To carry out an experiment in different regions and in a long period may found out an answer for the sustainable fertilizer level in different rainfall regions. However, it is a long-term endeavor that is both expensive and time-consuming. An alternative approach is to use computer model to simulate soil water content and crop yield under different fertilizer levels in different rainfall regions based on local situations (soil, crop and climate etcetera). Several models have been developed to simulate soil water and crop yield [7-9]. One such model is the EPIC model that simulates the soil water and crop yield simultaneously with the help of its two sub-models (growth model and hydrology model) [10]. Wang and Li [11] evaluated EPIC model for crop yield and soil water content among different cropping systems (spring maize, winter wheat and alfalfa) on the Loess Plateau. They found that EPIC model estimate soil water and crop yield well with the new database built up for the Loess Plateau.

The primary objective of this study was to determine a sustainable fertilizer level for winter wheat in different rainfall regions on the Loess Plateau. The secondary objective was to evaluate EPIC model for crop yield and soil water among different fertilizer levels, using a long-term experimental data at Changwu Agricultural Station.

1   Materials and Methods

1.1   Field Experiment

The field experiment was carried out at Changwu Agricultural Station from 1985 to 2000. It consisted of three fertilization treatments (table 1) with three replications in 9 plots of 10.26×6.5 m (with a buffer zone of 1m between plots). Plots were arranged as a randomized complete block design. All fertilizers were mixed and applied at sowing, and winter wheat was sown at the rate of 19.5 kg/hm^2, using a no-till disk drill with the row space of 0.25 m. For grain yield determination, the plots were harvested manually. Soil samples were taken by core break method [12] in 0.1 m layers to the depth of 3 m soil. Soil water content was measured (gravimetrically) for each soil sample by the oven-drying method [13].

Table 1. Fertilizer treatments for the winter wheat at Changwu Agricultural Station from 1985 to 2000

Treatments	N (kg/hm^2)	P$_2$O$_5$ (kg/hm^2)
CK	0	0
N	120	
NP	120	60

1.2   EPIC Model

EPIC is a widely tested and adopted process-based agro-ecological model originally built to quantify the effects of soil erosion on productivity [10, 14]. Currently the model has evolved into a comprehensive model capable of simulating photosynthesis, evapo-transpiration and other major plant and soil processes [15]. The model runs on a daily time step and needs daily weather data as well as information on soil properties, specific crop growth parameters and farm management practices [16].

Based on crop parameters and other related parameters, the EPIC model can calculate the uptakes of soil water and nutrients by crop, estimate the impacts of temperature, water, nutrients (N, P and k), air and salt stresses on crop biomass accumulation and crop yield by daily step [17]. For soil water, the EPIC model contains algorithms that allow for a description of the hydrological balance at the small watershed [18]. Calculated hydrological processes include snowmelt, surface runoff, infiltration, soil water content, percolation, lateral flow, water table change, and evapo-transpiration at a daily time step, details was given out by Sharply and Williams [19] and Williams et al. [15].

1.3   Methods

1.3.1   Evaluation of the EPIC Model

In this study, six statistical values were used to evaluate the model performance as followings. Root Mean Square Error (RMSE), Relative Root Mean Square Error (RRMSE), Relative Error (RE), Model Efficiency (ME), Correlation Coefficient (R), Determination Coefficient (R^2), detailed information for these equations can found in Wang and Li [11] and Ko et al. [20].

1.3.2   Calculation for WUEG and FUEG

Water-use efficiency was calculated as equation 1 [21]:

$$\mathrm{WUEG} = \frac{\mathrm{GY}}{\mathrm{ET}} \tag{1}$$

Where WUEG was water-use efficiency for the grain yield, GY was the grain yield and ET was the cumulative evapo-transpiration over the growing season which was calculated using the water balance equation (equation 2):

$$\mathrm{ET} = (P + I + \mathrm{C}) - (R + D) - \mathrm{D}S \tag{2}$$

Where P was precipitation, I was irrigation, C was upward flow into the root zone, R was surface runoff, and D was downward drainage out of the root zone. DS was the change of available soil water in winter wheat field, which was calculated by equation 3.

$$\mathrm{DS} = \mathrm{ASP} - \mathrm{ASH} \tag{3}$$

Where ASP was available soil water in 0-7 m soil when winter wheat was sown, ASH was available soil water in 0-7 m soil when winter wheat was harvested.

Since the experimental field was terraced, and located in the Loess Plateau, surface runoff was ignored. The groundwater table was very low, so the upward flow into the root zone and the downward drainage out of the root zone were negligible. Consequently, the soil water balance equation was reduced to equation 4.

$$ET = P -(\ ASP - ASH \ ) \tag{4}$$

Increase rate of grain yield for different fertilizer level was calculated by equation 5, 6 and 7.

$$IRFG = \frac{DG}{DN} \tag{5}$$

$$DG = GN_m - GN_{m-1} , 6 \geq m \geq 1 \tag{6}$$

$$DN = N_m - N_{m-1} , 6 \geq m \geq 1 \tag{7}$$

Where DG was increased yield from fertilizer level N_{m-1} to fertilizer N_m, DN was increased fertilizer of N from fertilizer level N_{m-1} to fertilizer N_m; GN_m was grain yield of winter wheat under fertilizer level of N_m, GN_{m-1} was grain yield of winter wheat under fertilizer level of N_{m-1}; N_m was amount of N in fertilizer level of N_m, N_{m-1} was amount of N in fertilizer level of N_{m-1}.

1.3.3   Description of the Development of Dry Soil Layer

In order to compare the development of dry soil layer among different fertilizer levels and different regions, parameters as following were considered in this study. 1) SMDDT Period, a Period in which soil water decreased and dry soil layers thickened continuously; 2) SSMD Year, a year in which steady dry soil layer built; 3) Max DSL, Maximum depth of dry soil layer occurred; 4) SSL Range, a depth range in which soil water content was stable after steady dry soil layer built; 5) SMC Range, a depth range in which soil water content was unstable after steady dry soil layer built; 6) SWU Depth, Maximum depth in which soil water was used; 7) DSLD Range, a depth range in which dry Soil Layers occurred; 8) SD Speed, dry soil layer building speed.

Parameters of SMDDT Period and SSMD Year described the time needed to building dry soil layer; Max DSL and DSLD Range presented the distribution depth of dry soil layer; SSL Range and SMC Range told us the stability of dry soil layer; SD speed indicated the building speed of dry soil layer.

1.3.4   Statistical Method

Analysis of variance (ANOVA) was used to test the difference in grain yield, water-use efficiency and fertilizer-use efficiency among different treatments and different regions. Mean comparisons were made by the LSD (the least significant difference) method with $P < 0.05$ and $P<0.01$ respectively. The analyses were conducted using the SPSS program [22].

1.4   Design for the Simulation

Mean annual rainfall at Luochuan and Changwu were 622 mm and 584 mm respectively; at Yan'an and Shouyang were 535 mm and 455 mm respectively. In this study, Luochuan and Yan'an were selected as the representation of semi-humid region and semi-arid region respectively; Changwu as the representation of semi-humid and drought-prone region; Shouyang as the semi-arid and drought-prone region on the Loess Plateau. Fertilizer treatments for each regions consisted of 7 fertilizer treatments as table 2.

Table 2. Fertilizer treatments for winter wheat in different rainfall regions on the Loess Plateau of China

Fertilizer level	Luochuan (kg/hm^2) N	P	Changwu (kg/hm^2) N	P	Yan'an (kg/hm^2) N	P	Shouyang (kg/hm^2) N	P
N_0	0	0	0	0	0	0	0	0
N_1	90	45	90	45	90	45	90	45
N_2	120	60	120	60	120	60	120	60
N_3	150	75	150	75	150	75	150	75
N_4	180	90	180	90	180	90	180	90
N_5	210	105	210	105	210	105	210	105
N_6	240	120	240	120	240	120	240	120

First, we input soil data, crop data, meteorological data and management data into the EPIC model for each selected region. Second we evaluate the EPIC model using the long-term experimental data at Changwu Agricultural Station. Third we run the EPIC model from 1961 to 2000, and out put the simulated winter wheat yield year by year and soil water day by day. At last we analyzed the fluctuation of crop yield and soil water during a long period, based on the simulation results. Considering the sustainable using of soil water and production of winter wheat, we pointed out a sustainable fertilizer level for different rainfall regions on the Loess Plateau.

2   Results

2.1   Evaluation Results of the EPIC Model

2.1.1   Winter Wheat Yield

Measured mean annual winter wheat yield were 1.40, 2.28 and 2.98 t/hm^2; Simulated mean annual winter yield were 1.46, 2.30 and 2.77 t/hm^2 for CK, N and NP respectively. Table 3 showed that simulated winter wheat yield were slightly higher for CK and N treatment and were slightly lower for NP treatment, comparing with measured yield. Paired-t test demonstrated the difference between simulated and

measured winter wheat yield for each treatment was not significant with P=0.05. EPIC model simulated mean annual winter wheat yield well, with the R vale of 0.91, 0.93 and 0.89 for CK, N and NP respectively.

Table 3. Comparison of simulated and measured winter wheat yield for different fertilizer levels at Changwu Agricultural Station

	Annual mean yield(t/hm^2)		RE (%)	R	RMSE (t/hm^2)	RRMSE (%)
	Measured	Simulated				
CK	1.40	1.46	4.4	0.91**	0.41	29
N	2.28	2.30	1.2	0.93**	0.53	23
NP	2.98	2.77	-7.1	0.89**	0.86	28

Little difference between ME and R^2 value (figure 1) presented the variance of winter wheat yield in different years was simulated well by EPIC model with the RRMSE value of 29 %, 23 % and 28 % respectively (table 3). Comparing with that for NP, winter wheat yield was simulated better with a higher value of R^2 for CK and N (figure 1).
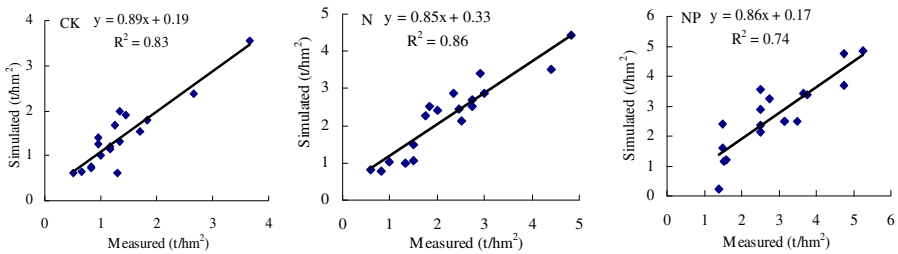
Fig. 1. Comparison of simulated and measured winter wheat yield for different fertilizer levels at Changwu Agricultural Station

2.1.2   Soil Water

Measured mean available soil water in 0-3 m soil were 120, 96 and 85mm for CK, N and NP respectively; simulated mean value were 113, 97 and 92 mm respectively. The value of mean available soil water estimated by EPIC model was slightly higher than that of simulated for N and NP, and slightly lower for CK (table 4). Paired-t test indicated the difference between simulated and measured available soil water in 0-2 m soil for each treatment was not significant with the P value of 0.05. EPIC model estimated mean annual available soil water well, with the R vale of 0.90, 0.96 and 0.81 for CK, N and NP respectively.

Table 4. Comparison of simulated and measured available soil water in 0-3m soil for different fertilizer levels at Changwu Agricultural Station

	Annual available soil water (mm)		RE (%)	R	RMSE (mm)	RRMSE (%)
	Measured	Simulated				
CK	120	113	-5.8	0.90	22.38	19
N	96	97	1.0	0.96	16.44	17
NP	85	92	8.2	0.81	18.37	23

Little difference between ME and R^2 value (figure 2) showed the variance of available soil water in different years for each fertilizer level was simulated well by EPIC model with the RRMSE value of 19%, 17% and 23% respectively (table 4). Comparing with that for NP, available soil water in 0-3 m soil was simulated better with a higher value of R^2 for CK and N (figure 2).
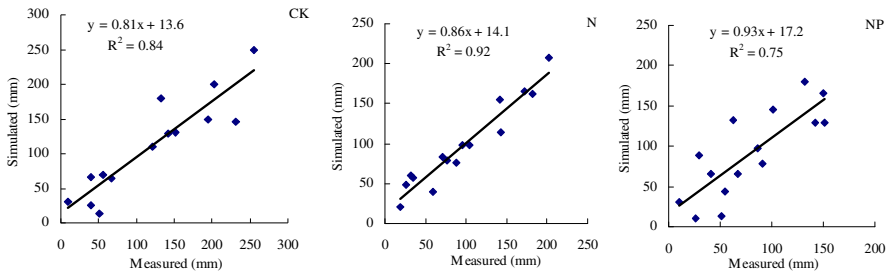
Fig. 2. Comparison of simulated and measured annual mean available soil water in 0-3m soil for different fertilizer levels at Changwu Agricultural Station

2.2 Simulation Results of Winter Wheat Yield and Soil Water

2.2.1 Winter Wheat Yield

The difference of mean annual winter wheat yield was highly significant (P=0.01) in different rainfall regions (table 5). Mean value of annual winter wheat yield were 2.77, 2.70, 1.97 and 1.00 t/hm², with standard deviation of 1.52, 1.59, 1.40 and 0.84 t/hm². With the increase amount of precipitation, more winter wheat yield was founded at Luochuan and Changwu.

The difference of winter wheat yield among N_0, N_1, N_2, and N_3 fertilizer level was significant (table 5). The difference of winter wheat yield between N_5 and N_6 fertilizer level was not significant. Before N3 fertilizer level, winter wheat yield increased with the increase amount of fertilizer application; after N4 fertilizer level, the impact of fertilizer on winter wheat was different in different rainfall regions (table 5).

Table 5. Difference of winter wheat yield among different fertilizer levels in different rainfall regions on the Loess Plateau of China

Fertilizer level	Luochuan	Changwu	Yan'an	Shouyang
N_0	1.20 a	1.44 a	0.96 a	0.53 a
N_1	1.83 b	1.94 b	1.36 b	0.69 b
N_2	2.53 c	2.45 c	1.76 c	0.90 c
N_3	3.00 d	2.87 d	2.08 d	1.07 d
N_4	3.44 e	3.24 e	2.43 e	1.22 de
N_5	3.59 f	3.38 ef	2.54 ef	1.25 ef
N_6	3.78 f	3.56 f	2.67 f	1.33 f

The correlation between winter wheat yield and precipitation over growing season was highly significant (P=0.01); and it was found significant (P=0.05) between winter wheat yield and precipitation over growing year (table 6). Table 6 showed that winter wheat yield was highly significant correlated with the available soil water in 0-3 m soil before sowing. This indicated that available soil water in 0-3 m soil before sowing and precipitation in growth period was the key impact factors to influence winter wheat yield on the Loess Plateau. Since the correlation index between winter wheat yield and available soil water in 0-7 m soil before sowing was significant (P=0.05), soil water in 0-7 m soil should be considered when to determine a sustainable fertilizer level for the winter wheat on the Loess Plateau.

Table 6. Winter wheat yield, precipitation and available soil water in different rainfall regions on the Loess Plateau of China

Fainfall region	Mean annual yield (t/hm^2)	Correlation index between yield and			
		Precipitation over (mm)		Available soil water before sowing in (mm)	
		Growing season	Growing year	0-3 m soil	0-7 m soil
Luochuan	2.77 a	0.66**	0.42*	0. 81**	0.57*
Changwu	2.70 b	0.63**	0.38*	0. 78**	0.60*
YanAn	1.97 c	0.80**	0.43*	0. 80**	0.70*
Shouyang	1.00 d	0.51**	0.34*	0. 79**	0.68*

Growing season of winter wheat was from September to June; growing year was from September to August.

2.2.2   Soil Water

2.2.2.1   Available Soil Water in 0-7 m Soil. With the increase amount of precipitation, more available soil water was founded in winter wheat field at luochuan and Changwu. Mean value of monthly available soil water in 0-7 m soil were 740,

558, 512 and 412 mm, with standard deviation of 102, 140, 137 and 99 mm, at luochuan, Changwu, Yan'an and Shouyang respectively. Considering among different rainfall regions, precipitation was one of the key factors to impact the available soil water in winter wheat field.

The value of DS in different rainfall regions was different significantly. Table 7 showed more soil water decreased at Changwu and Yan'an, with the mean DS value of 113 and 114mm respectively. Comparing with that in Shouyang, more precipitation and less DS value was founded at Changwu and Yan'an. Analysis of the winter wheat growth indicated the leaf area index (LAI) of winter wheat at Shouyang was significantly less than that at Changwu and Yan'an. It means that less trans-evaporation at Shouyang, comparing with that at Changwu and Yan'an. A highest value of LAI and precipitation was found at Luochuan. Though higher trans-evaporation (indicated by the high value of LAI) took place and may be decreased soil water, more precipitation added more water to the soil and decreased the DS value at Luochuan.

Comparing with N_1, N_2 and N_3 fertilizer level, a higher value of DS was founded in N_4, N_5 and N_6 fertilizer level. This indicated that more soil water decreased under the higher fertilizer level in winter wheat field on the Loess Plateau.

Table 7. Mean value of decreased available soil water from winter wheat was sowed to winter wheat was harvested

	N_0	N_1	N_2	N_3	N_4	N_5	N_6
Luochuan (mm)	61	71	78	82	91	93	94
Changwu (mm)	85	99	109	116	125	126	130
Yan'an (mm)	86	100	111	111	127	128	131
Shouyang (mm)	76	86	93	99	104	106	111

2.2.2.2   Development of Dry Soil Layer in 0-7m Soil. Fertilizer and rainfall were two key factors to effect the development of dry soil layer in winter wheat field on the Loess Plateau of China. Table 5 indicated that higher fertilizer level increased the winter wheat yield; at the same time increase the SD speed and the thickness of dry soil layer (table 8). Highest value of SD speed and highest SSL range were founded in N6 fertilizer level at four selected regions. With the increase amount of fertilizer applied to winter wheat, the value of SD and Max DSL increased (table 8). With the increasing amount of fertilizer level, more soil water was depleted and soil desiccation degree increased. More rainfall resulted in lower SD speed and SSL range in Luochuan and Changwu, comparing with that in Shouyang and Yan'an.

Table 8. Statistical value of soil water distribution in 0-7 m soil in winter wheat field under different fertilizer levels in different rainfall regions on the Loess Plateau

Rainfall region	treat ment	SMDDT Period	SSMD Year	Max DSL (m)	DSLD Range (m)	SSL Range (m)	SMC Range (m)	SD Speed (m/a)	SWU Depth (m)
Luochuan	N_0	1960~1968	1968	3	1~3	2~3	0~1	0.38	2~3
	N_1	1960~1968	1969	3	1~3	2~3	0~2	0.33	2~3
	N_2	1960~1967	1971	4	1~4	2~4	0~2	0.36	>4
	N_3	1960~1967	1970	4	1~4	2~4	0~2	0.4	>4
	N_4	1960~1969	1971	5	1~5	2~5	0~2	0.42	>5
	N_5	1960~1968	1971	5	1~5	2~5	0~2	0.45	>5
	N_6	1960~1968	1970	5	1~5	2~5	0~1	0.5	>5
Changwu	N_0	1960~1968	1969	3	1~3	2~3	0~2	0.33	2~3
	N_1	1960~1967	1968	3	1~3	2~3	0~2	0.38	2~3
	N_2	1960~1967	1971	4	1~4	2~4	0~2	0.36	>4
	N_3	1960~1968	1970	4	1~4	2~4	0~2	0.4	>4
	N_4	1960~1966	1972	5	1~5	2~5	0~2	0.45	>5
	N_5	1960~1966	1971	5	1~5	2~5	0~1	0.45	>5
	N_6	1960~1966	1970	5	1~5	2~5	0~1	0.5	>5
Yan'an	N_0	1960~1966	1967	3	1~3	2~3	0~2	0.43	2~3
	N_1	1960~1966	1967	3	1~3	2~3	0~2	0.43	2~3
	N_2	1960~1966	1969	4	1~4	2~4	0~2	0.44	>4
	N_3	1960~1965	1968	4	1~4	2~4	0~2	0.5	>5
	N_4	1960~1965	1969	5	1~5	2~5	0~2	0.56	>5
	N_5	1960~1965	1968	5	1~5	2~5	0~1	0.63	>5
	N_6	1960~1965	1968	5	1~5	2~5	0~2	0.63	>5
Shouyang	N_0	1960~1966	1966	3	1~3	2~3	0~1	0.5	2~3
	N_1	1960~1965	1966	3	1~3	2~3	0~2	0.5	2~3
	N_2	1960~1965	1968	4	1~4	2~4	0~2	0.5	>4
	N_3	1960~1965	1967	4	1~4	2~4	0~2	0.57	>5
	N_4	1960~1965	1968	5	1~5	2~5	0~2	0.63	>5
	N_5	1960~1965	1967	5	1~5	2~5	0~2	0.71	>5
	N_6	1960~1965	1967	5	1~5	2~5	0~1	0.71	>5

2.2.3   WUEG and IRFG

Table 9 showed the value of WUEG increased, with increase application of fertilizer to winter wheat field. Mean value of WUEG among different regions for N_0, N_1, N_3, N_2, N_4, N_5 and N_6 fertilizer level were 2.76, 3.80, 4.83, 5.57, 6.32, 6.59 and 6.92 t/(hm^2·mm) respectively. Among different rainfall regions, WUEG increased with the increasing of annual rainfall at Luochuan, Changwu, Yan'an and Shouyang (table 9).

From N_1 to N_4 fertilizer level, the value of IRFG increased with the increase of fertilizer applied to the winter wheat field (table 9). From N_4 to N_5 fertilizer level, a contrary changing trend was founded for the value of IRFG. The value of IRFG became lower, when the fertilizer was more than 180 kg/hm^2. Among different rainfall regions, the value of IRFG was the highest at Luochuan, it was the second at Changwu, and the lowest value was at Shouyang (table 9). Annual rainfall was one of the key factors to influence fertilizer using efficiency on the Loess Plateau of China.

Table 9. Comparison of WUEG among different fertilizer levels and different rainfall regions on the Loess Plateau of China

Rainfall regions	items	N_0	N_1	N_2	N_3	N_4	N_5	N_6
Luochuan	WUEG (t/(hm^2·mm))	3.14	4.73	6.37	7.22	8.38	8.76	9.19
	IRFG (t/(hm^2·kg))	--	0.70	1.11	1.20	1.24	1.14	1.08
Changwu	WUEG (t/(hm^2·mm))	3.53	4.61	5.63	6.50	7.24	7.55	7.93
	IRFG (t/(hm^2·kg))	--	0.56	0.84	0.95	1.00	0.92	0.88
Yan'an	WUEG (t/(hm^2·mm))	2.67	3.66	4.58	5.33	6.07	6.35	6.66
	IRFG (t/(hm^2·kg))	--	0.43	0.67	0.75	0.82	0.75	0.71
Shouyang	WUEG(t/(hm^2·mm))	1.71	2.20	2.73	3.23	3.58	3.70	3.91
	IRFG (t/(hm^2·kg))	--	0.18	0.30	0.36	0.38	0.34	0.33

3   Discussions

3.1   Fertilizer, Dry Soil Layer and Rainfall

Nitrogen fertilizer stimulated the winter wheat root to develop into the deeper soil, and using more deep soil water [23]. However, once dry soil layer built in deep soil, it was difficult to recover on the Loess Plateau [24]. Therefore, the use of high fertilization rates has been found to increase the thickness of dry soil layer [25] and resulted the fluctuation of winter wheat yield in different rainfall years on the Loess Plateau [6]. This paper showed that more soil water in deep soil depleted and the SD speed (soil desiccation speed) increased, with the increasing application of fertilizer for winter wheat. The amount of annual rainfall impacted the develop speed of dry soil layers in different regions. Comparing with Luochuan and Changwu, higher value of SD speed was founded at Yan'an and Shouyang.

3.2   Fertilizer, WUEG and IRFG

The increasing use of deep soil water during critical crop development stages increased the value of WUEG (water use efficiency for grain yield) for high fertilizer winter wheat [5]. Various study found that water and fertilizer related and impacted each other [26-27]. This study indicated higher value of IRFG and WUEG in higher fertilizer level field, when no dry soil layer build in deep soil; while lower IRFG value in higher fertilizer level field, when dry soil layers had been built in deep soil.

3.3   Sustainable Fertilizer Level for Winter Wheat Field

At Luochuan, the difference of winter wheat yield for N_5 and N_6 fertilizer level was not significant; there was not significant difference between winter wheat yield for N_5 and N_6 at Changwu and Yan'an, and also for N_4 and N_5 fertilizer level was not significant; At Shouyang, there was no significant difference between winter wheat yield for N_5 and N_6, N_4 and N_5, N_3 and N_4. These statistical results indicated that fertilizer level should not be higher than N_5, N_4, N_4 and N_3, at Luochuan, Changwu, Yan'an and Shouyang respectively.

When dry soil layers develop to more than 5 m depth, it will become more difficult to recover [28] on the Loess Plateau of China. Table 8 showed that the depth of dry soil layer was deeper than 5m for N_4, N_5 and N_6 at Luochuan and Changwu; for N_3, N_4, N_5 and N_6 at Yan'an and Shouyang. Therefore, considering on the recovery of dry soil layers, fertilizer level for winter wheat should not be higher than N_4 at Luochuan and Shouyan; and it should not be higher than N_3 at Yan'an and Shouyang. Table 9 indicated the value of IRFG became lower when the fertilizer level was higher than N_4 for winter wheat field, either at Luochuan and Changwu or at Yan'an and Shouyang. This result indicated the fertilizer using efficiency decreased when fertilizer level was higher than N_4 for winter wheat field.

Considering on the sustainable production of winter wheat and the sustainable using of soil water, the fertilizer level for the different rainfall regions on the Loess Plateau should as followed. 1) In the semi-humid region (Luochuan), fertilizer level should be from N_4 to N_5; 2) in the semi-humid and drought-prone region (Changwu) and in the semi-arid region (Yan'an), it should be from N_3 to N_4; 3) in the semi-arid and drought-prone region (Shouyang), it should be lower than N_3.

4   Conclusions

The EPIC model simulated the variance of winter wheat yield among different fertilizer levels well, with the mean R value of 0.91; and it estimated mean annual available soil water well, with the mean R vale of 0.89, on the Loess Plateau of China.

Soil desiccation speed increased with the increasing of fertilizer applied to winter wheat field; and it decreased with the increasing of annual rainfall in different rainfall regions on the Loess Plateau of China.

With the increase application of fertilizer, the value of IRFG and WUEG became higher, when soil water in deep soil was not be used excessively; the value of IRFG became lower, when soil water in deep soil was used excessively.

Sustainable fertilizer levels for winter wheat on the Loess Plateau were different for different rainfall regions. 1) In the semi-humid region (Luochuan), it should be from N_4 to N_5; 2) in the semi-humid and drought-prone region (Changwu) and in the semi-arid region (Yan'an), it should be from N_3 to N_4; 3) in the semi-arid and drought-prone region (Shouyang), it should be lower than N_3.

References

[1] Zhu, X.: Soil and Agriculture in the Loess Plateau. Agricultural Science Press, Beijing (1989) (in Chinese)

[2] Brown, P.L.: Water use and soil-water depletion by dryland winter wheat as affected by nitrogen fertilization. Agron. J. 63, 43–46 (1971)

[3] Read, D.W.L., Warder, F.G., Cameron, D.R.: Factors affecting fertilizer nitrogen response of wheat insouthwestern Saskatchewan. Can. J. Soil Sci. 62, 577–586 (1982)

[4] Ritchie, J.T., Johnson, B.S.: Soil and plant factors affecting evaporation. In: Stewart, B.A., Nielsen, D.R. (eds.) Irrigation of Agricultural Crops, Agronomic Monograph, ASA, CSSA, SSSA, Madison, WI, USA, vol. 30, pp. 363–390 (1990)

[5] Huang, M.B., Dang, T.H., Jacques, G., Monique, G.: Effect of increased fertilizer applications to wheat crop on soil-water depletion in the Loess Plateau, China. Agricultural Water Management 58, 267–278 (2003)

[6] Li, Y.S.: Fluctuation of yield on high-yield field and desiccation of the soil on dryland. Chin. J. Acta Pedol. Sin. 38(3), 353–355 (2001)

[7] Zhang, L., Dawes, W.: WAVES—An Integrated Energy and Water Balance Model, CSIRO Land and Water Technical Report no. 31/98, Australia (1998)

[8] Van Genuchten, M.T.: A Numerical Model for Water and Solute Movement in and Below the Root Zone. Research Report No. 121. US Salinity Laboratory, USDA, ARS, Riverside, CA (1987)

[9] Jones, J.W., Batchelor, W.D., Hoogenboom, G., Porter, C.H., Boote, K.J., Hunt, L.A., Wilkens, P.W., Singh, U., Gijsman, A.J., Ritchie, J.T.: The DSSAT cropping system model. Europ. J. Agron. 18, 235–265 (2003)

[10] Williams, J.R., Jones, C., Dyke, P.T.: A modeling approach to determining the relationship between erosion and soil productivity. Trans. ASAE 27, 129–144 (1984)

[11] Wang, X.C., Li, J.: Evaluation of crop yield and soil water estimates using the EPIC model for the Loess Plateau of China. Math. Comput. Model 51, 1390–1397 (2010)

[12]   Bennie, A.T.P., Taylor, H.M., Georgen, P.G.: An assessment of the core-break method for estimating root density of different crops in the field. Soil Tillage Res. 9(24), 343–347 (1987)

[13]   Blake, G.R., Hartge, K.H.: Bulk density. In: Klute, A. (ed.) Methods of Soil Analysis. Part I. Physical and Mineralogical Methods, pp. 363–382. American Society of Agronomy, Madison (1986)

[14]   Williams, J.R., Jones, C.A., Kiniry, J.R., Spanel, D.A.: The EPIC crop growth model. Trans. ASAE 32, 497–511 (1989)

[15]   Williams, J.R.: The EPIC model. In: Singh, V.P. (ed.) Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch (1995)

[16]   Robert, A.B., Norman, J.R.: Sensitivity of crop yield and water use to change in a range of climatic factors and CO, concentrations: a simulation study applying EPIC to the central USA. Agr. Forest. Meteorol. 83, 171–203 (1997)

[17]   Niu, X.Z., Easterling, W., Hays, C.J., Jacobs, A., Mearns, L.: Reliability and input-data induced uncertainty of the EPIC model to estimate climate change impact on sorghum yields in the U.S. Great Plains. Agric. Ecosys. Environ. 129, 268–276 (2009)

[18]   Izaurralde, R.C., Williams, J.R., McGill, W.B., Rosenberg, N.J., Quiroga Jakas, M.C.: Simulating soil C dynamics with EPIC: Model description and testing against long-term data. Ecol. Model. 192, 362–384 (2006)

[19]   Williams, J.R., Jones, C.A., Dyke, P.T.: The EPIC Model. In: Williams, J.R. (Eds.), EPIC—Erosion Productivity Impact Calculator. 1. Model Documentation. pp. 3–86, U.S. Department of Agriculture Technical Bulletin No. 1768 (1990)

[20]   Ko, J., Piccinni, G., Stelich, E.: Using EPIC model to manage irrigated cotton and maize. Agric. Water Manag. 96, 1323–1331 (2009)

[21]   Hussain, G., Al-Jaloud, A.A.: Effect of irrigation and nitrogen on water use efficiency of wheat in Saudi Arabia. Agric. Water Manage. 27, 143–153 (1995)

[22]   SPSS Inc., SPSS for Windows Base System User's Guide Release 6.0. Marija J. Norusis/SPSS Inc. (1977)

[23]   Nielsen, D.C., Halvorson, A.D.: Nitrogen fertility influence on water stress and yield of winter wheat. Agron. J. 83, 1065–1070 (1991)

[24]   Wang, X.L., Chen, M.C., Li, F.M., Li, Y.J.: Water restoration of dry soil layers in the Loess Plateau and crop yield response. Res. Soil Water Conserv. 14(3), 1–4 (2007)

[25]   Wang, X.C., Li, J., Jiang, B., Hu, W.: Simulation of yield and soil desiccation effects of continuous spring maize in different precipitation areas of the Loess Plateau. Acta Ecol. Sin. 29(4), 2053–2066 (2009)

[26]   Mitchell, C.C., Westerman, R.L., Brown, J.R., Peck, T.R.: Overview of long-term agronomic research. Agron. J. 83, 24–29 (1991)

[27]   Sandor, J.A., Eash, N.S.: Significance of ancient agriculture soils for long-term agronomic studies and sustainable agriculture research. Agron. J. 83, 29–37 (1991)

[28]   Wang, X.C., Muhammad, T.N., Hao, M.D., Li, J.: Sustainable recovery of soil desiccation in semi-humid region on the Loess Plateau. Agric. Water Manage. 98, 1262–1270 (2011)

Based on Vague Sets of Strawberry Varieties Resistance Comparison

Yan Zhang[1], Hongxu Wang[2], and Hongbin Zhang[1]

[1] College of Biological Science and Technology,
Qiongzhou University,
Sanya, P.R. China 572022
`zy-wzs1991@163.com`
[2] Electronic Information Engineering College,
Qiongzhou University,
Sanya, P.R. China 572022

Abstract. Proposing the similarity measures formula between Vague sets E and G. Concluding Vague resistance analysis. The concrete application steps were: ①Establishing comprehensive characters set; ②Screening excellent varieties set; ③ Extracting theory optimal varieties set; ④ The single-value date transformed into Vague date, obtaining different varieties Vague sets; ⑤Vague resistance analysis 1,the similarity measures were calculated between the excellent varieties Vague sets and the theory optimal varieties Vague sets, obtaining more suitable for people's needs varieties among excellent varieties; ⑥ Vague resistance analysis 2, the weighted similarity measures were calculated between the excellent varieties Vague sets and the theory optimal varieties Vague sets, obtaining more suitable for people's needs resistance varieties among excellent varieties. According to specific needs of the problem, selecting and applying the steps ⑤ or ⑥. Vague resistance analysis was applied to strawberry varieties resistance analysis, the analysis result was satisfactory. The similarity measures formula between Vague sets E and G was the application basis of Vague resistance analysis.

Keywords: Vague sets, Similarity measures formula, Vague resistance analysis, Strawberry varieties, Resistance analysis.

1   Introduction

Strawberry is rich nutrition and has a variety of organic acids, minerals and vitamins, especially rich in vitamin c. It is red tender, juicy, sweet delicious and fragrant. It is often known as the empress of the fruit and health food enjoyed by young and old. In recent years, Hainan has actively introduced strawberry fine varieties from home and abroad. This paper intends to use vague resistance analysis to research strawberry varieties resistance analysis, so as to screening fine strawberry varieties more suitable for planting in Hainan.

2   Basic Concepts

2.1   Vague Sets Definition

Definition 1[1]. Set non-empty universe Z. For $z \in Z$, regulations interval $[t_E(z), 1 - f_E(z)]$ for Vague sets E of Z at dot z Vague value or Vague membership, among them $0 \le t_E(z) \le 1, 0 \le f_E(z) \le 1,$ and meets constraints $t_E(z) + f_E(z) \le 1$. $t_E(z)$ 、 $f_E(z)$ 、 $\pi_E(z)(= 1 - t_E(z) - f_E(z))$ is called respectively truth-membership function, false-membership function and uncertain function of Vague set E.

When $Z = \{z_1, z_2, \cdots, z_n\}$ for discrete universe, its Vague sets E can be written

as $E = \sum_{i=1}^{n} [t_E(z_i), 1 - f_E(z_i)] / z_i$, or $E = \sum_{i=1}^{n} [t_{e_i}, 1 - f_{e_i}] / z_i$.

2.2   Creating Vague Environment

Creating Vague environment is that original data is transformed into Vague data. It is the application premises of Vague resistance analysis. The following only introduces the definition that the single-value data transformed into theVague data.

Definition 2[2]. Set $Z = \{z_1, z_2, \cdots, z_n\}$ as discrete universe, Z has sets $E_i (i = 1, 2, \cdots, m)$, E_i as the single-value data $z_{ij} (\ge 0)$ that represents the original data of   comprehensive characters $z_j (j = 1, 2, \cdots n)$.

a. Vague conditions. $0 \le t_{ij} \le 1 - f_{ij} \le 1$;

b. Output conditions. If $0 \le z_{kj} < z_{ij}$, the single-value data z_{ij} and z_{kj} are respectively transformed into the Vague data $E_i(z_j) = z_{ij} = [t_{ij}, 1 - f_{ij}]$ and $E_k(z_j) = z_{kj} = [t_{kj}, 1 - f_{kj}]$ meet: $t_{kj} \le t_{ij}, 1 - f_{kj} \le 1 - f_{ij}$. Called single-value data $z_{ij} (\ge 0)$ that meets Vague conditions and output conditions transformed into the Vague data $E_i(z_j) = z_{ij} = [t_{ij}, 1 - f_{ij}]$ type of conversion formula for the output-based conversion formula.

c. Investment conditions. If $0 \le z_{kj} < z_{ij}$, the single-value data z_{ij} and z_{kj} are respectively transformed into the Vague data $E_i(z_j) = z_{ij} = [t_{ij}, 1 - f_{ij}]$ and $E_k(z_j) = z_{kj} = [t_{kj}, 1 - f_{kj}]$ meet: $t_{kj} \ge t_{ij}, 1 - f_{kj} \ge 1 - f_{ij}$. Called single-value data $z_{ij} (\ge 0)$ that meets Vague conditions and investment conditions transformed

into the Vague data $E_i(z_j) = z_{ij} = [t_{ij}, 1 - f_{ij}]$ type of conversion formula for the investment-based conversion formula.

Annotation: When the value of comprehensive characters is bigger always better, the output-based conversion formula is suitable used. And when the value of comprehensive characters is smaller always better, the investment-based conversion formula is suitable used.

2.3   A Kind Vague Membership Data Mining

Definition 3[3]. The method of a kind Vague membership data mining is: Vague membership $e = [t_e, 1 - f_e]$, denoted by $t_e^{(0)} = t_e, f_e^{(0)} = f_e, \pi_e^{(0)} = \pi_e = 1 - t_e - f_e$. When $m = 1, 2, \cdots$, and regulations $t_e^{(m)} = t_e \cdot (1 + \pi_e + \pi_e^2 + \cdots \pi_e^m), f_e^{(0)} = f_e \cdot (1 + \pi_e + \pi_e^2 + \cdots + \pi_e^m), \pi_e^{(m)} = \pi_e^{m+1}$.

Lemma 1[3]. $e^{(m)} = [t_e^{(m)}, 1 - f_e^{(m)}]$ is Vague membership.

Annotation: Definition 3 has put forward the method of Vague membership data mining, it regards Vague membership $e = [t_e, 1 - f_e]$ mining into Vague membership $e^{(m)} = [t_e^{(m)}, 1 - f_e^{(m)}]$ ($m = 1, 2, \cdots$) . This paper by means of Vague membership data mining constructs the new similarity measures between Vague membership.

2.4   Similarity Measures between Vague Membership

Definition 4[4]. Set $e = [t_e, 1 - f_e]$ and $g = [t_g, 1 - f_g]$ as two Vague membership. The formula $M(e, g)$ is called similarity measures between Vague membership e and g . If the formula meets the following conditions:

    a. Trivial conditions. $M(e, g) \in [0,1]$;

    b. Symmetric conditions. $M(e, g) = M(g, e)$;

    c. Reflexive conditions. $M(e, e) = 1$;

    d. Minimum conditions. When $e = [0,0], g = [1,1]$ or $e = [1,1], g = [0,0]$, they all guarantee $M(e, g) = 0$.

The definition of the similarity measures and the weighted similarity measures between Vague sets may be similar to definition 4, here is omitted.

Annotation: $M(e, g)$ expresses similar degree between Vague value e and g . Its meaning is that the larger of the value of $M(e, g)$ expresses more similar between Vague value e and g ; Especially when $M(e, g)$ takes maximum 1, expressing most similar between Vague value e and g ; The smaller of the value of

$M(e, g)$ expresses more dissimilarity between Vague value e and g; Especially when $M(e, g)$ takes minimum 0, expressing most dissimilarity between Vague value e and g.

3   New Theorem and New Method

Theorem 1. $z_{j\min} = \min\{z_{1j}, z_{2j}, \cdots, z_{mj}\}, z_{j\max} = \max\{z_{1j}, z_{2j}, \cdots, z_{mj}\}$. then

$$\text{a.} \quad E_i(z_j) = z_{ij} = [t_{ij}, 1 - f_{ij}] = \left[\frac{z_{ij} - z_{j\min}}{z_{j\max} - z_{j\min}}, \left[\frac{z_{ij} - z_{j\min}}{z_{j\max} - z_{j\min}} \right]^{\frac{1}{2}} \right] \tag{1}$$

It is conversion formula $E_i(z_j) = z_{ij} = [t_{ij}, 1 - f_{ij}]$ of single-value data $z_{ij} (\geq 0)$ transformed into Vague data for output-based conversion formula.

$$\text{b.} \quad E_i(z_j) = z_{ij} = [t_{ij}, 1 - f_{ij}] = \left[1 - \left[\frac{z_{ij} - z_{j\min}}{z_{j\max} - z_{j\min}} \right]^{\frac{1}{2}}, 1 - \frac{z_{ij} - z_{j\min}}{z_{j\max} - z_{j\min}} \right] \tag{2}$$

It is conversion formula $E_i(z_j) = z_{ij} = [t_{ij}, 1 - f_{ij}]$ of single-value data $z_{ij} (\geq 0)$ transformed into Vague data for investment-based conversion formula.

Theorem 2. Set $e = [t_e, 1 - f_e]$ and $g = [t_g, 1 - f_g]$ for two Vague membership. The following formula is similarity measures between Vague membership e and g ($m = 0, 1, 2, \cdots$):

$$M_m(e, g) = \frac{1 + \min\{t_e^{(m)} - t_g^{(m)}, f_e^{(m)} - f_g^{(m)}\}}{1 + \max\{t_e^{(m)} - t_g^{(m)}, f_e^{(m)} - f_g^{(m)}\}}. \tag{3}$$

Application literature[2] method, not difficult checking out that resolution of formula (3) is higher. Similar theorem 2 obtains the following results.

Theorem 3. Set $Z = \{z_1, z_2, \cdots, z_n\}$ as the universe, Z has Vague sets $E = \sum_{i=1}^{n} [t_E(z_i), 1 - f_E(z_i)] / z_i$ and $G = \sum_{i=1}^{n} [t_G(z_i), 1 - f_G(z_i)] / z_i$.

Abbrevd. $E = \sum_{i=1}^{n} [t_{e_i}, 1 - f_{e_i}] / z_i, G = \sum_{i=1}^{n} [t_{g_i}, 1 - f_{g_i}] / z_i$.

The following formula is similarity measures between Vague sets E and G ($m = 0, 1, 2, \cdots$):

$$M_m(E,G) = \frac{1}{n}\sum_{i=1}^{n}\frac{1+\min\left\{t_{e_i}^{(m)}-t_{g_i}^{(m)}, f_{e_i}^{(m)}-f_{g_i}^{(m)}\right\}}{1+\max\left\{t_{e_i}^{(m)}-t_{g_i}^{(m)}, f_{e_i}^{(m)}-f_{g_i}^{(m)}\right\}}. \tag{4}$$

Theorem 4. Set element z_i weight $0 \le w_i \le 1$, and $\sum_{i=1}^{n} w_i = 1$. On conditions theorem 3, the following formula is weighted similarity measures between Vague sets E and G $(m = 0,1,2,\cdots)$:

$$WM_m(E,G) = \sum_{i=1}^{n} w_i \cdot \frac{1+\min\left\{t_{e_i}^{(m)}-t_{g_i}^{(m)}, f_{e_i}^{(m)}-f_{g_i}^{(m)}\right\}}{1+\max\left\{t_{e_i}^{(m)}-t_{g_i}^{(m)}, f_{e_i}^{(m)}-f_{g_i}^{(m)}\right\}}. \tag{5}$$

Vague Resistance Analysis

Concluding literature[4] Vague sets comprehensive decision rules for Vague resistance analysis. The concrete application steps are: ①Establishing comprehensive characters set; ② Screening excellent varieties set; ③ Extracting theory optimal varieties set; ④ The single-value date transformed into Vague date, obtaining different varieties Vague sets; ⑤Vague resistance analysis 1,the similarity measures are calculated between the excellent varieties Vague sets and the theory optimal varieties Vague sets, obtaining more suitable for people's needs varieties among excellent varieties; ⑥Vague resistance analysis 2, the weighted similarity measures are calculated between the excellent varieties Vague sets and the theory optimal varieties Vague sets, obtaining more suitable for people's needs resistance varieties among excellent varieties. According to specific needs of the problem, selecting and applying the steps ⑤ or ⑥.

4   Strawberry Varieties Resistance Comparison

Strawberry cultivation range is very wide, in recent years, Hainan has actively introduced strawberry fine varieties from home and abroad. Selection excellent comprehensive characters strawberry varieties is main work for improving strawberry cultivation yield, therefore, we screened five varieties for Frandy, Maiterli, Kinuama, Rafi, Fengxiang and developed five varieties resistance comparison experiment. We applied Vague resistance analysis to analyse, in order to select excellent varieties more appropriate Hainan plastic greenhouse cultivation, to improve strawberry cultivated benefit and economic benefit.

4.1   Establishing Comprehensive Characters Set

Establishing comprehensive characters set $Z = \{z_1, z_2, \cdots, z_8\}$:

Index z_1 : survival rate of seedlings (%); z_2 : survival rate of plant (%); z_3 : yield(t/hm^2); z_4 : fruit commodity rate(%); z_5 : leaves heating-damage rate(%);

z_6: young fruit freezing-damage rate(%); z_7: blight incidence(%); z_8: grey cinerea incidence(%).

4.2   Screening Excellent Varieties Set

Selecting Hainan plastic greenhouse cultivation main fine varieties for excellent varieties sets $E = \{E_1, E_2, E_3, E_4, E_5\}$, among them E_1 : Frandy; E_2 : Fengxiang; E_3 : Maiterli; E_4 : Kinuama; E_5 : Rafi. They all are sets of comprehensive characters set $Z = \{z_1, z_2, \cdots, z_8\}$.   Resistance comparison experiment original data shown in Table 1.

4.3   Extracting Theory Optimal Varieties Set

Because the value of z_1 , z_2 , z_3 , z_4 is bigger always better, the value of z_5, z_6, z_7, z_8 is smaller always better. So extraction theory optimal varieties G concrete data shown in Table 1.

Table 1. The original data of resistance comparison experiment

	E_1	E_2	E_3	E_4	E_5	G
z_1	96.60	92.20	97.30	95.30	94.80	97.30
z_2	93.80	90.70	96.60	93.80	91.10	96.60
z_3	15.79	11.38	16.00	11.94	13.91	16.00
z_4	87.80	71.40	81.20	72.20	86.40	87.80
z_5	9.20	28.10	6.60	24.50	19.70	6.60
z_6	52.10	60.80	54.80	49.40	55.50	49.40
z_7	16.60	8.50	5.80	5.80	27.20	5.80
z_8	8.00	18.80	11.60	15.20	6.90	6.90

4.4   The Single-Value Date Transformed into Vague Date, Obtaining Different Varieties Vague Sets

Can be seen from 4.3, application formula(1) to comprehensive characters z_1 , z_2, z_3, z_4 , application formula(2) to comprehensive characters z_5, z_6, z_7, z_8 , they make the original data in Table 1 transform into Vague data and obtain different varieties Vague sets(see table 2).

4.5   Vague Resistance Analysis 1

Application formula(4)(take $m = 2$) calculates the similarity measures between the excellent varieties Vague sets and the theory optimal varieties Vague sets, the results show:

$M_2(E_1, G) = 0.70, M_2(E_2, G) = 0.09, M_2(E_3, G) = 0.77,$

$M_2(E_4, G) = 0.42, M_2(E_5, G) = 0.41$. The preferential order of strawberry varieties adaptation Hainan is: Maiterli(E_3), Frandy(E_1), Kinuama(E_4), Rafi(E_5), Fengxiang(E_2). First choice Maiterli.

4.6   Vague Resistance Analysis 2

The cold resistant strawberry varieties need to be considered at slightly cold incidental heavy fetch in middle of Hainan.

Application formula(5) (take $m = 2$), take weight:

$w_1 = 0.1, w_2 = 0.1, w_3 = 0.1, w_4 = 0.1, w_5 = 0.1, w_6 = 0.3, w_7 = 0.1, w_8 = 0.1$

the weighted similarity measures are calculated between the excellent varieties Vague sets and the theory optimal varieties Vague sets, the results show:

$WM_2(E_1, G) = 0.67, WM_2(E_2, G) = 0.07, WM_2(E_3, G) = 0.66, WM_2(E_4, G) = 0.53,$

$WM_2(E_5, G) = 0.37$. The preferential order of strawberry varieties adaptation slightly cold incidental heavy fetch in middle of Hainan is: Frandy(E_1), Maiterli(E_3), Kinuama(E_4), Rafi(E_5), Fengxiang(E_2). First choice Frandy.

Table 2. Vague data of resistance comparison experiment

	E_1	E_2	E_3	E_4	E_5	G
z_1	[0.92,0.96]	[0.00,0.00]	[1.00,1.00]	[0.61,0.78]	[0.51,0.71]	[1.00,1.00]
z_2	[0.53,0.73]	[0.00,0.00]	[1.00,1.00]	[0.53,0.73]	[0.07,0.26]	[1.00,1.00]
z_3	[0.96,0.98]	[0.00,0.00]	[1.00,1.00]	[0.12,0.35]	[0.55,0.74]	[1.00,1.00]
z_4	[1.00,1.00]	[0.00,0.00]	[0.60,0.77]	[0.05,0.22]	[0.92,0.96]	[1.00,1.00]
z_5	[0.65,0.88]	[0.00,0.00]	[1.00,1.00]	[0.09,0.17]	[0.22,0.39]	[1.00,1.00]
z_6	[0.51,0.76]	[0.00,0.00]	[0.31,0.53]	[1.00,1.00]	[[0.27,0.47]	[1.00,1.00]
z_7	[0.29,0.50]	[0.65,0.87]	[1.00,1.00]	[1.00,1.00]	[0.00,0.00]	[1.00,1.00]
z_8	[0.70,0.91]	[0.00,0.00]	[0.37,0.61]	[0.17,0.30]	[1.00,1.00]	[1.00,1.00]

5   Conclusion

Through strawberry varieties Vague resistance analysis, it gave a new method to study such problems, but also it enriched Vague pattern recognition theory. The method was a kind of pattern recognition method, it made pattern recognition between excellent varieties and theory optimal varieties. The recognition tool was the similarity measures formula between Vague sets and the weighted similarity measures formula between Vague sets. New formula(1)～(5) was used in Vague resistance analysis, especially new formula(4) ～ (5) was corroborated, such formula was indispensable for Vague pattern recognition theory[5].

References

[1] Gau, W., Buehrer, D.J.: Vague Sets. IEEE Transactions on Systems. Man and Cybernetics 23(2), 610–614 (1993)
[2] Wang, H.: Comprehensive evaluation of new wheat varieties applying Vague optimized decision-making method. Computer Engineering and Applications 47(12), 210–212 (2011)
[3] Liu, H., Wang, F.: Transformations and Similarity Measures of Vague Sets. Computer Engineering and Applications 40(32), 79–81, 84 (2004)
[4] Wang, H.: Synthesis decision rule of vague sets and its application in scheme optimum seeking. Computer Engineering and Applications 46(27), 145–147 (2010)
[5] Liu, H.: Basis of fuzzy pattern recognition-similarity measures. Pattern Recognition and Artificial Intelligence 17(2), 141–145 (2004)

Evaluating New Varieties of Wheat with the Application of Vague Optimization Methods

Hongxu Wang[1], FuJin Zhang[2], and Yunsheng Xu[3]

[1,3]College of Science and Engineering,
[2]College of Electronic Information Engineering, Qiongzhou University,
Sanya Hainan 572022, China
zfj56801@163.com, whx16233@yahoo.cn

Abstract. The Vague optimization method sorted out is a special case of the pattern recognition method of the Vague set. Its specific application steps are as follows: 1 the set-up of characters to determine the evaluation; 2 the establishment of a collection of new varieties of wheat to be optimized; 3 the extract of the ideal set of new wheat varieties ; 4 the construction of Vague environment to collect all varieties of Vague sets; 5 the Vague optimization: to calculate similarity measures between Vague sets to obtain the new wheat varieties based on numerical similarity measures and to propose the similarity measures formula between the Vague sets. This optimization formula is supported by Vague class techniques. Breeding new varieties of wheat with comprehensive quality traits is one of the directions in wheat breeding. It is a new attempt to study the wheat assessment of the new wheat varieties with the application of optimization methods. The wheat assessment case of the new wheat varieties shows that both the formula and the method are practical.

Keywords: Vague set, Vague optimization, Vague environment, similarity measure, assessment of new varieties of wheat.

1   Introduction

Vague sets [1] were put forward as Fuzzy Sets' population[1]. Though it was put forward 28 years later than Fuzzy set theory, Vague set theory has been widely used in many fields such as in pattern recognition, automatic control, intelligent reasoning, for it uses the interval numbers of Vague set membership to indicate degree of membership, which can more fully reflect the fuzzy information. This paper aims to apply the Vague set theory to the evaluation of new varieties of wheat and to provide the new research methods to the agriculture-related issues.

2   Basic Concept

Definition 1[1]. Suppose L be a point of space, any one of the elements can be indicated with l. L on a Vague set is S, it is a membership function with a true t_s and

a false membership function f_S said. $t_S(l)\, l$ is evidence to support the export of certain lower bound of the membership, and l is the evidence from the opposition $f_S(l)$ derived the lower bound of the negative membership. $t_S(l)$ and $f_S(l)$ will be a real number in the range of point of contact with the L together. That is, map $t_S : L \rightarrow [0,1], f_S : L \rightarrow [0,1]$, and satisfies the constraints: $1 \geq t_S(l) + f_S(l)$. Vague set can be recorded L points l in the membership or Vague is simplified by $S(l) = [t_S(l), 1 - f_S(l)]$ or $l = [t_l, 1 - f_l]$.

It can be seen from the definition of $l = [t_l, 1 - f_l] \subseteq [0,1]$. $\pi_S = 1 - t_S - f_S$ that on the set of S elements in l pairs is uncertain function Vague, also called the degree of uncertainty or the degree of Vague. If L is a discrete domain of $L = \{l_1, l_2, \cdots, l_n\}$ and L on Vague set S can be written as $S = \sum_{i=1}^{n} \frac{[t_S(l_i), 1 - f_S(l_i)]}{l_i}$,

also denoted by $S = \sum_{i=1}^{n} \frac{[t_{l_i}, 1 - f_{l_i}]}{l_i}$.

3   Vague Environmental Construction

The so-called construction of Vague environment is to turn the raw data into Vague data. This step is a prerequisite for the application of Vague sets. The single-value data are turned into the data definition Vague proposed in Reference [3], here again the formula for turning single-value data into a Vague data is provided in the assessment of new wheat varieties.

Definition 2. Let the domain $L = \{l_1, l_2, \cdots, l_n\}$, L, has a collection of $S_i (i = 1, 2, \cdots m)$, $S_i\, l_j (j = 1, 2, \cdots n)$ data set of characters with non-negative single-value data l_{ij}. Suppose when the single-value data when the non-negative l_{ij} Vague data turning into the formula to meet the output of the conversion of $S_i(l_j) = l_{ij} = [t_{ij}, 1 - f_{ij}]$ conditions and Vague terms, we call this formula the formula for the output type, among which:

a. Output Conditions
If $0 \leq l_{kj} < l_{ij}$, l_{ij} and l_{kj} single-value data turn into the Vague data respectively, $S_i(l_j) = l_{ij} = [t_{ij}, 1 - f_{ij}]$ and $S_k(l_j) = l_{kj} = [t_{kj}, 1 - f_{kj}]$ satisfy   the   condition: $t_{kj} \leq t_{ij}, 1 - f_{kj} \leq 1 - f_{ij}$.

b. Input Conditions
If $0 \leq l_{kj} < l_{ij}$, l_{ij} and l_{kj} single-value data turn   into   the   Vague   data, $S_i(l_j) = l_{ij} = [t_{ij}, 1 - f_{ij}]$ and $S_k(l_j) = l_{kj} = [t_{kj}, 1 - f_{kj}]$ satisfy   the   condition: $t_{kj} \geq t_{ij}, 1 - f_{kj} \geq 1 - f_{ij}$

c. Vague Terms $0 \le t_{ij} \le 1 - f_{ij} \le 1$.

Note: Output type transformation formula for numerical characters takes "the bigger, the better" in use; input type transformation formula for numerical characters takes "the smaller the better" in use.

Theorem 1: If $l_{j\max} = \max\{l_{1j}, l_{2j}, \cdots, l_{mj}\}$, then

$$S_i(l_j) = l_{ij} = [t_{ij}, 1 - f_{ij}] = \left[\left(\frac{l_{ij}}{l_{j\max}} \right)^2, \left(\frac{l_{ij}}{l_{j\max}} \right)^{\frac{1}{2}} \right] \tag{1}$$

is non-negative single-value data l_{ij} into the output data type Vague conversion formula.

$$S_i(l_j) = l_{ij} = [t_{ij}, 1 - f_{ij}] = \left[1 - \left(\frac{l_{ij}}{l_{j\max}} \right)^{\frac{1}{2}}, 1 - \left(\frac{l_{ij}}{l_{j\max}} \right)^2 \right] \tag{2}$$

is non-negative single-value data l_{ij} into the input data type Vague conversion formula.

4   The New Similarity Measures between Vague Values

If **Lemma** [4] is to Vague value of $l = [t_l, 1 - f_l]$, then $l^{(m)} = [t_l^{(m)}, 1 - f_l^{(m)}]$ is Vague value. Vague value $l = [t_l, 1 - f_l]$ of the formula is known as the first m to (t_l, f_l) data mining Vague value, among which:

$$f_l^{(m)} = f_l(1 + \pi_l + \pi_l^2 + \cdots + \pi_l^m), \alpha_l^{(m)}$$
$$= (t_l + f_l)(1 + \pi_l + \pi_l^2 + \cdots + \pi_l^m), \beta_l^{(m)}$$
$$= (t_l - f_l)(1 + \pi_l + \pi_l^2 + \cdots + \pi_l^{(m)}), (m = 0, 1, 2, \cdots).$$

A simple and practical Vague measure of similarity between the values is defined by Reference [5].

Definition 3. If $s = [t_s, 1 - f_s], h = [t_h, 1 - f_h]$ is two Vague values, the formula $M(s, h)$ between s and h is similarity measures of the Vague value, if $M(s, h)$ satisfies the following conditions:

   a. 0-1 conditions $0 \le M(s, h) \le 1$
   b. Symmetry conditions $M(s, h) = M(h, s)$
   c. Reflexive conditions $M(s, s) = 1$
   d. Minimum condition

When $s = [0,0], h = [1,1]$ or $s = [1,1], h = [0,0]$, the $M(s,h) = 0$. Vague, which is also known as the numerical value of $M(s,h) = 0$ similarity between s and h.

Note: the similarity measure of Vague value between s and h, means that the greater the similarity values of $M(s,h)$, s and h, the more similar Vague values between s and h, especially when the maximum similarity value of $M(s,h)$ is up to 1, the value of Vague s and h is the most similar; conversely, the smaller the similarity value of $M(s,h)$, s and h, less similar Vague values between s and h, especially when the minimum similarity value of $M(s,h)$ reaches 0, value of Vague s and h is the least similar.

Theorem 2. Note $m = 0,1,2,\cdots$, the formula

$$M_m(s,h) = \left[1 - \left|f_s^{(m)} - f_h^{(m)}\right|\right] \cdot \left[1 - \frac{\left|\beta_s^{(m)} - \beta_h^{(m)}\right| + \left|\alpha_s^{(m)} - \alpha_h^{(m)}\right|}{2}\right] \tag{3}$$

is the similarity measure of Vague value between s and h.

5   Vague Similarity Measures between the New Vague Sets of Metrics

From the definition of similarity measures between Vague Sets and Vague Sets similarity between the weighted measure, similar to Definition 3, theorem 2 can also be obtained from the following theorem.

Theorem 3. Let the domain $L = \{l_1, l_2, \cdots, l_n\}$, L on $S = \sum_{i=1}^{n} \frac{[t_S(l_i), 1 - f_S(l_i)]}{l_i}$ and $H = \sum_{i=1}^{n} \frac{[t_H(l_i), 1 - f_H(l_i)]}{l_i}$ Vague sets are simply denoted as $S = \sum_{i=1}^{n} \frac{[t_{s_i}, 1 - f_{s_i}]}{l_i}$ and $H = \sum_{i=1}^{n} \frac{[t_{h_i}, 1 - f_{h_i}]}{l_i}$ respectively. And note $m = 0,1,2,\cdots$, The formula

$$M_m(S,H) = \frac{1}{n} \sum_{i=1}^{n} \left[1 - \left|f_{s_i}^{(m)} - f_{h_i}^{(m)}\right|\right] \cdot \left[1 - \frac{\left|\beta_{s_i}^{(m)} - \beta_{h_i}^{(m)}\right| + \left|\alpha_{s_i}^{(m)} - \alpha_{h_i}^{(m)}\right|}{2}\right] \tag{4}$$

is the similarity measures of Vague sets between s and h of the metrics.

Theorem 4. If the element weight $l_i (i = 1,2,\cdots,n)$ is $b_i \in [0,1]$, and $\sum_{i=1}^{n} b_i = 1$, and note $m = 0,1,2,\cdots$. then in the conditions in Theorem 3, the formula

$$M_m(S,H) = \sum_{i=1}^{n} b_i \cdot \left[1 - \left|f_{s_i}^{(m)} - f_{h_i}^{(m)}\right|\right] \cdot \left[1 - \frac{\left|\beta_{s_i}^{(m)} - \beta_{h_i}^{(m)}\right| + \left|\alpha_{s_i}^{(m)} - \alpha_{h_i}^{(m)}\right|}{2}\right] \tag{5}$$

is the weighted similarity measures between Vague sets S and H.

6   Vague Optimization

The comprehensive decision-making rules of Vague sets in Reference [6] are reorganized into Vague optimization method fit for the research on the assessment of new varieties of wheat. Their specific application steps are as follows: 1 the set-up of characters to determine the evaluation; 2 the establishment of a collection of new varieties of wheat to be optimized; 3 the extract of the ideal set of new wheat varieties; 4 the construction of Vague environment to collect all varieties of Vague sets; 5 the Vague optimization: to calculate similarity measures between Vague sets to obtain the new wheat varieties based on numerical similarity measures. If Vague sets of pattern recognition is for the purpose of optimization, then the pattern recognition method of Vague sets can be called the Vague optimization method. So Vague set optimization method is a special case of the pattern recognition. It has one standard model and several to be recognized over the Vague sets of pattern recognition methods.

7   Optimal Assessment of New Varieties of Wheat

Fuzzy Comprehensive Evaluation in Reference [7] is used to study the assessment of new varieties of wheat. In this paper, Vague optimization method is applied to re-examine the issue.

7.1   To Establish the Evaluation Set of Traits

Take l_1 as the "yield (kg/667m)"; l_2 as "strain rate (%)"; l_3 as the" drought index "; l_4 as a" valid spike number (million / 667m) "; l_5 as" grains per spike (tablets) "; l_6 as" grain weight (g) ", then the evaluation of the set of characters is $L = \{l_1, l_2, \cdots, l_6\}$.

Table 1. The characteristics of new varieties of wheat, the average data

Factor	The characteristics of new varieties of wheat, the average data				
	S_1	S_2	S_3	S_4	H
l_1	285.8	278.1	271.3	256.7	285.8
l_2	6.4	3.3	10.3	7.8	3.3
l_3	1.203	1.413	1.039	1.101	1.413
l_4	38.80	40.45	34.5	35.60	40.45
l_5	33.1	32.3	32.4	32.0	33.1
l_6	37.7	32.5	38.8	37.4	38.8

7.2   To Set Up the Optimal Set of New Varieties of Wheat

Take S_1 as the "Liao 97 Kam 30",S_2 as "Zhemai 10", S_3 as "Kyrgyzstan Spring 9806",S_4 as the "Liao Chun 9 (ck)".   They are composed of a collection of traits evaluated on a set of $L = \{l_1, l_2, \cdots, l_6\}$. They consist of a collection of new varieties of wheat to be optimized. The specific data are from Reference [7], as shown in Table 1.

7.3   To Extract the Ideal Project

Reference [7] tells that if the new varieties of wheat are expected to be all "optimal" in all traits, then the bigger the l_1 is, the better should be; the smaller the l_2 is, the better; the bigger the l_3 is, the better; the more the l_4 is, the better; the more the l_5 is the better; the heavier l_6 is, the better. Extracting the best data of various traits available to evaluate the set of characters composed of a collection of $L = \{l_1, l_2, \cdots, l_6\}$. on the H,   which is called the ideal of new wheat varieties. The data for each trait are shown in Table 1.

7.4   Enter the Vague Environment

Formula (1) can be applied to Characters in $l_1, l_3, l_4,$ l_5 and l_6 ; Formula (2) is applied to l_2. Table 1 can be turned into Table 2.

Table 2. The characteristics of new varieties of wheat Vague data

Factor	The characteristics of new varieties of wheat Vague data				
	S_1	S_2	S_3	S_4	H
l_1	[1.000,1.000]	[0.947,0.986]	[0.901,0.974]	[0.806,0.948]	[1.000,1.000]
l_2	[0.212,0.614]	[0.434,0.898]	[0.000,0.000]	[0.271,0.438]	[0.434,0.898]
l_3	[0.724,0.922]	[1.000,1.000]	[0.540,0.857]	[0.607,0.880]	[1.000,1.000]
l_4	[0.920,0.979]	[1.000,1.000]	[0.728,0.924]	[0.774,0.938]	[1.000,1.000]
l_5	[1.000,1.000]	[0.953,0.988]	[0.958,0.989]	[0978,0.994]	[1.000,1.000]
l_6	[0.945,0.986]	[0.702,0.915]	[1.000,1.000]	[0.929,0.982]	[1.000,1.000]

Table 2 shows the selection of the new wheat varieties of S_1, S_2, S_3,S_4,and H new varieties of wheat ideal Vague collection.

7.5   Vague Optimization

New varieties of wheat are sorted out through calculating the similarity measures between Vague sets, according to the size of numerical similarity measure. The

formula (4) is as follows: take $m = 2$, calculate the similarity measures between Vague sets $S_i (i = 1, 2, 3, 4)$ and H. The results are:

$$M_2(S_1, H) = 0.843, M_2(S_2, H) = 0.956, M_2(S_3, H) = 0.729, M_2(S_4, H) = 0.770.$$

Thus, according to the size of similarity measures, the order is:

$$M_2(S_2, H) > M_2(S_1, H) > M_2(S_4, H) > M_2(S_3, H).$$

The similarity of Vague sets is that the larger the number is, the more similar these two Vague sets are; the smaller the value is, the more dissimilar the two Vague sets are. So the conclusion of Vague optimization decision: among which the best variety is S_2: Zhemai 10. The second is S_1: Liao 97 Kam 30. and poorer one is S_4: Liao Chun 9 (ck) and S_3: Ji Chun 9806,S_4: Liao Chun 9 (ck).

8   Conclusion

The Fuzzy comprehensive evaluation method used applied in Reference [7], although it is able to "make up for the lack of analysis of variance," is relatively more complicated, in particular it needs to calculate fuzzy matrix, whose speed is rather slow. The optimization method, by means of wheat Vague Selection of new varieties, is feasible. It is more reasonable and easier to use than the existing one, and it provides a more practical approach in optimizing the new wheat varieties. tend to provide. From the example mentioned-above, Vague optimization method is an alternative to the fuzzy comprehensive evaluation. As we all know, the application of fuzzy comprehensive evaluation process is more cumbersome, and with a large number of calculation. Vague optimization method not only provides a new method for discussing the similar problems, and its application in the fields such as species optimization is of a great potential area. And the formula (1), (2), (3), (4) and (5) put forward in this paper provide the technical support for it.

References

1. Gau, W.-L., Buehrer, D.J.: Vague Sets. IEEE Transactions on Systems, Man and Cybernetics 23(2), 610–614 (1993)
2. Zadeh, L.A.: Fuzzy Sets. Information and Control (8), 338–353 (1965)
3. Wang, H.-x.: Definition and transforming formulas from the single valued data to the vague valued data. Computer Engineering and Applications 46(24), 42–44 (2010)
4. Liu, H.-W., Wang, F.-Y.: Transformations and Similarity Measures of Vague Sets. Computer Engineering and Applications 40(32), 79–81 (2004)

5. Wang, H.-X.: Similarity measure between vague sets and their application. Computer Engineering and Applications 46(26), 198–199 (2010)
6. Wang, H.-X.: Synthesis decision rule of vague sets and its application in scheme optimum seeking. Computer Engineering and Applications 46(27), 145–147 (2010)
7. Modeleto, Erdunga, Bayilatu, et al.: The fuzzy comprehensive evaluation to the application of new wheat evaluation. Inner Mongolia Agricultural Science and Technology (1), 37–38 (2007)

Research on Effect of Ventilated Region to Grape Rain Shed Performance Based on CFD

Jian Wang[1] and Hongbing Yang[2,*]

[1] College of Horticulture, Nanjing Agricultural University, Nanjing, China 210095
[2] College of Engineering, Nanjing Agricultural University, Nanjing, China 210031
{wangjian,hbyang}@njau.edu.cn

Abstract. Two different kinds of multi-span grape rain sheds are constructed and the wind field and rainfall field distribution are simulated based on CFD and computer simulation techniques. The airflow distribution, temperature variation, wind pressure distribution and rainfall vectors distribution are compared under two conditions with ventilated region and without ventilated area in multi-span grape rain shed. The research results show that set the ventilated area on the top of grape rain shed can improve effectively the high temperature and high humidity climate environment, and it has a benefit to grapes fertility.

Keywords: CFD, Multi-span grape rain shed, ventilated region, simulation.

1 Introduction

Rainproof cultivation is one important form of grape protected cultivation in southern China. Grape rainproof cultivation has two main merits. On the one hand, the grape fruit set is promoted, and the grape quality is improved. On the other hand, the soil moisture is controlled effectively, and the fungus spread and the amount of pesticide spraying are decreased significantly [1]. However, under the effects of the temperature rise rapidly and grape transpiration, the temperature and humidity are very high on the top of grape rain shed during May to July, and it can be damaged the grape fruit set. If the ventilated region is set on the top of shelter shed, the high temperature and humidity conditions may be improved. Nevertheless, the construction of grape rain shed is changed when ventilated area is set, and the climate environments of shelter shed how to change when ventilated region is set. These new problems need to be researched in detail. In recent years, the computer simulation techniques have a wide range of applications in agricultural facilities. Briassoulis and Mistriotis [2] proposed the integrated structural design methodology for agricultural protecting structures covered with nets based on computer simulation method. Harmanto [3] simulated the airflow distribution of greenhouse under different mesh sizes of nets and predicted the air exchange rate. M. Teitela[4]researched the effect of

---

wind direction on greenhouse ventilation rate, airflow patterns and temperature distributions based CFD.

The aims of this research were to analysis the effect of ventilated region to climate environment, and evaluate the possibility of installing ventilated region in the grape rain shed. The airflow distribution, temperature variation, wind pressure distribution and water vectors distribution are compared under two conditions of ventilated region and non-ventilated area in grape rain shed.

2   Material and Method

2.1   Grape Multi-span Greenhouse

The grape multi-span greenhouse was located in the coastal areas of Jiangsu province. The multi-span grape rain shed is built by a new kind of coated PE plastic steel pipe and the schematic diagram is shown in figure 1a. The four-span greenhouse (length is12m and height is 2.374m) is covered with polyethylene film. One meter wide vent is set each two spans on the top of multi-span grape rain shed. The schematic diagram with ventilated region is shown in figure 1b.

(a) without ventilated region   (b) with ventilated region
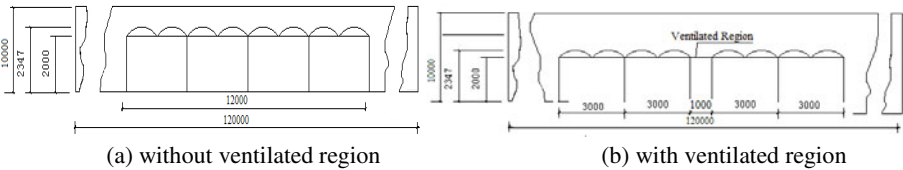
Fig. 1. Schematic diagram multi-span grape rain shed (mm)

Wind pressure formula [5] is give by

$$p_v = \frac{1}{2}C\rho_{ao}v_0^2 \tag{1}$$

Where C is pressure shape factor, ρ_{ao} is outdoor air density(kg/m^3)and v_0 is outdoor wind speed (m/s).

Boulard and Baille [6, 7] put forward the ventilation rate formula. When only considering the role of wind pressure, the ventilation rate G is

$$G = (S/2)AC_w^{0.5}\bar{v} \tag{2}$$

Where S is vent cross-sectional area(m^2), A is vent distribution coefficient, C_w is wind pressure effect coefficient and $\bar{v}$ is average wind speed(m/s).

2.2   CFD Numerical Simulation Method

Computational fluid dynamics (CFD) is a sophisticated design and analysis tool that allows us to study different design limitations based on a computational model. The CFD software used in the present work was FLUENT 6.2. Its code uses a discretization procedure of finite volumes and it incorporates the equations that govern heat flow and transfer, such as the continuity equation, momentum conservation and energy conservation. The overall methodology used to carry out a complete study with the commercial software used in this work is summarized as follows: generating the geometry of the fluid domain, meshing, defining the boundary conditions and the physical properties of the model (pre-processing), solving and finally analysis (post-processing) [8]. The classical mass, momentum, energy and concentration equations can be represented for a steady-state, three-dimensional flow with the following conservation equation.

$$\frac{\partial \phi}{\partial t} + \frac{\partial}{\partial x_j}\left(u_j \phi\right) = \frac{\partial}{\partial x_j}\left(\Gamma_\phi \frac{\partial \phi}{\partial x_j}\right) + S_\phi \tag{3}$$

Where Φ stands for the variables of interest, i.e. the three velocity components μ_j (ms^{-1}), the temperature T (K), Γ_Φ and S_Φ represent the diffusion coefficient and source term of Φ and a description of their forms can be found in Boulard [9-10]. The system of equations built with these variables is numerically solved with the finite volume method. Algorithms and methods for the resolution of this system of equations can be found in the fluent 6.2 user's guide. To model the turbulent constraints using the standard k-ε turbulence model , in Eq. 1 Φ stands also for the turbulent kinetic energy k (m^{2}s^{-2}) and the dissipation of the turbulent kinetic energy ε (m^{2}s^{-3}).

The airflow field and rain field of multi-span grape rain shed are researched differently. The airflow field distribution is simulated when the wind direction is 0 degree, wind velocity is 20m/s and temperature is 308K. The rain field distribution is researched when rain direction 0 degree, rain velocity is and 5m/s and temperature is 308K.

3   Results and Discussion

3.1   Airflow Distribution Simulation of Multi-span Grape Rain Shed Based on CFD

Wind pressure distribution contours of two kinds of multi-span grape rain shed are shown in figure 2. From figure 2 we can see that outer surface of two different kinds of shelter sheds are lift state under the action of wind pressure, and their external flow field distribution is almost the same. However, the internal flow field of two kinds of rain shelters is different. The flow field distribution of span 1 and 2 with ventilated
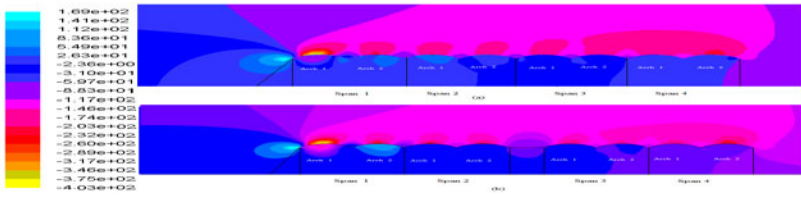
Fig. 2. Pressure distribution contours of two kinds of grape rain sheds

region is uniform, and the flow field distribution of span 3 and 4 with ventilated region is more uneven than non-ventilated region.

The wind pressure distribution on the top of shelter shed is given in figure 3. The abscissa axis represents the span of grape rain shed, and the ordinate axis stands for the wind pressure distribution. It can be seen that the wind pressure distribution of Span 1 and span 2 is similar to two different kinds of grape rain shed. The wind pressure distribution of span 3 and span 4 is different to two different kinds of grape rain shed. From figure 3(a) we can know that the top of shelter shed is lift mainly. The wind lift maximal is 370 Pa and wind pressure maximal is 72Pa in the span 1and span 2, and from Span 3 to span 4, the value range of wind of upward lift is 10~189Pa. It can be seen from figure 3(b) that wind lift maximal is 354Pa and wind pressure maximal is 216Pa in the span 1and span 2, and from Span 3 to span 4, the value range of wind pressure is 0~74Pa. Simulation results show that the wind pressure value is decreased on the top of grape rain shed when the ventilated region is installed. According to formula (1)we can see that it is proportional relationship between wind speed and air pressure, So the wind speed is decreased, the wind pressure is also decreased. The airflow velocity vectors distribution of multi-span grape rain shed with ventilated region is shown in figure 4.When the ventilated region is set, the velocity of external grape rain shed is decreased, so the wind pressure value is decreased correspondingly.
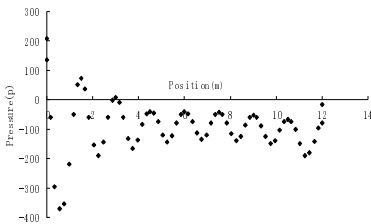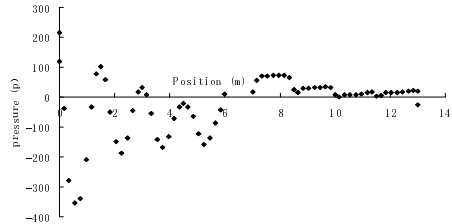
(a) without ventilated region   (b) with ventilated region

Fig. 3. Pressure distribution curve on the top of grape rain shed
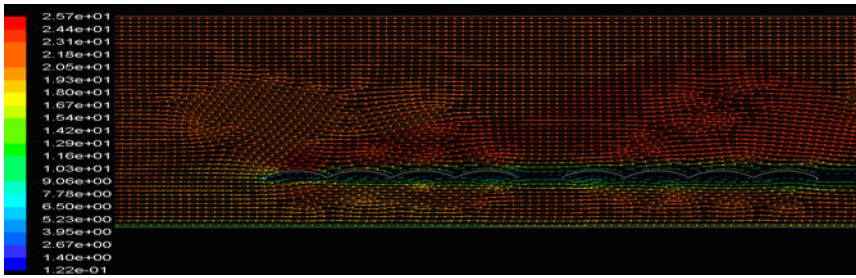
Fig. 4. Velocity vectors distribution of multi-span grape rain shed with ventilated region

3.2 Temperature and Rainproof Performance Simulation of Multi-span Grape Rain Shed Based on CFD

The internal temperature distribution of two different shelter sheds is shown in figure 5. It can be seen that the inner temperature of grape rain shed with ventilated region is
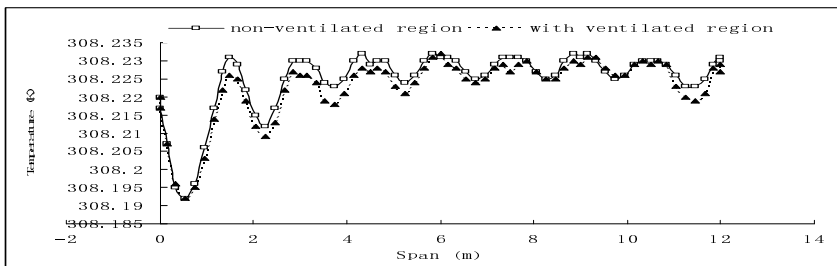
Fig. 5. Internal Temperature distribution of two different shelter sheds
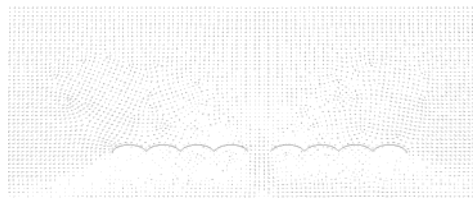
(a) without ventilated region   (b) with ventilated region

Fig. 6. Rainfall vector distribution of two different grape rain shed

decreased significantly. This is consistent with formula (2), and when the internal wind speed is increased, the ventilation rate is increased accordingly, so temperature and humidity of grape rain shed with ventilated region are deceased clearly. The rainproof performance has some impact when the ventilated region is set in grape rain shed. When rainfall vertical ground and the velocity is 5m/s, the rainproof effect of two kinds of grape rain shed is simulated in figure 6. Compared figure 6(a) and

figure 6(b), we can see that the rainproof performance without ventilated region is better than grape rain shed with ventilated region. However, it can be seen that rainfall distribution mainly at the near ground in grape rain shed clearly and these rainfall has little effect on the grape growth.

4   Summary

The two kinds of models of grape rain shed are constructed, and the airflow distribution, temperature variation, wind pressure distribution and rainfall vectors distribution are compared under two conditions of ventilated region and non-ventilated area in grape rain shed based on computer simulation. The research results show that set the ventilated area on the top of grape rain shed can improve effectively the high temperature and high humidity climate environment, and it has a benefit to grapes fertility. Of course, the construction costs of grape rain shed is increased when installing ventilated region.

References

1. Li, J.C., Shun, D.L., Zhang, X.S.: Effect of rainproof cultivation on grape disease incidence and light intensity under the shelter. Journal of Fruit Science 26(6), 847–850 (2009)
2. Briassoulis, D., Mistriotis, A.: Integrated structural design methodology for agricultural protecting structures covered with nets. J. Biosystems Engineering 105, 205–220 (2010)
3. Harmanto, H., Tantau, J., Salokhe, V.M.: Microclimate and air exchange rates in greenhouses covered with different nets in the humid tropics. J. Biosystems Engineering 94(2), 239–253 (2006)
4. Teitel, M., Ziskind, G., Liran, O., et al.: Effect of wind direction on greenhouse ventilation rate, airflow patterns and temperature distributions. J. Biosystems Engineering 101(3), 351–369 (2008)
5. Zhou, C.J.: Modern Greenhouse Engineering. Publishing House of Chemistry and Industry, Beijing (2003)
6. Boulard, T., Baille, A.: Modeling of air exchange rate in greenhouse equipped with continuous roof vents. Journal of Agricultural Engineering Research 61(1), 37–48 (1995)
7. Boulard, T., Feuilloley, P., Kittas, C.: Natural Ventilation Performance of Six Greenhouse and Tunnel Types. Journal of Agricultural Engineering Research 67(4), 249–266 (1997)
8. Franco, A., Valera, D.L., Pena, A., Pérez, A.M.: Aerodynamic analysis and CFD simulation of several cooling pads used in Mediterranean greenhouses. J. Computers and Electronics in Agriculture 76(2), 218 (2011)
9. Bartzanas, T., Boulard, T., Kittas, C.: Effect of Vent Arrangement on Windward Ventilation of a Tunnel Greenhouse. J. Biosystems Engineering 88(4), 479–490 (2004)
10. Boularda, T., Royb, J.C., Fatnassi, H., et al.: Computer fluid dynamics prediction of climate and fungal spore transfer in a rose greenhouse. J. Computers and Electronics in Agriculture 74(2), 280–292 (2010)

Research on Evaluation of Agricultural Planting Security in Hebei Province

Li Deng and Yali Li

Department of Economic and Management, Tangshan College,
Hebei, P.R. China
Dengli20072005@126.com

Abstract. Hebei agricultural planting security is relative to the agricultural products of the consumer safety of the whole province and the beijing-tianjin region of. This paper selects agricultural planting safety as the research object of products and focus on the characteristics of the agricultural products plant. This paper applies theory and method of system comprehensive evaluation to screen index of agricultural planting evaluation, in order to building evaluation model of agricultural planting safety. This paper abides by the principle of qualitative analysis with quantitative analysis, applies the method of AHP scheduling comprehensive ranking of each index, scientifically evaluates planting security system of agricultural products in Hebei province. The purpose of this paper is to provide the scientific basis in Hebei province for improving agricultural planting security policy.

Keywords: Agricultural planting security, Evaluation system, AHP.

1 Introduction

Along with the progress of science and technology of agriculture, the unit output and total production of agricultural commodities improve continuously, agricultural ended its economic shortage, appeared the situation of total balance and surplus in good years. The demand of consumers for agricultural production has been changed from pursuing the number to the pursuiting of quality and safety. Hebei is a big agricultural province with agricultural product variety, large output, links, chain long, low degree of organization, the heavy responsibility of safety supervision, and the special geographic position of encircling the beijing-tianjin and special status of supplying beijing-tianjin agricultural commodities. Therefore, evaluate agricultural planting scientific and objectively, strengthen planting safety of agricultural products globally, to guarantee the safety of agricultural planting, and ensure that the province and the beijing-tianjin region agricultural products consumer safety has the important practical significance.

2   Model Design

2.1   Specification of a Model

According to the requirements of good agriculture practice, combine with the actual situation of evaluation of agricultural planting site. From the angle of farm produce safety management, the evaluation index system will include five aspects these are land usage, soil analysis detection, irrigation water, management of environment problem and the validity of the planting site management system. This paper also establish further evaluation index of pass class time structure (as it is shown in table 1).

Table 1. Evaluation System of Agricultural Planting Security in Hebei Province

Target layer	Primary index	Secondary index
Evaluation	Soil analysis detection (B1)	Controlling of pesticides residues(C1)
		Controlling of heavy metal(C2)
System		Nutrient content(C3)
		Adequacy of the water(C4)
	Irrigation water (B2)	Quality of water (C5)
of		Controlling of pollution(C6)
	Management of	Surrounding environment (C7)
Agricultural	environment problem(B3)	Protection of ecological environment(C8)
		Protection of resources(C9)
Planting		Technology management(C10)
	Validity of land	Identification and traceability management（C11）
	management system (B4)	Fertilizer usage(C12)
Security		Protection of crops（C13）

2.2   Build Judgment Matrix

The effects of these four indicators are different to the agricultural planting security in Hebei Province. Based on this, considering the differences of Hebei province, this paper used questionnaires to visit different experts and relevant management department in order to get the data of evaluation system of agricultural planting safety in Hebei. First order judgment matrix and second order judgment matrix was built based on the indicators above.

2.3   Weight Calculation of Single Hierarchy and Consistency Test

Eigenvalues and eigenvectors were calculated by AHP Software. The weight sorting results of each effect factor were listed in the table 2 and table 3. The results of each matrix of the consistency test were listed in the bottom of each table.

Table 2. First Order Judgment Matrix

A	B1	B2	B3	B4	Sort Weight
B1	1	1/2	2	3	0.278
B2	2	1	3	4	0.467
B3	1/2	1/3	1	2	0.160
B4	1/3	1/4	1/2	1	0.095

The effect size of various indicators in evaluation system of agricultural planting security in Hebei province could be determined from table 2. The result of operation was λmax=4.031, CI=0.010, RI=0.900, CR=0.011<0.1, so the test of consistency passed. Results showed that the top one of weight sort was the irrigation water (with the value of 0.467), which played an important part in this group. The second one of weight sort was the Soil analysis detection, and it was 0.278. The sum of two was 0.745.

Table 3. Second Order Judgment Matrix

B1	C1	C2	C3	Sort Weight
C1	1	2	1/3	0.230
C2	1/2	1	1/5	0.122
C3	3	5	1	0.648

From table 3, we can determine the contribution size of controlling of pesticides residues, controlling of heavy metal, nutrient content, λmax=3.004, CI=0.002, RI=0.580, CR=0.003<0.1, so the test of consistency passed. Results showed that the greatest impact on soil analysis detection is nutrient content, and the weight sort is 0.648. The second one of weight sort was the soil analysis detection, the weight sort is 0.230. The sum of the two is 0.878.

Table 3. (Continuation Sheet 1)

B2	C4	C5	C6	Sort Weight
C4	1	3	4	0.614
C5	1/3	1	3	0.268
C6	1/4	1/3	1	0.117

The contribution size of adequacy of the water, quality of water and controlling of pollution in irrigation water could be determined from table 3 (Continuation Sheet 1). λmax=3.074, CI=0.037, RI=0.508, CR=0.064<0.1, so the test of consistency passed. Results showed that the greatest impact in irrigation water is adequacy of the water, and the weight sort is 0.614, the sum of the other two is 0.385.

Table 3. (Continuation Sheet 2)

B3	C7	C8	C9	Sort Weight
C7	1	2	5	0.582
C8	1/2	1	3	0.309
C9	1/5	1/3	1	0.109

The contribution size of surrounding environment, protection of ecological environment and protection of resources in management of environment problem could be determined from table 3(Continuation Sheet 2). λmax =3.004, CI=0.002, RI=0.580, CR=0.003<0.1, the test of consistency passed. Results showed that the greatest impact in management of environment problem is surrounding environment, the weight sort is 0.582.

Table 3. (Continuation Sheet 3)

B4	C10	C11	C12	C13	Sort Weight
C10	1	1/2	1/3	1/4	0.095
C11	2	1	1/2	1/3	0.160
C12	3	2	1	1/2	0.278
C13	4	3	2	1	0.467

The contribution size of technology management, identification and traceability management, fertilizer usage and protection ofcrops could be determined from table 3 (Continuation Sheet 3). λmax =4.031, CI=0.010, RI=0.900, CR=0.011<0.1, the test of consistency passed. Results showed that the greatest impact in validity of land management system is protection of crops. he weight sort is 0.467. The second one of weight sort was fertilizer usage, the weight sort is 0.278, the sum of the other two is 0.745.

2.4   Weight Calculation of Overall Hierarchy and Consistency Test

CI=0.019, RI=0.610, CR=0.031<0.1, so the test of consistency passed. From table 4, the conclusion that the top three indexes are adequacy of the water, quality of water and nutrient content could be drawn. Results showed that the important position in agricultural planting security system of Hebei province is ensuring sufficient irrigation water, and water quality is also very important, both of them become the most important factors in evaluation system.

Table 4. Overall Hierarchy

A	B1	B2	B3	B4	Sort Weight
C1	0.230				0.064
C2	0.122				0.034
C3	0.648				0.108
C4		0.614			0.287
C5		0.268			0.125
C6		0.117			0.055
C7			0.582		0.093
C8			0.309		0.050
C9			0.109		0.018
C10				0.095	0.009
C11				0.160	0.015
C12				0.278	0.026
C13				0.467	0.044

3   Conclusions and Recommendations

Lake of irrigation water resource directly leads to reduce agricultural water and restricts to establish the system of agricultural products plant safety; the more serious of water pollution makes water quality can't satisfy the requirements of agricultural irrigation and influences the growth of agricultural products quality. So agricultural planting should stay away from the pollution sources, in order to ensuring the quality of irrigation water source. Organic production technology and regulatory rules should be used gradually to improve water resources environment and make them comply with the agriculture standard; in the production practice. The surrounding water pipe buried should strengthen in order to ensuring the clean water of irrigation.

Strengthen the supervision of agricultural input products. With pesticides and additives for key, the government should put an end to the production and sales illegal pesticides and poisonous and harmful chemicals. Strengthened supervision and inspection which is the quality of pesticide products in production enterprise and standard management of pesticide labels and effective component content should also be emphasized.

Focus on the requirements which are the environment quality of agricultural products whether meet the production safety of agricultural products should be focus on. Analysis the pollution source of agricultural products, based on geographic information system establish regional soil database. Professional evaluation model of regional soil for environmental quality assessment should be used in order to conclude that fit for a regional development of the species of agricultural products.

Compulsory for agricultural planting process setting up production records, in order to laying the foundation of back of agricultural planting security; guide and supervise of quality and safety of agricultural products self inspections through each production

main body, ban unqualified agricultural products on the market, adopt technology of bar code, consumers can rapidly inquire agricultural planting files, So as to realize the rest assured consumption.

References

1. Zhang, H., Sun, X., Liu, Y., Liu, J.: Research on the Feedback System for Quality and Safety of Agricultural Products. J. Hubei Agricultural Sciences (2010)
2. Zhou, Y., Deng, C.: Quality-safety Situation and Development Suggestions of Agricultural Products in Liaoning Province. J. Agricultural Economy (2011)
3. Gao, S.: Problems and CountermEasures in Agricultural Product Quality Security Management System in China. Agricultural University of Henan (2009)
4. Zhang, Y.: Study on Status and Development of Agro-food Quality and Safety Management in Jiangsu Province. University of Yangzhou (2009)

Evaluation of Temporal Resolution Effect in Remote Sensing Based Crop Phenology Detection Studies

Hu Zhao[1,2], Zhengwei Yang[3], Liping Di[2], and Zhiyuan Pei[1]

[1] Agriculture Resource Monitoring Station, Chinese Academy of Agriculture Engineering,
41 Maizidian Street, Chaoyang, Beijing 100125, China
zhaohu@whu.edu.cn, peizhiyuan@tom.com
[2] Center for Spatial Information Science and Systems, George Mason University,
6301 Ivy Lane, Suite 620, Greenbelt, MD 20770, USA
ldi@gmu.edu
[3] National Agricultural Statistics Service, United States Department of Agriculture,
3251 Old Lee Highway, Room 305, Fairfax, VA 22030, USA
Zhengwei_Yang@nass.usda.gov

Abstract. Remote sensing based phenology detection method has been employed to study agriculture, forestry and other vegetations for its potential to reflect the variations in climate change. These studies usually utilized time series Normalized Difference Vegetation Index (NDVI) generated from various sensors through a Maximum Value Compositing (MVC) process, which minimized the contamination from cloud and simultaneously introduce degradation of temporal accuracy. In this study, we assess the impact of temporal resolution on crop phenology derivation researches by comparing three different Moderate Resolution Imaging Spectroradiometer (MODIS) datasets: daily surface reflectance, 8 day composited surface reflectance and 16 day composited NDVI. The surface reflectance data were first filtered by employing auxiliary data which contained quality and viewing geometry information, and then used to calculate NDVI with specific date. A least square method was taken to fit the survival data points to double logistic function. And finally, seven time-related metrics were obtained and matched with field observation crop phenology stages. These remote sensing derivate phenology dates were compared to National Agricultural Statistics Service (NASS) weekly crop progress reports to evaluate the capability of these datasets in temporal sensitive studies. The results illustrated that daily surface reflectance datasets were the most accurate source for time-sensitive studies. However, extra ancillary datum and appropriate denoising techniques should be applied to reconstruct the time series curve. Phenology matching process is a necessary step before detecting phenological information from remote sensing imagery for specific land cover type since same phenological stages of different crop types might have different counterparts on time series curve.

Keywords: Crop phenology, temporal resolution evaluation, least square, double logistic function fitting.

1   Introduction

Remote sensing based phenology detection techniques offer great opportunity for many climate change related researches, such as land surface vegetation dynamics (Beck et al., 2006), agriculture management (Garrigues et al., 2008), forestry monitoring (Ahl et al., 2006) and global warming (Parmesan, 2007). Phenology is the study of the timing of cyclical events in nature such as springtime vegetation budburst or seasonal bird migrations (Badeck et al., 2004), while crop phenology refers to the developments, differentiation and initiation of organs of a crop (Hodges, 1991). Many researchers have studied phenological change using field-collected data, which is prohibitive for large scale monitoring due to the need for repeat data collection in order to monitor change from previous years (Thayn and Price, 2008). The situation has been totally changed after satellite sensor imagery, which record multi-scale (spatial or temporal) information of ground surface, was applied to monitor land vegetation dynamics.

A number of research works have been conducted to monitor 'generic or mixed' vegetation phenological events (Justice *et al.*, 1985; Reed *et al.*, 1994; Zhang *et al.*, 2003; Fisher *et al.*, 2006; White and Nemani, 2006; Stöckli *et al.*, 2008), and many efforts have been made to develop a general procedure for remote sensing based crop phenology detection. Sakamoto et al. (Sakamoto et al., 2005) generated planted, heading and harvest dates of rice paddy in Japan using time series MODIS Enhanced Vegetation Index (EVI) datasets. Vina et al. (Vina et al., 2004) presented a visible atmospherically resistant index (VARI) and accumulated growing degree days (AGDD) to detect maize physiological transitions that undetectable from NDVI. Duchemin et al. (Duchemin et al., 2006) monitored wheat phenology through the analysis of relationship among evapotranspiration, crop coefficients, leaf area index (LAI) and NDVI. Islam and Bala (Islam and Bala, 2008) made use of NDVI and LAI to evaluate potato phenological characteristics by modeling its essential phenological metrics. Each of these pioneering works has its special contribution to remotely-sensed crop phenology estimation research.

However, the majority of above-mentioned studies utilized multi-temporal NDVI/EVI/LAI values processed with the MVC technique (Holben, 1986), which effectively removed the cloud-contaminated data points in original observations series. The maximum value of the composited period (usually range from 8 to 16 days) was assigned to this multiday period rather than the date when this value was captured (Thayn and Price, 2008). This processing step imbedded a temporal error of 8 (or 4) days in average and 15 (or 7) days maximum because a maximum value in 16 (or 8) days is selected and assigned to all of 16 (or 8) days. This could not be neglected in time-sensitive studies, especially for crop phenology detection. Therefore, it is essential to assess the effect of this composited process on crop phenology estimation accuracy and seek out an appropriate strategy for the remote sensing based crop phenology estimation.

The objectives of this paper are to: a) check the feasibility of employing daily surface reflectance datasets and its corresponding ancillary data to generate NDVI curve and consequently calculate crop phenology dates; b) make a quantitative comparison between different temporal resolutions (daily, 8 day and 16 day composited) with fixed spatial scale (250m*250m / pixel) about their effects on phenological stages detection studies; c) attempt to develop a framework for matching 'specific' phenology stages with curve-generated Julian dates. Two crop types (corn and soybean) have been tested on 50 sites distributing over US Corn Belt in this experiment. For each test site, seven Julian dates related potential phenology stages have been generated and matched with specific crop phenological events.

2   Data Description

2.1   Study Sites

In this study, the State of Iowa (range from 40°23′N to 43°30′N and 90°8′W to 96°38′W) in U.S. Corn Belt is selected for experiment because it is a predominant corn and soybean state and it is one of the most important corn and soybean production areas in United States. Iowa locates in the middle of American and its terrain is mainly dominated by plain. All these facts will help us to simplify the study and to keep focus on the method for phenological information extraction from the MODIS products. In this study, 50 sites (25 sites each for corn and soybean) within Iowa State, as shown in figure 1, were selected for method and data experiments.

2.2   Datasets Description and Preprocessing

Three MODIS products have been used in this experiment: Surface Reflectance Daily L2G Global 250m (MOD09GQ), Surface Reflectance 8-Day L3 Global 250m (MOD09Q1) and Vegetation Indices 16-Day L3 Global 250m (MOD13Q1). The MOD09GQ product provides an estimate of the surface spectral reflectance as it would be measured at ground level in the absence of atmospheric scattering or absorption. Low-level data can be corrected for atmospheric gases and aerosols, yielding a level-2 basis for several higher-order gridded level-2 (L2G) and level-3 products. Bands 1 (RED) and 2 (near infrared (NIR)) are provided at a 250-meter resolution in a daily gridded L2G product in the Sinusoidal projection. This product is meant to be used in conjunction with the Surface Reflectance Daily L2G Global 1km and 500m (MOD09GA) where important quality and viewing geometry information is stored. MOD09Q1 is a 8-day gridded level-3 product that stores an estimate of the surface spectral reflectance. Each MOD09Q1 pixel contains the best possible L2G observation during an 8-day period as selected on the basis of high

observation coverage, low view angle, the absence of clouds or cloud shadow, and aerosol loading. MOD09GQ and MOD09Q1 are first calculated NDVI through the following formula: NDVI = (NIR-RED) / (NIR+RED). MOD13Q1, unlike MOD09GQ and MOD09Q1, is designed to provide consistent spatial and temporal comparisons of vegetation conditions. Blue, red, and near-infrared reflectance are used to determine the MODIS daily vegetation indices. The MODIS NDVI complements NOAA's Advanced Very High Resolution Radiometer (AVHRR) NDVI products and provides continuity for time series historical applications. Although MODIS also includes EVI that minimizes canopy background variations and maintains sensitivity over dense vegetation conditions, we only extract NDVI values for comparison purpose (USGS, 2009). All these datasets are captured between March 31 and December 11 from 2006 to 2008. This period includes 256 year days and totally covers the growing season for both corn and soybean in each year according to the statistical information published by United State Department of Agriculture (USDA).
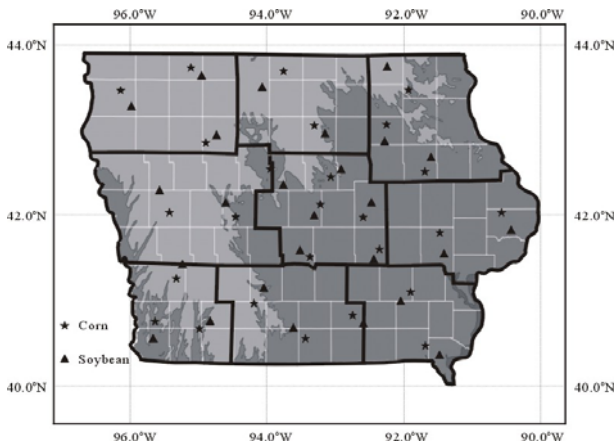
Fig. 1. Distribution of test sites in Iowa State, United States. Filled symbol pentagram (★) and triangle (▲) represented corn and soybean respectively. The background color gradient indicates the general condition of elevation.

Many criterions, besides cloud status information recorded in quality flag, have been presented previously to remove noisy data points affected by thick clouds, gaseous and aerosol scattering as well as bidirectional reflectance distribution function (BRDF) distortion. For example, Sakamoto et al. (Sakamoto et al., 2005) considered the pixels whose blue band reflectance was greater than 10% or whose sensor zenith angle was more than 32.25°should be removed as abnormal data. Points with a random NDVI increase greater than 0.4 during 20 days are also rejected as misleading data (Chen et al., 2004). We adopt these extra criterions to reserve

the pixels in time-series of a test site for above-mentioned three datasets: a) the cloud and cloud shadow status should be 'clear'; b) the blue band value should be lower than 0.1; c) the sensor zenith angle is smaller than 30°and d) difference between survival data point and its Inverse Distance Weighted (IDW) estimation should be small than 0.1(Lu et al., 2007). These standards could ensure that the survival sensor recorded reflectance values are not influenced by cloud or other noises.
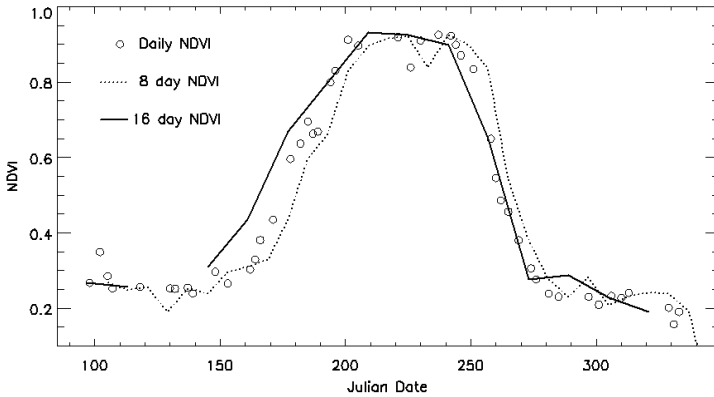
Fig. 2. Time-series datasets of MOD09GQ (open circle ' ○ '), MOD09Q1 (dotted line) and MOD13Q1 (solid line) after preprocessing. Many MOD09GQ cloud-contaminated data points are removed and replaced with NAN so the time series looks like discrete.

The National Agricultural Statistics Service (NASS), the USDA data collection arm,conducts hundreds of surveys every year and prepares weekly crop progress report containing crop planting and emergence percentage estimates. Crop progress data to be used for validation were acquired from the weekly 2006-2008 Crop Progress reports produced by the NASS for Iowa. (These reports are publicly available at http://www.nass.usda.gov/Statistics_by_State/Iowa/Publications/Crop_Progress_&_Co ndition/index.asp.) According to these reports, there are nine districts in Iowa State: northwest (NW), north center (NC), northeast (NE), west center (WC), center (C), east center (EC), southwest (SW), south center (SC) and southeast (SE). Each of these districts has its own validation information (planted, emerged, silking, etc.) for different kinds of crops. Each stage was described as a period as its percentage covers from 0 to 100. Generally, the 50 percent corresponding Julian date of each stage could be selected as final validation information (Wardlow et al., 2006). For corn, the report offered nine phenology stages: planted, emerged, tasseled, silking, milk, dough, dent, mature and harvest. For soybean, there are seven stages: planted, emerged, blooming, setting pods, leaf turning color, mature and harvest. All phenological stage definitions could be accessed at http://www.nass.usda.gov/Charts_and_Maps/Crop_ Progress_ &_Condition/Terms/index.asp. The Cropland Data Layer (CDL) program (http://www. nass.usda.gov/research/Cropland/SARS1a.htm), also offered by NASS, provides the information about what kind of crop has been planted on a specific field with

coordinate information. USDA NASS online products are the main source of our experiment validation information.

3   Methodologies

3.1   Curve Fitting Strategy for Daily Data

After preprocessing steps, the time series are usually not equal-spaced or continuous since some abnormal pixels have been removed (as shown in Figure 2). Meanwhile, even those survival points are still contain some undefined noise (soil background, water content, moisture, saturation, pre-crop vegetation, etc.). Curve fitting is proved to be one of the most effective strategies to eliminate those noises (Bradley et al., 2007).

Various techniques have been implemented to model the survival data points and to reconstruct the removed values, such as Best Index Slope Extraction (BISE) (Viovy et al., 1992), Fourier analysis (Roerink et al., 2000; Geerken et al., 2005), Savitzky–Golay filter and its improvement (Savitzky and Golay, 1964; Chen et al., 2004), Support Vector Machine (Sun et al., 2006), asymmetric Gaussian function (Jönsson and Eklundh, 2002), mean-value iteration (MVI) (Ma and Veroustraete, 2006), wavelet transformation and its improvements (Sakamoto et al., 2005; Lu et al., 2007) and logistic function (Zhang et al., 2003; Jönsson and Eklundh, 2004; Beck et al., 2006; Beck et al., 2007). Hird and McDermid (Hird and McDermid, 2009) made a model-based empirical comparison of some frequently-used NDVI time series denoising techniques and concluded that double logistic function fitting strategy demonstrated the superiority over other methods. The formula could be expressed as equation (1), where $wNDVI$ stands for minimum NDVI value (often arose in winter) of the final curve fitting time series. In this test, $wNDVI$ represents bare soil or pre-crop vegetation NDVI value. This value is often represented as geometric mean of NDVI values in non-growing season (winter or bare soil background)(Beck et al., 2006). According to the NASS crop progress reports; there is no crop or plant available in the corn or soybean fields before April from 2006-2008. This means that we could set the non-growing season as the date before the 120th day of the year. $mNDVI$ denotes the maximum NDVI and its difference with $wNDVI$ represents the amplitude of NDVI time series; mS and S are variables relating to the inflection point in vegetative growing season, while mA and A are referred to senescence season. Their relationship as shown in Figure 3 could be expressed (Beck et al., 2007).

$$NDVI_t = wNDVI + (mNDVI - wNDVI) \times (\frac{1}{1+e^{-mS(t-S)}} + \frac{1}{1+e^{mA(t-A)}} - 1) \qquad (1)$$
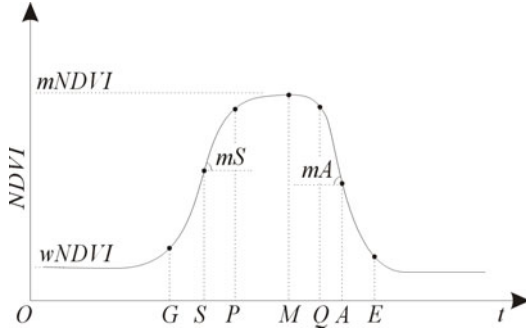
Fig. 3. Example of the double logistic function $NDVI_t$ (equation (1)) used to model the yearly NDVI time series. It is defined by six parameters: the minimum and maximum NDVI ($wNDVI$ and $mNDVI$), two inflection points (S and A), and the rate of increase or decrease in NDVI at S and A, respectively (mS and mA) (Beck et al., 2007). Feature points G, P, Q and E stand for four characteristic curvature values, while M is the point when the NDVI reaches its maximum (or the central points of NDVI saturation period in case of maximum value continues during a period).

Least square method is commonly applied in function fitting to select the 'best fit' parameters (Steinier et al., 1972; Sun et al., 2006; Markwardt, 2008) and often expressed as follows:

$$\chi^2 = \sum_{i=1}^{n} \left[w_i \left(NDVI_i - NDVI_{ti} \right) \right]^2 \tag{2}$$

Where w_i is the weight of ith NDVI value. When χ^2 reaches its minimum, the parameters are optimal. Beck et al. (Beck et al., 2006) presented a two-step schema to determine weight, which was supposed to get the upper envelop of the survival data. First, all weights were set equally as 1 to obtain the reference curve line. In the second step, the weights were reassigned different values according to relative relationship between original values and the reference curve line: if the original value is above (below) the reference curve line, it gets a higher (lower) weight. $NDVI_i$ is the ith originally calculated NDVI value, and $NDVI_{ti}$ is the corresponding estimated value. Since there are contaminated data in daily NDVI, we have made a little adjustment to this schema: set all contaminated points as zero and then applied Beck's two-step weighting strategy. The weights of those contaminated data were set to zero throughout the whole weighting process. This ensures that those contaminated points will not affect the phenology estimation. To get more accurate phenology estimation, the composited products have been linearly interpolated to reconstruct contaminated

daily data with ENVI Interactive Data Language (IDL) routine 'interpol'. However, all interpolated values are weighted as zero during the weighting process to keep the 'original' survival points' affect on resultant NDVI curve.

3.2   Phenology Stage Estimation Based on NDVI Curve Feature Points

As shown in Figure 3, a NDVI curve could be characterized with seven curve feature points: *G, S, P, M, Q, A, E*. The inflection points *S* and *A* could be obtained from the maximum and minimum of the first derivate of estimated NDVI curve. The point *M* represents the date of the maximum NDVI value. If there exist more than one date, *M* will be the central point of these maximum-value dates. Points *S, P, M, Q,* and *A* divided the estimated NDVI curve into four parts. In each of these four parts, the maximum curvature could be obtained and marked as *G, P, Q* and *E* according to the following formula:

$$\kappa = \frac{\left| NDVI_t'' \right|}{\left(1 + \left(NDVI_t' \right)^2 \right)^{3/2}} \tag{3}$$

Julian date t is the independent variable. These seven temporal points will be used to match with actual phenological dates.

After calculating seven feature point dates for each test site, these dates should be matched with statistical stages to identify the relationship between curve-derived dates and specific statistical phenological event dates (treated as actual phenology stages). Before matching, these remote sensing derived dates could be grouped as set G, set S, set P, set M and set Q. We also marked the statistical validation phenology dates (planted, emerged, silking, milk, mature and harvest, etc.) as phenology stage A, stage B, stage C, stage D, etc. By calculating root-mean-square error (RMSE) between each set and stage, an error matrix was produced through formula (4), in which εi,j was the RMSE between ith set $Set_{i,k}$ and jth phenology stage Stagej (includes m test sites, k was the serial number of test sites).

$$\varepsilon_{i,j} = \sqrt{\frac{\sum_{k=1}^{m} \left(Set_{i,k} - Stage_j \right)^2}{m}} \tag{4}$$

All RMSE matrices for nine districts were first computed with test sites that falls into each district, and the state-level RMSE matrix was then obtained by weighting every district matrix. The weight of each district-level matrix is the proportion of district test site numbers to the total test sites. Figure 4 illustrates the whole process of final matching pair generation and error analysis.

Here are three basic principles to determine the intermediate match pairs for each dataset: 1) smaller RMSE, the minimum $\varepsilon_{i,j}$ was selected to determine the first match pair, for example, Setp and Stageq; 2) time line consistency, the sets before (or after) Setp will only be matched with stages before (or after) Stageq, which is proposed to keep the consistency over time line. The final matching pair is generated by considering daily, 8-day and 16-day intermediate matching results, which should abide by the 3) majority rule: first we pick out the matching pairs in which these three datasets have three same intermediate results. Then we choose the matching pairs that two of these three datasets have same matching pairs and the other dataset has different result which will be re-decided same as the former two counterparts. If one stage refers to three different datasets, the result will be determined by the minimum RMSE value. The second rule, time line consistency, is simultaneously applied during this process.
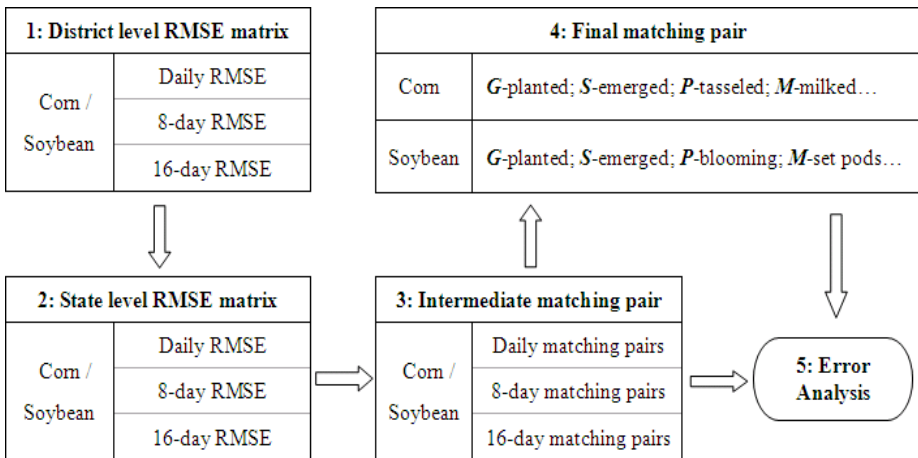
Fig. 4. Flowchart of matching pairs determination and error analysis. It includes 5 steps, step 1: calculating district level RMSE matrix, step 2: integrating 9 district results into state level RMSE matrix; step 3: determining intermediate matching pairs for daily, 8-day and 16-day datasets; step 4: reaching a final matching pair for corn and soybean and step 5: conducting error analysis by using matching pair results in step 3 and 4.

3.3   Evaluation Methods

We used two complementary approaches to evaluate the performance of each data set: phenology matching accuracy and phenology estimation error. First, we define three intermediate matching results and the final matching pair as: Imp01, Imp08, Imp16 and Fmp. They record the matching pairs as phenology "milk" to metric "M'', phenology "mature"   to metric "A'', etc. We interpreted phenology matching accuracy ($\varepsilon_{i,j}$) as ratio between number of final matching pairs (Fmp) and the number of intersection of intermediate matching pairs (Imp01, Imp08 or Imp16) and the final

matching pairs (Fmp). The matching accuracy reflects the degree of matching between intermediate matching results and the final matching pairs, εmi equal to 1 means the intermediate result is the same as the final result. The bigger the εmi is, the worse they coincide with each other. The formula could be expressed as:

$$\varepsilon_{mi} = \frac{N(Fmp)}{N(Fmp \cap Imp_i)} \tag{5}$$

where εmi stands for the phenology matching accuracy of datasets i, N is the function that extract the number of a set. Impi represents Imp01, Imp08 or Imp16 and symbol $\cap$ means the intersection of the two datasets.

The phenology estimation error (εi,j) refers to the summation of final matching pair estimation errors, which is determined by the ratio of RMSE and the mean of the matching pairs' RMSE. For phenology estimation error εi,j, it happened between different datasets and could be listed as the ratio of final matching RMSE value RMSEij to the mean of its counterparts RMSEij (of different datasets). Formula 6 defines phenology estimation error.

$$\varepsilon_{ei} = \sum_{j=0}^{m} \frac{RMSE_{ij}}{\frac{1}{3} \times \sum_{i=01,08,16} RMSE_{ij}} \tag{6}$$

The final error is the summation of matching error and estimation error. And it can be expressed as:

$$\varepsilon_i = a \bullet \sum \varepsilon_{mi} + b \bullet \sum \varepsilon_{ei} \tag{7}$$

Where a and b are the weight for phenology matching error and phenology estimation error. Here we consider these two kinds of error have the same contribution to the final total error, which sets both a and b as 1.0.

4   Results and Discussion

According to the basic rules described in section 3.2, we got the final matching pairs for these three different temporal resolution datasets in 2006-2008 (figure 5). For corn, the final matching pairs in 2007 and 2008 are the same: field observation phenology stage 'planted' corresponds to remote sensing derived first metric 'G'; emerged to 'S'; tasseled to 'P'; milked to 'M'; dough to 'Q'; mature to 'A' and harvest to 'E'. There is no metrics corresponded to phenology stage corn silking and dent in these two years. However, in 2006, the matching pairs seem slightly different: the stage emerged refers to metric 'G' and there is no stage matching with metrics 'S'; and for metric 'A', the correspondent phenological stage is dent, not dough. Other matching pairs are the same with the following two years. According to the statistical observation information obtained from NASS, the maize dough phenology stage is

recorded 5 (South center of IOWA, 2007) to 15 (North center of IOWA, 2007) days before dent stage. When comparing to the RMSE error between remote sensing derived phenological dates and the validation information (in table 1 and 2), the duration between dough and dent is somewhat negligible. This indicates that the matching results are not   varying greatly across different years since phenology changes every year. Comparing with vegetation types, comprehensive human being participation (seed selection, field management, irrigation status, etc.) is the most important factor of crop land cover development, which is sensitive to both natural and manmade microenvironment.
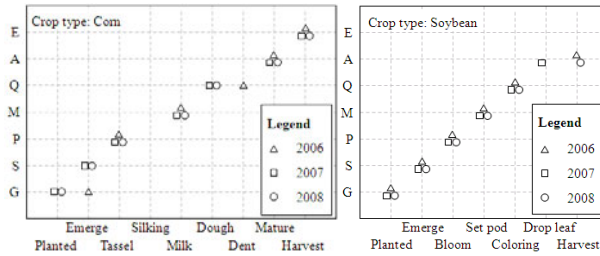
Fig. 5. Final matching pairs for corn (left) and soybean (right) in 2006-2008

Table 1. State level RMSE result for corn in 2007 (days)

	Planted	Emerged	tasseled	Silking	milk	Dough	Dent	Mature	Harvest
01G	7.6	11.3	64. 9	69.4	84.2	96.2	108. 6	127.4	171.0
01S	26.4	16.6	39. 8	44.3	59.1	71.1	83. 5	102.4	144.4
01P	53.1	43.3	15.3	19.3	33.3	45.1	57.3	76.1	118.0
01M	78.0	68.2	15.4	12.3	11.4	21.0	32. 9	51.4	93.2
01Q	99.3	89.4	34.5	30.1	16.3	9.9	15.5	31.4	72.4
01A	133.3	123.4	68.0	63.6	48.9	37.1	25.3	11.5	39.3
01E	127.9	125.7	94. 6	90.1	75.4	63.6	51.6	33. 7	22.8
08G	13. 8	9. 6	55. 8	60.3	75.0	87.0	99.3	118.2	150.2
08S	35.3	25.6	31.8	36.2	50.9	62.8	75.1	94.0	135.9
08P	61.1	51.3	11.5	13.9	26.7	38.1	50.1	68.7	110. 5
08M	85.2	75.4	21. 4	17.4	9.8	15.7	26.5	44.6	86.3
08Q	103.7	93.7	38.7	34.3	20.2	10.6	10.6	26.5	67.7
08A	135.6	125.7	70.2	65.7	51.0	39.1	27.0	10.5	36.2
08E	140.8	134.5	95.1	90.6	75.9	64.0	51.9	33.7	19.7
16G	8.2	14.1	68.7	73.2	88.1	100.1	112.5	131.4	152.0
16S	22.0	12.8	44.4	48.9	63.7	75.8	88.1	107.0	149.0
16P	46. 7	36.9	20.9	25.1	39.6	51.5	63.7	82.5	124.5
16M	76.4	66.6	15.8	13.3	15.4	24.5	54.0	95. 6	35.8
16Q	97.5	87.7	34.0	30.1	17.7	11.5	17.5	34.4	75.3
16A	129.5	119.6	64. 2	59. 7	44.9	33.1	21.2	7.8	42.2
16E	149.4	144.1	88. 6	84.1	69.3	57.3	45.2	27.0	21.2

Table 2. State level RMSE result for soybean in 2007 (days)

	Planted	Emerged	Blooming	Set pods	Coloring	Drop leaf	Harvest
01G	14.5	21.4	60.4	76.9	119.2	128.5	147.5
01S	20. 5	11.0	32.7	49.2	91.6	100.9	120.0
01P	52.2	42.1	13.3	20. 9	60.7	69.8	88.7
01M	76.6	66.3	25.3	11.6	35.0	44.2	63.2
01Q	99.7	89.4	48.4	31. 9	16.9	22.6	40.7
01A	132.6	122.4	81.4	64.8	24.4	19.3	18.9
01E	173.0	148.9	107.9	91.4	49.8	41.0	28.5
08G	14.5	24.3	65.1	81.8	124.4	133.6	152.7
08S	14. 5	4.9	37. 6	54.2	96. 9	106.2	125.3
08P	44.8	34. 7	14.0	26.3	67.6	76. 8	95. 8
08M	71.7	61.5	21. 5	10.6	40.8	49.9	68.8
08Q	93. 7	83.5	43.3	27.8	23.7	31.1	48.9
08A	128.1	117.8	76.9	60.5	22.3	19.3	22.3
08E	129.4	120.0	103.6	87.1	45.8	37.3	28.8
16G	18.7	28. 6	69.4	86.1	128.7	137.9	157.0
16S	12.0	3.0	39.6	56.3	99.0	108.3	127.4
16P	53.1	43.2	20.9	27. 2	61.3	70.3	89.0
16M	73.2	63.2	24.1	15. 2	41.0	49.8	68. 5
16Q	94. 9	84. 7	44.3	28.6	24. 0	29.6	47. 4
16A	126.4	116.2	75.2	58.7	20.1	17.2	21.4
16E	128. 8	141.4	100.3	83. 7	41.8	32.9	22.4

For soybean, year 2006 and 2008 have the same matching pairs: planted to 'G'; emerged to 'S'; blooming to 'P'; setting pods to 'M'; leaves turning color to 'Q' and harvest to 'A'. Remote sensing derived metric 'A' refers to dropping leaf rather than harvest in 2007. The inter-annual matching error often happened at beginning or end of growing season. We believe that, besides the different cultivate conditions mentioned above, the limitation of NDVI at low vegetation density might be responsible for this experiment matching results. Another matter of concern was that, even using the same definition, stage 'harvest' is matched with metric 'E' for corn and 'A' for soybean. This indicates that soybean was harvested when NDVI descended dramatically. This is consistent with the reality: soybean is harvestable after it dries completely (soybean has less than 15% moisture). At this stage, soybean plant turns completely brown. However, when corn reaches harvestable stage, its stem is usually still green. For corn and soybean, the 2007 state level RMSE results are listed in table 1 and 2 respectively, in which normal italic RMSE values (with or without underlines) are intermediate matching pairs for each datasets, while final matching results are highlighted by underlines. The RMSE range for corn is from 7.6 (between planted and 'G', daily surface reflectance) to 25.6 (between emerged and 'S', 8-day composited surface reflectance) days. The average difference between NASS observation and remote sensing results for these three datasets are 13.5, 14.4

and 13.9 days.   For soybean, the RMSE ranges from 3.0 (between emerged and 'S', 16 day composited vegetation product) to 23.9 (between turning color and 'Q', 16 day composited vegetation product) days. And the average difference between statistical phenology stages and NDVI curve based estimates for these datasets are 14.4, 14.5 and 16.5 days. MODIS Daily surface reflectance improves the remote sensing estimation results in 1-2 days when compared with MODIS MVC products.

The total errors for three datasets are displayed in figure 6. Generally the daily surface reflectance product performs better than other two datasets when being applied to crop phenology estimation studies in 2006-2008. The 8-day composited surface reflectance dataset works better than MODIS vegetation production (MOD13Q1) overall except a little irregularity for corn in 2007. Moreover, there is no obvious conclusion could be reached for crop-specific difference between corn and soybean.
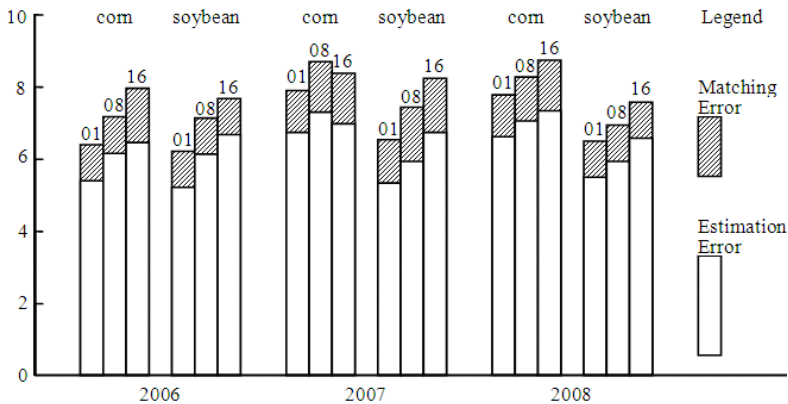
Fig. 6. RMSE result for corn and soybean calculates form different datasets

5   Conclusion

Remote sensing based phenology detection is highly related to the temporal resolution of datasets. We have evaluated the effect of different datum which has different temporal resolution for two crop types: corn and soybean. USDA NASS crop progress reports have been used as ground truth data to assess the performance of different datum. Several conclusions could be drawn from above analysis: 1) Daily surface reflectance dataset could be employed to detect phenology information and had slight advantage to the composited data products, however, extra ancillary data and appropriate denoising and reconstruct techniques should be applied to eliminate the contaminated points. Double logistic function fitting method would be highly recommended; 2) The MVC process brought additional temporal error to final time series trend curve by removing contaminated points. This had a negative effect on time-sensitive remote sensing applications, for example, crop phenology estimation; 3) phenology matching is a necessary step when detecting specific stage for specific

land cover type. Even some phenological stage have the same name, for example, emerged or harvest, they might corresponded to different segments of the final time series curve because of differences in cropping practice. Furthermore, phenology is a multi-factor dependent index to describe plant growth; NDVI has its own limitations to work as the only factor, other factors, like precipitation, temperature and soil moisture, et al. should be considered as part of the phenology model.

References

Ahl, D.E., Gower, S.T., Burrows, S.N., Shabanov, N.V., Myneni, R.B., Knyazikhin, Y.: Monitoring spring canopy phenology of a deciduous broadleaf forest using MODIS. Remote Sensing of Environment 104(1), 88–95 (2006)

Badeck, F.W., Bondeau, A., Bottcher, K., Doktor, D., Lucht, W., Schaber, J., Sitch, S.: Responses of spring phenology to climate change. New Phytologist. 162(2), 295–309 (2004)

Beck, P.S.A., Atzberger, C., Hgoda, K.A., Johansen, B., Skidmore, A.K.: Improved monitoring of vegetation dynamics at very high latitudes: a new method using MODIS NDVI. Remote Sensing of Environment 100(3), 321–334 (2006)

Beck, P.S.A., Jönsson, P., Högda, K.-A., Karlsen, S.R., Eklundh, L., Skidmore, A.K.: A ground-validated NDVI dataset for monitoring vegetation dynamics and mapping phenology in Fennoscandia and the Kola peninsula. International Journal of Remote Sensing 28(19), 4311–4330 (2007)

Bradley, B.A., Jacob, R.W., Hermance, J.F.: A curve fitting procedure to derive inter-annual phenologies from time series of noisy satellite NDVI data. Remote Sensing of Environment 106(2), 137–145 (2007)

Chen, J., Josson, P., Tamura, M., Gu, Z., Matsushita, B., Eklundh, L.: A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky - Golay filter. Remote Sensing of Environment 91(3-4), 332–344 (2004)

Duchemin, B., Hadria, R., Erraki, S., Boulet, G., Maisongrande, P., Chehbouni, A., Escadafal, R., Ezzahar, J., Hoedjes, J.C.B., Kharrou, M.H.: Monitoring wheat phenology and irrigation in Central Morocco: On the use of relationships between evapotranspiration, crops coefficients, leaf area index and remotely-sensed vegetation indices. Agricultural Water Management 79(1), 1–27 (2006)

Fisher, J.I., Mustard, J.F., Vadeboncoeur, M.A.: Green leaf phenology at Landsat resolution: Scaling from the field to the satellite. Remote Sensing of Environment 100(2), 265–279 (2006)

Garrigues, S., Allard, D., Baret, F.: Modeling temporal changes in surface spatial heterogeneity over an agricultural site. Remote Sensing of Environment 112(2), 588–602 (2008)

Geerken, R., Zaitchik, B., Evans, J.P.: Classifying rangeland vegetation type and coverage from NDVI time series using Fourier Filtered Cycle Similarity. International Journal of Remote Sensing 26(24), 5535–5554 (2005)

Hird, J.N., McDermid, G.J.: Noise reduction of NDVI time series: An empirical comparison of selected techniques. Remote Sensing of Environment 113(1), 248–258 (2009)

Hodges, T. (ed.): Predicting crop phenology. CRC Press (1991)

Holben, B.N.: Characteristics of maximum-value composite images from temporal AVHRR data. International Journal of Remote Sensing 7(11), 1417–1434 (1986)

Islam, A.S., Bala, S.K.: Assessment of Potato Phenological Characteristics Using MODIS-Derived NDVI and LAI Information. GIScience & Remote Sensing 45(4), 454–470 (2008)

Jönsson, P., Eklundh, L.: Seasonality extraction by function fitting to time-series of satellite sensor data. IEEE Transactions on Geoscience and Remote Sensing 40(8), 1824–1832 (2002)

Jönsson, P., Eklundh, L.: TIMESAT-A program for analyzing time-series of satellite sensor data. Computers and Geosciences 30(8), 833–845 (2004)

Justice, C.O., Townshend, J.R.G., Holben, B.N., Tucker, C.J.: Analysis of the phenology of global vegetation using meteorological satellite data. International Journal of Remote Sensing 6(8), 1271–1318 (1985)

Lu, X., Liu, R., Liu, J., Liang, S.: Removal of Noise by Wavelet Method to Generate High Quality Temporal Data of Terrestrial MODIS Products. Photogrammetric Engineering and Remote Sensing 73(10), 1129–1139 (2007)

Ma, M., Veroustraete, F.: Reconstructing pathfinder AVHRR land NDVI time-series data for the Northwest of China. Advances in Space Research 37(4), 835–840 (2006)

Markwardt, C.B.: Non-linear Least Squares Fitting in IDL with MPFIT. In: Astronomical Data Analysis Software and Systems XVII ASP Conference Series, vol. XXX (2008)

Parmesan, C.: Influences of species, latitudes and methodologies on estimates of phenological response to global warming. Global Change Biology 13(9), 1860–1872 (2007)

Reed, B.C., Brown, J.F., Vander Zee, D., Loveland, T.R., Merchant, J.W., Ohlen, D.O.: Measuring phenological variability from satellite imagery. Journal of Vegetation Science, 703–714 (1994)

Roerink, G.J., Menenti, M., Verhoef, W.: Reconstructing cloud free NDVI composites using Fourier analysis of time series. International Journal of Remote Sensing 21(9), 1911–1917 (2000)

Sakamoto, T., Yokozawa, M., Toritani, H., Shibayama, M., Ishitsuka, N., Ohno, H.: A crop phenology detection method using time-series MODIS data. Remote Sensing of Environment 96(3-4), 366–374 (2005)

Savitzky, A., Golay, M.J.E.: Smoothing and differentiation of data by simplified least squares procedures. Analytical chemistry 36(8), 1627–1639 (1964)

Steinier, J., Termonia, Y., Deltour, J.: Comments on Smoothing and differentiation of data by simplified least square procedure. Analytical Chemistry 44(11), 1906–1909 (1972)

Stöckli, R., Rutishauser, T., Dragoni, D., O'Keefe, J., Thornton, P.E., Jolly, M., Lu, L., Denning, A.S.: Remote sensing data assimilation for a prognostic phenology mode. Journal of Geophysical Research 113, 1–19 (2008)

Sun, J., Zhou, Y., Bai, Y., Luo, J.: Nonlinear Noise Reduction of Chaotic Time Series Based on Multi-dimensional Recurrent Least Squares Support Vector Machines. In: King, I., Wang, J., Chan, L.-W., Wang, D. (eds.) ICONIP 2006. LNCS, vol. 4232, pp. 900–906. Springer, Heidelberg (2006)

Thayn, J.B., Price, K.P.: Julian dates and introduced temporal error in remote sensing vegetation phenology studies. International Journal of Remote Sensing 29(20), 6045–6049 (2008)

USGS, Land Progresses Distribution Active Archive Center (2009),
       `https://lpdaac.usgs.gov/lpdaac/products/modis_products_table`
       (accessed July 23, 2009)

Vina, A., Gitelson, A.A., Rundquist, D.C., Keydan, G., Leavitt, B., Schepers, J.: Monitoring maize (Zea mays L.) phenology with remote sensing. Agronomy Journal 96(4), 1139–1147 (2004)

Viovy, N., Arino, O., Belward, A.S.: The Best Index Slope Extraction (BISE): A method for reducing noise in NDVI time-series. International Journal of Remote Sensing 13(8), 1585–1590 (1992)

Wardlow, B.D., Kastens, J.H., Egbert, S.L.: Using USDA crop progress data for the evaluation of greenup onset date calculated from MODIS 250-meter data. Photogrammetric Engineering and Remote Sensing 72(11), 1225–1234 (2006)

White, M.A., Nemani, R.R.: Real-time monitoring and short-term forecasting of land surface phenology. Remote Sensing of Environment 104(1), 43–49 (2006)

Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C.F., Gao, F., Reed, B.C., Huete, A.: Monitoring vegetation phenology using MODIS. Remote Sensing of Environment 84(3), 471–475 (2003)

Agricultural Landscape Dynamics and Its Response in Seasonal Vegetation Activities in the Loess Plateau, Northern Shaanxi, China

Zhengguo Li[1,2,*], Peng Yang[1,2], Yanglin Wang[3],
Qingbo Zhou[1,2], Huajun Tang[1,2], and Hsiaofei Chang[4]

[1] Key Laboratory of Resources Remote Sensing and Digital Agriculture, Beijing 100081, China
[2] Institute of Agricultural Resources and Regional Planning,
Chinese Academy of Agricultural Sciences, Beijing 100081, China
[3] Department of Geography, Peking University, Beijing 100871, China
[4] Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences,
P.O. Box 2871, Beijing 100085, China
zhengguoli@caas.net.cn, (yangp,zhouqb,hjtang)@mail.caas.net.cn,
ylwang@urban.pku.edu.cn, lzg.123@263.net

Abstract. The ecological and environmental conditions in semiarid areas are closely linked to landscape dynamics. This study examined the seasonal vegetation activities of landscape classes and dynamics in the Loess Plateau in North Shaanxi Province, China. Landscape dynamics were studied by comparing Landsat images in 1987 and 2007, and classified as landscape transitions, landscape variations and unchanged landscape. The characteristics of seasonal vegetation activities, such as dates for onset-of-greenness, peaks, and onset-of-senescence for landscapes were determined using Moderate Resolution Imaging Spectroradiometer (MODIS) 10-day vegetation index composite products (2002–2009). The landscape dynamics showed that some sloping croplands were converted to woodlands and grasslands. Vegetation activities responded diversely and seasonally to landscape classes and dynamics. Temporal analysis of the Normalized Difference Vegetation Index (NDVI) showed that the time of onset-of-greenness and onset-of-senescence were similar for all major landscape classes, while the time of peak vegetation activity was longer for tree than grass species. The analysis of seasonal vegetation activities and landscape dynamics indicated that seasonal vegetation activities were not only relative to landscape classes, but also affected by landscape dynamics.

Keywords: Agricultural Landscape, Landscape structure, Normalized Difference Vegetation Index (NDVI), The Loess Plateau of China.

1   Introduction

Serious soil erosion in Loess Plateau area of China and the increase in human disturbances in ecosystem have resulted in a range of ecological and land use

---

* Corresponding author.

problems, such as the depletion of land resources and the degradation of eco-environments. An urgent need for agricultural landscape and ecological planning appears to ensure the sustainable development of the Loess Plateau area (Fu et al., 2000). The formulation of policies relating to land use requires adequate understanding on the landscape dynamics. Considerable attention is paid, therefore, to the research and monitoring of the characteristics and trends of landscape and spatial changes (Chen et al., 2001; Fu et al., 2006). The findings are helpful to future landscape development and land use.

In this context, landscape structure, function and its dynamic changes constitute the fundamental contents of landscape ecology (Forman and Godron, 1986; Forman, 1995). Thus, change in spatial patterns is an important aspect of landscape dynamics, which will change landscape function (Turner and Ruscher, 1988; Turner, 1990a; Turner, 1990b). Landscape dynamics is always an active area of landscape ecology research (Li, 2000). Characterizing landscape dynamics and identifying its driving mechanism should be the pre-requisite to understand its ecological impact (Forman and Godron, 1986; Forman, 1995). Since the 1980s many methods have been proposed to quantify the mosaic patterns of landscapes (O'Neill et al., 1988; Turner, 1988; Forman, 1995). Among them, there is considerable interest in comparative landscape metrics with multi-date satellite imagery (Nagaike and Kamitani, 1999; Li, 1999); or investigating the mechanism of landscape dynamics by drive force analysis by which the change rules can be established (Wu et al., 2000) and simulated (Chang and Wu, 1998). However, few efforts have been made to evaluate landscape dynamics according to change intensity or range, to further explore their effect on ecological conditions.

Information on spatial and temporal dynamics of land cover is critical for regional and global change research (Senay and Elliott, 2000). Vegetation cover is an important ecological factor in semiarid and arid areas with great seasonal changes, due to functions such as intercepting rainfall, slowing runoff and preserving soil (Wang and Shao, 2001; Zhang et al., 2004). Research on applications of remotely sensed data for vegetation inventory and monitoring has focused on analysis of vegetation 'greenness' (Schwartz and Reed, 1999; Senay and Elliott, 2000), most often measured using the Normalized Difference Vegetation Index (NDVI). The NDVI is the most used vegetation index and has proved very useful in detecting vegetation change (Elmore et al., 2000), vegetation classification and deriving ecological parameters (Goward et al., 1985; Guo et al., 1999; Sobrino and Raissouni, 2000). The time series of NDVI may be analysed to generate a set of metrics that summarize the phenology of vegetation (Malingreau, 1986; Lloyd, 1990; Reed et al., 1994; Senay and Elliott, 2000; Michael and Graham, 2003).

Over the past 25 years or so, tremendous progress has been made for vegetation characterization by using the data from the Advanced Very High Resolution Radiometer (AVHRR) on board the National Oceanic and Atmospheric Administration (NOAA) meteorological satellites (Loveland et al., 1991). More recently, researchers have favoured higher spatial resolution products for regional scale vegetative studies (Michael and Graham, 2003). With the advent of the TERRA satellite, Moderate Resolution Imaging Spectroradiometer (MODIS) derived

vegetation indices have been used for monitoring temporal changes associated with vegetation, particularly for large regions (Salomonson et al., 2001; Townshend and Justice, 2002; David and Gregory, 2004). In the Loess Plateau area, the agro-climatic system is highly seasonal with dry winters and wet summers. The growing season is defined by moisture availability for plant growth and determinate, obligate annual plant forms (Fu et al., 2000; Fu et al., 2006). Within this system, the time series of NDVI can characterize the temporal extent of the growing season and productive potential of agricultural landscape at the regional scale.

Capability to assess the relationship between landscape dynamics and vegetation seasonal changes is required for recognizing the critical factors and modelling the ecological condition changes in the Loess Plateau area (Xiao and Zhong, 1998; Fu et al., 2000). The vegetation response to landscape dynamics can reveal any benefits or damage to eco-functions of the regional ecosystem (Wang et al., 1999).

The landscape dynamics and its response in seasonal vegetation activities in the Loess Plateau area, China were analysed for a period of 20 years. The specific objectives were: (1) to investigate the dynamic changes in the agricultural landscape and identify the main driving factors; (2) to determine the periods for onset-of-greenness, peaks and onset-of-senescence in each agricultural landscape, (3) to quantitatively and qualitatively assess the sensitivity of NDVI in discriminating landscape dynamics at the regional scale.

2 Study Area and Materials

2.1 Study Area

The study area was the northern part of Shaanxi Province, China between lat 35°21′N–39°34′N and long 107°28′E–111°15′E, in the middle part of the Loess Plateau. The area is 80,606 km^2 and the main rivers include the Kuye, Wuding, Qingqian, Yan, Fenchuan Rivers and part of the Luo River (Fig. 1). The terrain has significant topographic variability with typical Loess hills and gully slopes with elevations of 400–1900 m. The area has a semi-arid continental monsoon climate, with cold winters and warm summers. During the past decades, agriculture has been the primary economic activity and income source in the area, and the most important land use change was the clearing of forestland and grassland for agriculture (Chen et al., 2001). The hilly topography, intensive precipitation and especially long-term extensive human activity (i.e. removal of the natural vegetation and farming-accelerated deterioration) have caused serious soil erosion (Fu et al., 2000; Fu et al., 2006). The present land cover and vegetation types include cropland, shrub, woodland (mainly in southern mountainous areas) and grassland. The main crops are potatoes (Solanum tuberosum), beans (Phaseolus vulgaris), maize (Zea mays) and millet (Panicum miliaceum). The woodland is dominated by introduced vegetation, mainly locust trees (Robinia pseudoacacia). In this area, little leaf peashrub (Caragana microphylla) is the most important shrub species. The grassland is mainly covered by annuals such as wheatgrass (Agropyron cristatum Gaertn.), sweet wormwood (Artemisia annua), annual fleabane (Erigeron annuus) and sandy needlegrass (Stipa glareosa).

N1: Gushanchuan; N2: Kuye river; N3: Tuwei river; N4: Yuxi river;

N5: Jialu river;

W1: Toudaochuan; W2: Upper of Luo river; W3: Inland rivers;

W4: Upper of Wuding river;

M1: Qingjian river; M2: Dali river; M3: Middle of Wuding river;
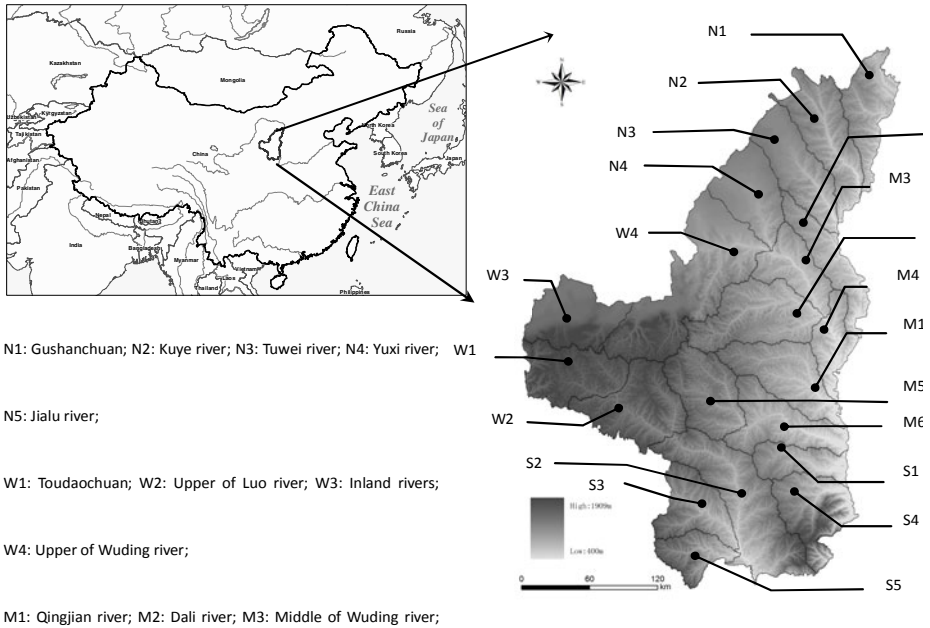
Fig. 1. Geographical location of study area and delineation of watersheds

2.2   Data Used

The landscape class data were derived from a series of Landsat-5 TM images acquired in 1987 and 2007. In total there were 12 images from each of the two time periods with the orbit sequence numbers 126-32, 34, 35 and 36; 127-33, 34, 35 and 36; and 128-34 and 35.

MODIS data were also used. MODIS 10-day vegetation index (VI) composite products in the period 2002–2009 were applied in the analyses. Due to limited availability of validated Landsat data coinciding with MODIS collects, the analysis was limited to 2007–2009, and to the major landscape classes(Table 1).

Table 1. Data source

Data source		Date/resolution
TM images	126-33/34/35/36	1987.07.29, 2007.06.02, with a resolution of 30m
	127-33/34/35	1987.06.02, 2007.06.25, with a resolution of 30m
	128-34/35	1987.05.24, 2007.08.03, with a resolution of 30m
MODIS images	VI products (MOD13)	10-Day composite from 2007 to 2009, with a resolution of 0.5km
Topographic map of Shaanxi province (1:50000)		Published by National Geomatics Center of China
Land use map of Shaanxi province		Made in 2000 by Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences

3   Methods

3.1   Data Pre-processing

Data pre-processing used remotely-sensed image processing software packages—ENVI 4.0, developed by RSI Company, and the geographical information system software ArcGIS 9.0, by ESRI Company. Prior to analysis, following atmospheric radiance calibration and initial geometric rectification of all the TM images using ENVI, a number of reference points were selected from a 1:50,000 scale topographic map. The TM images were rectified to a Gauss Kruger projection with a pixel resolution of 30 m × 30 m, by using nearest neighbour rules provided by ENVI, to ensure that the error was controlled in less than one-half pixel root mean square error. The MODIS products were also rectified to the Gauss Kruger projection by the above method, and resampled to 500 m × 500 m pixel resolution. Finally, the data was masked so that the analysis was limited to the administrative area of the northern Shaanxi Province.

The MODIS VI products included maximum value composites of the NDVI and enhanced VI, as well as the corresponding red and near-infrared reflectance (band 1 and band 2) and quality assessment (QA) flags (Huete et al., 2002). Only NDVI values from the MODIS composite products were used, along with the VI Usefulness Index in the QA data set. Reflectance values of NDVI for dates with a Usefulness Index value lower than "good quality" were replaced by linearly interpolated values from the two closest dates with good, high or perfect quality. Finally, an iterative median smoother was used to retain peaks and eliminate valleys of the temporal NDVI curve.

3.2   Classification of the 1987 and 2007 Landsat TM Images

Landscape classification and mapping with remotely sensed data is a common technique, mainly based on the ability to reflect land surface and human activity information (Ichoku and Karnieli, 1996; Metternicht and Fermont, 1998). We used supervised maximum likelihood classification methods. The accuracy of the 1987 and 2007 images was assessed using ground data points not used in the classification process. The equal control point methods in ENVI were used with at least 30 points for each class. The overall accuracy assessment was checked separately for each image and accepted if the accuracy was > 80% (determined with 580 random checkpoints). After accuracy assessment, all images were clumped and vectorized using ENVI. These coverages were pre-processed to eliminate areas < 0.27 ha (corresponding to 3 pixel × 1 pixel) using ArcGIS for faster spatial analysis.

Landsat TM images (1987) were successfully classified for 11 landscape classes; settlements and sparse woodland were classified with a lower accuracy than other classes. However, this was acceptable as the overall classification accuracy is much higher (87%) with the Kappa statistics value of 0.92. The classification of images

(2007) was also acceptable due to a higher overall classification accuracy (89%) and Kappa values of 0.93. Three main agricultural landscape classes were identified: cropland, woodland and grassland. A further subclassification of agricultural landscape was defined using the percentage of vegetation canopy and terrain characteristics (Table 2).

Table 2. Descriptions of landscape classes

Landscape classes	Landscape subclasses	Description
Cropland landscape	irrigated cropland (Cropland-I)	cropland less than 15° (AD15)
	terrace cropland (Cropland-T)	slope cropland more than 15° (AD15) and less than 25°(AD25)
	sloping cropland (Cropland-S)	slope cropland more than 25° (AD25)
Grassland landscape	dense grassland (Grassland-D)	75–100% grass cover
	moderate dense grassland (Grassland-M)	50-75% grass cover
	sparse grassland (Grassland-S)	< 50% grass cover
Woodland landscape	dense woodland (Woodland-D)	70–100% canopy cover
	moderate dense woodland (Woodland-M)	50-75% canopy cover
	sparse woodland (Woodland-S)	< 50% canopy cover
Residential landscape	settlements	residential areas
Water landscape	wetland	natural and artificial lake, river

3.3   Classifications of Landscape Dynamics

According to the intensity or range of landscape changes, landscape dynamics were simplified to three typical landscape changes. First, the original landscape was replaced by a different landscape, causing transition of landscape classes. Second, the landscape class was unchanged while landscape subclass changed with varying landscape characteristics. Third, the landscape was unchanged because neither landscape class nor subclass changed.

The complexity of landscape dynamics was based on the landscape classification results in 1987 and 2007. The landscape classification scheme was defined as unchanged landscape, landscape variation and landscape transition. In detail; unchanged landscape meant the class and subclass of landscape remained as original; transition meant a shift of landscape class; and variation meant landscape class changed but subclass remained unchanged(Table 3).

Table 3. Classifications of landscape dynamics

Classes		Cropland Landscape			Woodland Landscape			Grassland Landscape		
Subclasses		Cropland-I	Cropland-T	Cropland-S	Woodland-D	Woodland-M	Woodland-S	Grassland-D	Grassland-M	Grassland-S
Cropland landscape	Cropland-I	☆								
	Cropland-T	◇	☆							
	Cropland-S	◇	◇	☆						
Woodland landscape	Woodland-D	△	△	△	☆					
	Woodland-M	△	△	△	◇	☆				
	Woodland-S	△	△	△	◇	◇	☆			
Grassland landscape	Grassland-D	△	△	△	△	△	△	☆		
	Grassland-M	△	△	△	△	△	△	◇	☆	
	Grassland-S	△	△	△	△	△	△	◇	◇	☆

Notes: ☆ unchanged landscape, ◇ landscape variation, △ landscape transition.

3.4   Characteristics of Seasonal Vegetation Activities

A preliminary study on individual years indicated that although the relative magnitudes of the NDVI differed from year to year, partly due to different view angles and atmospheric conditions, the relative time periods for major vegetation activities could still be inferred (Senay and Elliott, 2000). It was anticipated that averaging of many years of data could eliminate sudden and unexplainable irregularities in the NDVI data (Michael and Graham, 2003). Thus, the source data were from the 3-year average 10-day maximum value MODIS-NDVI composites (2007–2009).

The key to defining seasonal characteristics is to identify the onset of the growing season. The method used was visual inspection of the seasonal NDVI curves accompanied by an evaluation of the difference between adjacent NDVI values (Senay and Elliott, 2000). Once the onset of the growing season was identified, additional seasonal characteristics were derived (Table 2). There were more than one local maximum NDVI values; two local maxima occurred, in the beginning and later part of the growing season. There were two families of seasonal characteristics: temporal and NDVI value. Temporal characteristics are time of onset of growing season (OnT), time of end of growing season (EndT), time of the first maximum NDVI (MaxT-1), and time of the second maximum NDVI (MaxT-2). NDVI characteristics are NDVI value at onset of growing season (OnV), NDVI value at end of growing season (EndV), the first maximum NDVI value (MaxV-1), and the second maximum NDVI value (MaxV-2).

Table 4. Description of NDVI metrics calculated for the period 2007-2009

Abbreviation	Definition	Metric
OnT	The period after which there was an increase of NDVI values from a seasonally low value	Starting date of NDVI high period (onset-of-greenness)
OnV	Value of NDVI at forwards intersection	NDVI at start of high period
MaxT-1	Time of the first maximum raw corrected NDVI	Date of the first maximum NDVI
MaxV-1	The first Maximum value of NDVI	The first maximum of NDVI
MaxT-2	Time of the second maximum raw corrected NDVI	Date of the second maximum NDVI
MaxV-2	The second Maximum value of NDVI	The second maximum of NDVI
EndT	the period after which a downward trend in NDVI began	End date of NDVI high period
EndV	Value of NDVI at backwards intersection	NDVI at end of high period (onset-of- senescence)

4   Results

4.1   Changes of the Landscape Pattern between 1987 and 2004

According to the classified Landsat images between 1987 and 2007, landscape changes have occurred in composition and structure of the Loess Plateau area (Table 5). Cropland was the dominant landscape class in this area. The proportion of

Table 5. Landscape structure compositions in study area (1987-2007)

Landscape type	Area (km^2)			Ratio (%)		
	1987	2004	Difference	1987	2004	Difference
Cropland-I	322.4	403.0	80.6	0.4	0.5	0.1
Cropland-T	24504.2	23375.7	-1128.5	30.4	29.0	-1.4
Cropland-S	5803.6	5723.0	-80.6	7.2	7.1	-0.2
Woodland-D	2015.2	2095.8	80.6	2.5	2.6	0.1
Woodland-M	7173.9	7335.1	161.2	8.9	9.1	0.2
Woodland-S	2579.4	2498.8	-80.6	3.2	3.1	-0.1
Grassland-D	2257.0	2095.8	-161.2	2.8	2.6	-0.2
Grassland-M	22730.9	23053.3	322.4	28.2	28.6	0.4
Grassland-S	12171.5	12977.6	806.1	15.1	16.1	1.0
settlements	322.4	806.1	483.6	0.4	1.0	0.6
wetland	725.5	241.8	-483.6	0.9	0.3	-0.6

cropland landscape decreased from 38 to 36.6%; especially, there was a net decline of 1128.5 km^2 in terrace cropland during 1987–2007. Grassland was also a significant landscape class, its area increased from 46.1 to 47.3%, with a net increase of 806.1 km^2 in sparse grassland. Overall, a large amount of cropland was transformed into woodlands and grasslands. Hence, over 1987–2004, there was a decrease in cropland and an increase in grassland and woodland.

4.2   Landscape Dynamics and Landscape Classes

Landscape dynamics among the major landscape classes, included landscape transition, variation and unchanged, were determined based on the classification of Landsat images. Certain landscape transitions or variations were observed and evaluated for each class (Table 6). For example, nearly 80.9 km^2 (25.1%) of irrigated cropland was converted into other landscape classes; however, there were no landscape variations. A large amount of terrace cropland (882.2 km^2) was transformed into grassland or woodland landscape, and 318.6 km^2 of terrace cropland developed into irrigated cropland or degraded to sloping cropland. Of sloping cropland, a net area of 510.7 km^2 shifted into grassland or woodland, and 116.1 km^2 upgraded into irrigated or terrace cropland.

Of woodland landscape, the unchanged dense woodland was 1954.7 km^2 (97.0%), only 30.2 km^2 was converted into other landscape classes, and a similar area of dense woodland degraded into medium or low coverage woodland. Most moderate dense woodland remained unchanged (95.6%) and 1.0% changed into other woodlands, while 243.9 km^2 (3.4%) shifted into other landscape classes. Of sparse woodland, there was an area of 123.8 km^2 in landscape transitions and 15.5 km^2 in variations.

Table 6. Landscape dynamics of landscape types (1987-2007)

Landscape type	Area (km^2)			Ratio (%)		
	Unchanged landscape	landscape variation	Landscape transition	Unchanged landscape	landscape variation	Landscape transition
Cropland-I	241.5	0.0	80.9	74.9	0.0	25.1
Cropland-T	23303.5	318.6	882.2	95.1	1.3	3.6
Cropland-S	5176.8	116.1	510.7	89.2	2.0	8.8
Woodland-D	1954.7	30.2	30.2	97.0	1.5	1.5
Woodland-M	6858.2	71.7	243.9	95.6	1.0	3.4
Woodland-S	2440.1	15.5	123.8	94.6	0.6	4.8
Grassland-D	2011.0	2.3	243.8	89.1	0.1	10.8
Grassland-M	18116.5	3977.9	636.5	79.7	17.5	2.8
Grassland-S	7059.5	3700.1	1411.9	58.0	30.4	11.6

Of grassland landscape, 2011 km^2 (89.1%) of dense grassland remained unchanged, 243.8 km^2 (10.8%) was converted to cropland and 2.3 km^2 (0.1%) was degraded. A large amount of moderate dense grassland 3977.9 km2 (17.5%) was heavily degraded into low coverage, and 636.5 km^2 (2.8%) changed into other landscapes. As for sparse grassland, an area of 1411.9 km^2 was in landscape transitions and 3700.1 km^2 in variations.

4.3   Landscape Dynamics and Spatial Distribution

The analysis of landscape composition in each watershed showed that the proportion of cropland landscape decreased gradually from north to south in the Loess Plateau area, while woodland and grassland obviously increased. Further landscape dynamics in watersheds were subsequently evaluated.

Table 7. Composition of landscape dynamics in watersheds

Watersheds	Ratio (%)		
	Unchanged landscape	landscape variation	Landscape transition
Gushan Chuan	98.8	0.2	1.0
Kuye River	86.9	7.4	5.7
Tuwei River	78.5	6.2	15.3
Yuxi River	82.1	4.6	13.3
Wuding River (upper)	72.4	18.2	9.4
Wuding River (middle)	79.4	19.4	1.2
Wuding River (lower)	94.3	1.5	4.2
Jialu RIver	83.9	13.1	3.0
Dali River	83.2	12.6	4.2
Toudao Chuan	60.6	36.9	2.5
Qingqian River	84.4	11.6	4.0
Yan River (upper)	81.1	17.8	1.1
Yan River (lower)	86.8	9.3	3.9
Luo River (upper)	82.1	17.1	0.8
Luo River (middle)	89.1	4.1	6.8
Yunyan River	98.2	0.1	1.7
Hulu River	92.4	4.4	3.2
Shiwang River	96.2	1.0	2.8
Ju River	97.7	1.2	1.1
Continental watersheds	76.7	4.1	19.2

The landscape classes in the north, where the classes were cropland and grassland, such as in the watersheds of Gushan Chuan River, the Kuye River and the middle reaches of the Wuding River, generally did not change (Table 7). In the middle area, the variations within landscape classes were significant; in Toudao Chuan River (36.9%), the upper (18.2%) and middle (19.5%) reaches of the Wuding River, the upper (17.8%) reaches of the Yan River and the upper (17.1%) reaches of the Luo River, where the dominant landscape classes were dense and moderate dense grassland. Landscape transitions occurred mainly in the northwest, such as the continental watersheds (19.2%), Tuwei (15.3%) and Yuxi Rivers (13.3%), where the landscape classes were mainly sparse grassland and sloping cropland. In the south, where woodland and grassland were the major landscape classes, almost no landscape transition took place; such as the watersheds of the upper reaches of the Luo (0.8%), Ju (1.1%), Yunyan (1.7%), Shiwang (2.8%) and Hulu (3.2%) Rivers.

4.4   Seasonal Vegetation Activities of Landscape Subclasses

Based on temporal plots of the 3-year average NDVI, the seasonal vegetation activities (onset, peak and senescence) of the various landscape subclasses were determined. For each landscape subclass, onset-of-greenness usually occurred about 15 April (Table 8). The peak-of-greenness occurred at different periods for different subclasses. The irrigated cropland was the first cropland to peak at the end of May; this was expected since this subclass was dominated by winter wheat. MaxT-1 occurred on approximately 15 July for terrace and sloping cropland, sparse woodland, moderate dense and sparse grassland; for dense woodland and grassland, moderately dense woodland, it occurred around 15 May. MaxT-2 was about 5 September for all cropland, sparse woodland, moderate dense and sparse grassland; and on about 25 September for dense and moderate dense woodland and dense grassland. Onset-of-senescence was about 25 October for all the subclasses.

Table 8. Summary of vegetation activity based on MODIS data (2007-2009)

Landscape class	OnT	OnV	MaxT-1	MaxV-1	MaxT-2	MaxV-2	EndT	EndV
Cropland-I	4/15	13.5	5/25	41.2	9/5	55.3	10/25	48.3
Cropland-T	4/15	6.1	7/15	34.5	9/5	40.3	10/25	30.9
Cropland-S	4/15	8.6	7/15	39.6	9/5	43.8	10/25	33.8
Woodland-D	4/15	18.3	5/15	62.2	9/25	61.2	10/25	57.3
Woodland-M	4/15	14.2	5/15	47.5	9/25	53.1	10/25	49.7
Woodland-S	4/15	11.9	7/15	50.6	9/5	50.5	10/25	44.2
Grassland-D	4/15	16.3	5/15	58.6	9/25	59.8	10/25	56.6
Grassland-M	4/15	8.6	7/15	39.9	9/5	43.5	10/25	36.0
Grassland-S	4/15	7.5	7/15	33.7	9/5	38.3	10/25	30.0

Landscape subclasses are compared in Table 8. All landscape subclasses in the Loess Plateau area had a high degree of variability in NDVI. Average NDVI values ranged from 6.1 for terrace cropland to 18.3 for dense woodland at the OnT. Furthermore, NDVI ranged from 33.7 and 38.3 for sparse grassland to 62.2 and 61.2 for dense woodland at the MaxT-1 and MaxT-2, respectively. Similarly, the NDVI values varied from 30.0 to 57.3 at the EndT.

4.5   Seasonal Vegetation Activity of Landscape Dynamics

The temporal plots of NDVI showed that vegetation activities responded diversely and seasonally to landscape classes and dynamics. As for unchanged landscape,
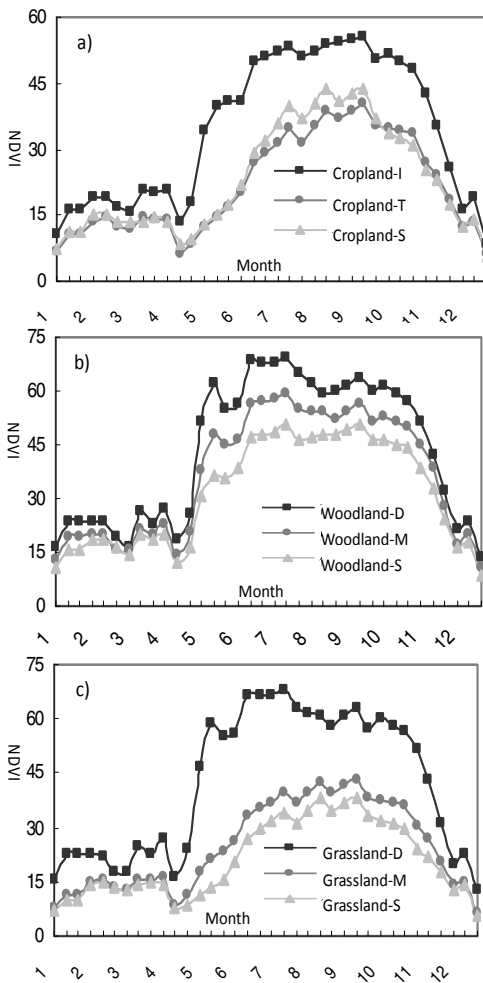
Fig. 2. vegetation seasonal activities of unchanged landscape

landscape subclasses in the Loess Plateau area could be categorized into two general groups of seasonal vegetation activities. The first group was irrigated cropland (Fig. 2a), dense and moderate dense woodland (Fig. 2b) and moderate dense grassland (Fig. 2c), which had an extended period between MaxT-1 and MaxT-2. The group could be influenced by the physiology of trees. The dense grassland and irrigated cropland subclasses also contained significant tree species inclusions (i.e. the MODIS pixels in a predominantly pasture or crop area might include signals from nearby forest). The second group was generally dominated by grass species. Sloping and terrace cropland (Fig. 2a), moderate dense and sparse grassland (Fig. 2c) had a relatively short period between MaxT-1 and MaxT-2. This could be due to these subclasses being mainly distributed in the northwest of the Loess Plateau area where rainfall is relatively low with a semi-arid climate.
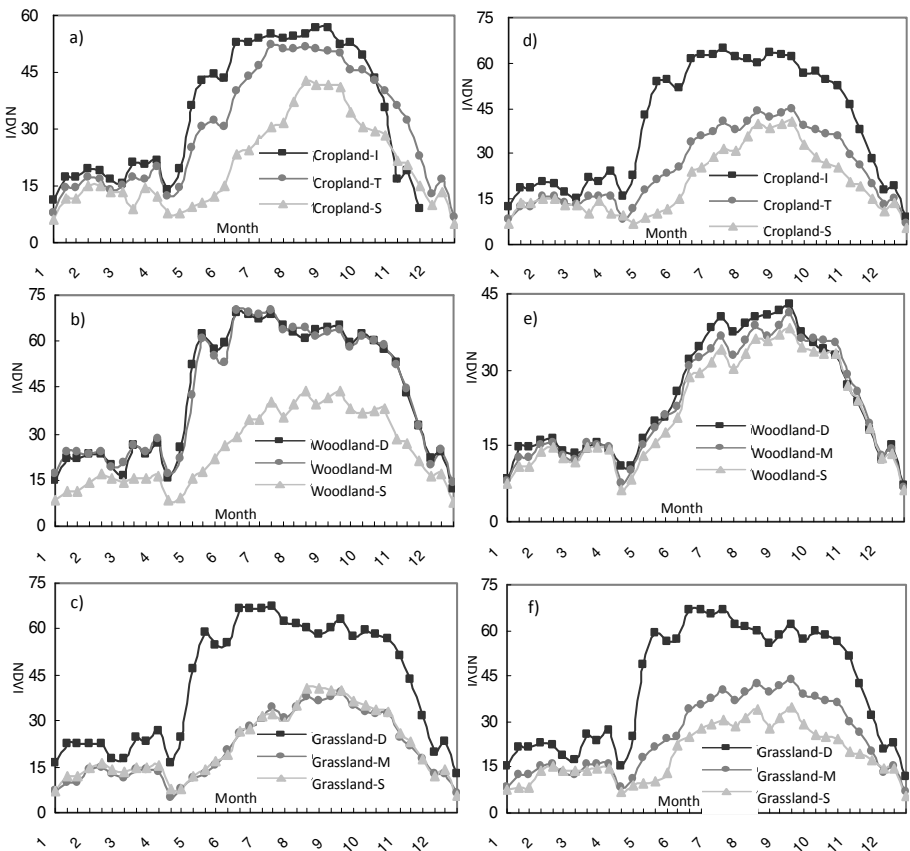
Fig. 3. Vegetation seasonal activities of landscape variations (left) and transitions (right)

Similar rules could be found in landscape variations. Differing to other croplands, the seasonal vegetation activities in irrigated cropland were characterized by higher maximum NDVI values at MaxT-1 and MaxT-2, shorter renewal period between OnT and MaxT-1, much more extended growing period between MaxT-1 and MaxT-2 (Fig. 3a). Similarly, the dense and moderate dense woodlands (Fig. 3b) and dense grasslands (Fig. 3c) had almost the same characteristics as each other of vegetation seasonal activities, while the sparse woodlands (Fig. 3b), moderate dense and sparse grasslands (Fig. 3c) had the opposite features with lower maximum NDVI values at MaxT-1 and MaxT-2, much longer renewal period between OnT and MaxT-1, shorter growing period between MaxT-1 and MaxT-2

Compared with irrigated cropland in landscape variations, there was a similar temporal pattern of seasonal vegetation activities in landscape transitions (Fig. 3d). However, the terrace and sloping croplands had lower maximum values of NDVI, longer renewal periods and shorter growing periods (Fig. 3d). Especially, MaxT-1 for these two subclasses was two months later than for irrigated cropland. At the same time, all woodlands converted from other landscape classes had much lower maximum NDVI values and shorter growing seasons than other woodlands (Fig. 3e). All dense grasslands in various landscape dynamics had high maximum values of NDVI and extended growing seasons (Fig. 3f), while other grasslands in landscape transitions and variations had the low maximum and short growing season

5   Discussion

The dynamics of landscape classes were identified by repeated satellite images coupled with GIS analyses of 1987 and 2007. The proportion of terrace and sloping cropland decreased. The proportion of moderate dense and sparse grassland, the dense and moderate dense woodland landscape increased. The quantitative evidence showed that some sloping croplands were converted to woodlands and grasslands in the past decades. Presumably, this is due to structural adjustment of the agricultural sector, the initiation of Chinese Family Planning Policy, the increase in regional income in recent years and implementation of large-scale eco-environment conservation programmes (e.g. the Grain-for-Green Programme in 1999). From an ecological viewpoint, the present landscape structure and pattern is more sustainable than the former.

In semiarid and arid areas, vegetation has significant ecological functions, such as soil and water conservation and providing habitats for all life forms. Vegetation activities (onset, peaks and senescence) could directly reflect soil resistance to erosion and the temporal pattern of water use. Most of all, the identification of two growing season peaks could be important to identify water use patterns of different agricultural landscape types. The period between the second peak and senescence could signify a gradual decrease in water use, while the period after senescence a rapid decline in evapotranspiration (Senay and Elliott, 2000).

The findings suggest that the spatial correspondence of MODIS and Landsat data sets was robust, and the combination of the two data sets to characterize spatial heterogeneity and temporal dynamics of vegetation was justified. This research has

shown that MODIS NDVI data has the potential to determine periods of important vegetative activities and relative 'greenness' magnitudes that can influence major energy and mass flux processes at regional scales.

6   Conclusion

On the regional scale, landscape compositions in the northern watersheds were mainly grassland and mixed cropland–grassland landscape, while in the southern watershed, landscape compositions mainly included mixed cropland–woodland and mixed woodland–grassland. Correspondingly, the vegetation cover ranged from low to high, which reflected the spatial pattern of vegetation cover resulting from natural forces and human activities on the macro regional scale.

Using a 3-year average of MODIS NDVI data, the temporal analysis of NDVI values showed that the onset-of-greenness (middle April) and onset-of-senescence (late October) were at similar times for all major landscape classes. The main difference among the classes was the NDVI magnitude in the period between onset-of-greenness and onset-of-senescence. Generally, the length of time for the peak vegetation activity (i.e. MaxT-1 to MaxT-2) was longer for tree than grass species.

The ranking of landscape subclasses in seasonal vegetation activities indicated that moderate dense grassland had landscape subclasses in two general groups. It is important to note that the moderate dense grassland was by far the dominant landscape in almost all watersheds. Landscape subclasses with seasonal average NDVI values less than moderate dense grassland included: all croplands, sparse woodland and sparse grassland; landscape subclasses with NDVI values larger than included dense and moderately dense woodland and dense grassland.

Taking landscape dynamics into account, the peak vegetation activity in unchanged landscapes was from high to low in the following sequence: dense and moderate dense woodland > dense grassland > sparse woodland and irrigated cropland > moderate dense grassland > terrace cropland > sloping cropland and sparse grassland. For landscape variations, the sequence was: dense and moderate dense woodland and irrigated cropland > dense grassland, sparse woodland and terrace cropland > moderate dense grassland > sparse grassland and sloping cropland. For landscape transitions, the sequence was: irrigated cropland > dense grassland > dense woodland, terrace cropland and moderate dense grassland > sloping cropland > moderate dense, sparse woodland and sparse grassland.

By analysing the relationships among the seasonal vegetation activities, landscape types and dynamics, it was evident that seasonal vegetation activities were not only relative to landscape classes, but also affected by landscape dynamics (e.g. landscape transitions and variations). Therefore, more attention should be paid to landscape spatial structure and temporal dynamics for the ecological recovery and construction in the Loess Plateau of North Shaanxi.

References

Chang, X.L., Wu, J.G.: Spatial analysis of pattern of sandy landscape in Kerqin, inner Mongolia. Acta Ecologica Sinica 18(3), 225–230 (1998) (in Chinese, with English abstract)

Chen, L.D., Wang, J., Fu, B.J., Qiu, Y.: Land use change in a small catchment of northern Loess Plateau, China. Agriculture Ecosystem Environment 86, 163–172 (2001)

David, B.L., Gregory, P.A.: Cropland distributions from temporal unmixing of MODIS data. Remote Sensing of Environment 93(1), 412–422 (2004)

Elmore, A.J., Mustard, J.F., Manning, S.J., Lobell, D.B.: Quantifying vegetation change in semiarid environments: Precision and accuracy of spectral mixture analysis and the Normalized Difference Vegetation Index. Remote Sensing of Environment 73(1), 87–102 (2000)

Forman, R.T.: Land Mosaics-The Ecology of Landscapes and Regions. Cambridge University Press, Cambridge (1995)

Forman, R.T., Gordron, M.: Landscape Ecology. Wiley, New York (1986)

Fu, B.J., Chen, L.D.: Agricultural landscape spatial pattern analysis in the semi-arid area of the Loess Plateau, China. Journal of Arid Environment 44, 291–303 (2000)

Fu, B.J., Zhang, Q.J., Chen, L.D., Zhao, W.W., Gulinck, H., Liu, G.B., Yang, Q.K., Zhu, Y.: Temporal change in land use and its relationship to slope degree and soil type in a small catchment on the Loess Plateau of China. Catena 65(1), 41–48 (2006)

Goward, S.N., Tucker, C.J., Dye, D.J.: North American vegetation patterns observed with the NOAA-7 advanced very high resolution radiometer. Vegetation 64, 3–14 (1985)

Huete, A., Didan, K., Miura, T., Rodriguez, E., Gao, X., Ferreira, L.: Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment 83(1-2), 195–213 (2002)

Ichoku, C., Karnieli, A.: A review of mixture modeling techniques for sub pixel land cover estimation. Remote Sensing Review 13, 161–186 (1996)

Lloyd, D.: A phenological classification of terrestrial vegetation cover using shortwave vegetation index imagery. International Journal of Remote Sensing 11, 2269–2279 (1990)

Li, X.Z.: Assessment of land use change using GIS: A case study in the Uanos de Orinoco. Wagemigen University Press, Wagemigen (1999)

Li, X.Z.: The focus and front of landscape ecology development from the review of 15th USA annual landscape ecology conference. Acta Ecologica Sinica 20(6), 1113–1115 (2000) (in Chinese, with English abstract)

Loveland, T.R., Merchant, J.W., Ohlen, D.O., Brown, J.F.: Development of a land-cover characteristics database for the conterminous US. Photogrammetric Engineering and Remote Sensing 57, 1453–1463 (1991)

Metternicht, G., Fermont, A.: Estimating erosion surface features by linear mixture modeling. Remote Sensing Environment 64, 254–265 (1998)

Malingreau, J.P.: Global vegetation dynamics: satellite observations over Asia. International Journal of Remote Sensing 9, 1121–1146 (1986)

Michael, J.H., Graham, E.D.: Estimating spatio-temporal patterns of agricultural productivity in fragmented landscapes using AVHRR NDVI time series. Remote Sensing Environment 84, 367–384 (2003)

Nagaike, T., Kamitani, T.: Factors affecting changes in landscape diversity in rural areas of the Fagus crenata forest region of central Japan. Landscape Urban Planning 43, 209–216 (1999)

O'Neill, R.V., Hunsaker, C.T., Timmins, S.P., Jackson, B.L., Jones, K.B., Riitters, K.H., Wickham, J.D.: Scale problems in reporting landscape pattern at the regional scale. Landscape Ecology 11, 169–180 (1996)

Reed, B.C., Brown, J.F., Vander Zee, D., Loveland, T.R., Merchant, J.W., Ohlen, D.O.: Measuring phenological variability from satellite imagery. Journal of Vegetation Science 5, 703–714 (1994)

Salmonson, V.V., Guenther, B., Masuoka, E.: A summary of the status of the EOS terra mission Moderate Resolution Imaging Spectroradiometer (MODIS) and attendant data product development after one year of on-orbit performance. In: IEEE 2001 International Geoscience and Remote Sensing Symposium, CDROM, July 9 -13 (2001)

Schwartz, M.D., Reed, B.C.: Surface phenology and satellite sensor-derived onset of greenness: an initial comparison. International Journal of Remote Sensing 20(17), 3451–3457 (1999)

Senay, G.B., Elliott, R.L.: Combining AVHRR-NDVI and landuse data to describe temporal and spatial dynamics of vegetation. Forest Ecology and Management 128(1), 83–91 (2000)

Sobrino, J.A., Raissouni, N.: Toward remote sensing methods for land cover dynamic monitoring: application to Morocco. International Journal of Remote Sensing 21(2), 353–366 (2000)

Townshend, J., Justice, C.: Towards operational monitoring of terrestrial systems by moderate-resolution remote sensing. Remote Sensing of Environment 83(1-2), 351–359 (2002)

Tunner, M.G.: Landscape changes in nine rural countries in Georgia. Photogrammetric Engineering and Remote Sensing 56(3), 379–386 (1990a)

Turner, M.G.: Spatial and temporal analysis of landscape patterns. Landscape Ecology 4(1), 21–30 (1990b)

Turner, M.G., Ruscher, C.L.: Changes in landscape patterns in Georgia, USA. Landscape Ecology 1(4), 241–251 (1988)

Wang, Y.L., Zhao, Y.B., Han, D.: The spatial structure of landscape eco-systems: concept, indices and case studies. Advance in Earth Sciences 14(3), 235–241 (1999) (in Chinese, with English abstract)

Wu, J.G., Jelinski, D.E., Luck, M., Tueller, P.T.: Multiscale Analysis of Landscape Heterogeneity: Scale Variance and Pattern Metrics. Geographic Information Sciences 6(1), 6–19 (2000)

Xiao, D.N., Zhong, L.S.: Ecological Principles for landscape classification and evaluation. Journal of Applied Ecology 9(2), 217–221 (1998); (in Chinese, with English abstract)

Zhang, Q.J., Fu, B.J., Chen, L.D., Zhao, W.W., Yang, Q.K., Liu, G.B., Gulinck, H.: Dynamics and driving factors of agricultural landscape in the semiarid hilly area of the Loess Plateau, China. Agriculture Ecosystem Environment 103, 535–543 (2004)

Shape Feature Extraction of Wheat Leaf Disease Based on Invariant Moment Theory

Zhihua Diao[*], Anping Zheng, and Yuanyuan Wu

Zhengzhou University of Light Industry, College of Electric and Information Engineering,
Dongfeng Road 5, 450002, Zhengzhou, China
{diaozhua,13613824027}@163.com, yuanyuan539@yahoo.com.cn

Abstract. Shape feature extraction is a key research direction on wheat leaf disease recognition. In order to resolve the problem of translation, scaling and rotation transformation invariance on shape matching, the invariant moment theory was introduced to shape feature extraction and seven Hu invariant moment parameters were defined as shape features. Meanwhile the present algorithm was used and new parameters were defined for shape feature extraction research on wheat leaf disease image. The shape features suitable for two types of wheat leaf disease recognition were received and applied in wheat disease intelligent recognition system. The results show that the system recognition rate is relatively high, and can meet the practical application requirements.

Keywords: feature extraction, shape feature, invariant moment, wheat disease.

1   Introduction

With the development of technology and the improvement of computer image processing technology, image processing and pattern recognition technology can be used as an effective way for the diagnosis and identification of wheat diseases. Feature extraction is a key research topic in image processing and pattern recognition, and broadly refers to a transformation which transforms the samples in high dimensional space to low-dimensional space through the map or transformation to achieve the purpose of dimensionality reduction [1]. In wheat leaf disease aspects, shape feature is an important feature to describe the image. And as a local feature, it describes the geometric properties of the local area. As the target shape of images obtained from different vision may be involved in a large difference, translation, scale and rotation transformation invariance problems need to be solved for accurate shape matching.

In recent years, many researchers have carried out the image shape feature extraction research and achieved certain results in agricultural field. Chen Jiajuan et al [2] extracted some shape features such as the length of the chain, chain code area and shape factor, tortuosity, irregularity and so on based on the holes and incomplete edges in cotton leaves to determine the extent of the victims of cotton pests. Xie Guojun et al [3] proposed a fruit shape feature extraction algorithm based on multi-dimensional vision system for fruit intelligent selection and grading. Cai Jianrong et al [4] obtained the best-fitting ellipse parameters to describe orange shape using genetic algorithm. Wang Xiaofeng et al [5] introduced a method to identify the leaf based on shape features of leaf images. Zhao Jinhui et al [6] extracted the lesion area width, squareness and roundness for the separation of disease lesion according to sugarcane seedling image characteristics of red rot and ring spot diseases. Yang Zengfu et al [7] detected the edge of tea buds according to the shape characteristics of tea buds, and which was used for tea buds identification. However, these studies did not consider the basic translation, scale and rotation transformation invariance problems, and had some flaws.

Invariant moment theory was introduced for shape feature extraction of wheat leaf disease image in this paper, and seven Hu Invariant moments were selected as shape parameters. Then shape features of wheat leaf disease image were extracted. A feature subset including twelve shape features was received, and was applied to wheat leaf disease identification system which acquires relatively high recognition efficiency.

2   Shape Feature Extraction

Shape feature is an important feature to describe the image. Firstly, shape often links to the target, which has a certain semantic meaning. So shape feature can be seen higher level feature than color and texture feature. In addition, the shape expression is much more complex than the expression of the color or texture in nature. Secondly, the description of the target shape is a very complex issue. In fact, we still have not found the exact mathematical definition of shapes, including geometric, statistical, or morphological definition, so that it can be consistent with people's feelings. People's feeling of shape is a general result between retina feelings and knowledge about the real world of people. In order to accurately match the shape, the problem of translation, scaling, rotation invariance needs to be solved. As the target shape of images obtained from different vision may be involved in a large difference, translation, scale and rotation transformation invariance problems need to be solved for accurate shape matching.

2.1   Invariant Moment

Invariant moment is a statistical image feature which meets the translation, rotation and scale invariance, and has been widely used in image recognition. There are many invariant moments Used to describe shape which are arithmetic invariant moment, Legendre moment and Zerrnike moments, etc. [8].

For the integral area of S , if we have a given two-dimensional continuous function $f(x, y)$, the $(p+q)$ order moment is defined as follows,

$$m_{pq} = \iint x^p y^q f(x, y)dxdy, \; p, q = 0, 1, 2, \cdots \; . \tag{1}$$

If the digital image $f(x, y)$ is a piecewise continuous function, and has non-zero value in a limited part of the xy plane, then we can prove the existence of various order moments. And the moment sequence m_{pq} could be uniquely determined by the image $f(x, y)$. At the same time the moment sequence m_{pq} also uniquely determine the image $f(x, y)$. The central moments can be expressed as follows,

$$\mu_{pq} = \iint (x - x_c)^p (y - y_c)^q f(x, y)dxdy \; . \tag{2}$$

For digital images, the above integral can be replaced by the following formula,

$$\mu_{pq} = \sum_x \sum_y (x - x_c)^p (y - y_c)^q f(x, y) \; . \tag{3}$$

The Standardized central moment is $\eta_{p,q}$, which can be expressed mathematically as follows,

$$\eta_{p,q} = \frac{\mu_{p,q}}{\mu_{0,0}^\gamma} \qquad where \quad \gamma = \frac{p+q+2}{2}, p+q = 2, 3, \cdots \; . \tag{4}$$

In 1962, Hu [9] proposed the invariant moment theory of image recognition, and established a statistical feature extraction method for image recognition which has been widely used. Li [10] provided a calculation of 52 invariant moments, and seven common invariant moments are presented. From the above equation (1), (2) we could get seven H_u moment invariants, which can be expressed as follows,

$$\phi_1 = \eta_{20} + \eta_{02} \; . \tag{5}$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \; . \tag{6}$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \; . \tag{7}$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \; . \tag{8}$$

$$\begin{aligned} \phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})\left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right] \\ + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right] \end{aligned} \quad . \tag{9}$$

$$\phi_6 = (\eta_{20} - \eta_{02})\left[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \ . \tag{10}$$

$$\begin{aligned}\phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})\left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right] \\ &+ (\eta_{03} - \eta_{12})(\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right]\end{aligned} \ . \tag{11}$$

Although the seven invariant moments can well describe shape character, but when the image database is large, only 7 scalar features are not enough for feature extraction. Because the information carried by their own is very limited, they usually need to be combined with other conventional and global shape features in order to well describe the actual shape properties of the object.

2.2   Definition of New Shape Feature

Seven H_u invariant moments were selected as shape features in this paper, while area, perimeter, density and other parameters were selected as initial features. And two parameters were defined as shape features in intelligent identification system, which are as follows,

1） Number of spots (NS), which mainly used to calculate the number of spots in a disease leaf.
2） Ratio of area and number of spots (RANS). This parameter is an area of measurement of single spot, which mainly used to distinguish between large and small lesion spots.

3   Result and Discussion

In this study two categories of wheat rust and wheat leaf blight are as the main object, and the features are calculated according to the above definition formula. The effect drawing of shape feature extraction is shown in Figure 1.

From experimental statistics we found that the above-defined parameters of moment invariants have a good translation, scale and rotation transformation invariance. While the defined three parameters of ratio of area and number of spots, density and number of spots can be identified as core features of identification of wheat rust and wheat leaf blight. When the value of ratio of area and number of spots is more than 1200, the wheat disease can be considered as the leaf blight disease. And when the parameter is less than 500, the disease can be considered as wheat rust disease. For the parameters of density and number of spots, they need to be used in combination with the parameter of ratio of area and number of spots. When the parameter of ratio of area and number of spots is used as identification basis, the recognition rate of wheat diseases is about 95.3%. When the three parameters of ratio of area and number of spots, density and number of spots are used as identification basis, the recognition rate of wheat diseases is about 97.1%. When invariant moments

are taken into consider, and support vector machine algorithm is used as the model of disease identification, the recognition rate of wheat diseases is up to 99.4%. Using the above-defined shape features as the input of wheat disease identification system, we could obtain a higher recognition efficiency which can fully meet the needs of practical application.
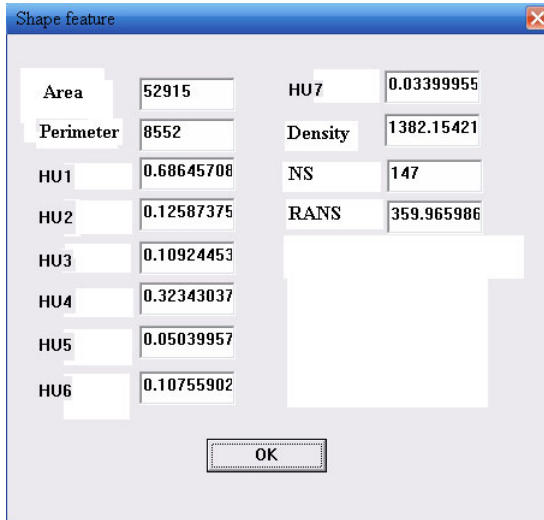
Fig. 1. Interface of shape feature extraction

4   Conclusion

Feature extraction process is a process to reduce the entropy of image information, and how to define and use features to express the image is the basis of image recognition [10]. In this study, combined with the current research of shape feature extraction, the invariant moment theory was introduced to shape feature extraction and seven Hu invariant moment parameters were defined as shape features. Meanwhile new parameters were defined for shape feature extraction research on wheat leaf disease image. The results showed that the system recognition rate is relatively high, and can meet the practical application requirements.

The shape features in this study is relatively easy to extract, and could reflect the morphological characteristics of two kinds of wheat disease images. But how to extract more features that can directly reflect the characteristics of different wheat disease images is the direction of future efforts.

References

1. Wang, J., Ci, L., Yao, K.Z.: A Survey of Feature Selection. Computer Engineering and Science 27, 68–71 (2005) (in Chinese)
2. Chen, J.J., Ji, S.W., Li, J., Zhao, X.D.: Automatic Measurement of Danger Degree of Cotton Insect Pests Using Computer Vision. Transactions of the CEAE 12, 157–160 (2001) (in Chinese)
3. Xie, G.J., Cao, Q.X., Liu, J.Z., Guo, F., Zhou, J.L.: Method for fruit shape feature acquisition based on multidirectional vision. Transactions of the CSAE 23, 127–132 (2007) (in Chinese)
4. Cai, J.R., Fan, J., Li, Y.L., Zhao, J.W., et al.: Shape feature extraction of on-tree citrus based on genetic algorithms. Journal of JiangSu University(Natural Science Edition) 28, 469–472 (2007) (in Chinese)
5. Wang, X.F., Huang, D.S., Du, J.X., Zhang, G.J.: Feature Extraction and Recognition for Leaf Images. Computer Engineering and Applications 3, 190–193 (2006) (in Chinese)
6. Zhao, J.H., Luo, X.W., Zhou, Z.Y.: Image Segmentation Method for Sugarcane Diseases Based on Color and Shape Features. Transactions of the Chinese Society for Agricultural Machinery 39, 100–103 (2008) (in Chinese)
7. Yang, F.Z., Yang, L.L., Tian, Y.N., Yang, Q.: Recognition of the Tea Sprout Based on Color and Shape Features. Transactions of the Chinese Society for Agricultural Machinery 40, 119–123 (2009) (in Chinese)
8. Ding, M.Y., Chang, J.L., Peng, J.X.: Research on moment invariants algorithm. Journal of Data Acquisition and Processing 7, 1–9 (1992) (in Chinese)
9. Hu, M.K.: Visual pattern recognition by moment invariant. IEEE Trans. Information Theory 8, 179–187 (1962)
10. Li, Y.J.: Reforming the theory of invariant moment for pattern recognition. Pattern Recognition 25, 723–730 (1992)
11. Yu, X.W., Shen, Z.R., Gao, L.W., Li, Z.H.: Feature Measuring and Extraction for Digital Image of Insects. Journal of China Agricultural University 8, 47–50 (2003) (in Chinese)

Nitrogen Status Estimation of Winter Wheat by Using an IKONOS Satellite Image in the North China Plain

Liangliang Jia[1,2], Zihui Yu[1], Fei Li[1], Martin Gnyp[3], Wolfgang Koppe[3], Georg Bareth[3], Yuxin Miao[1], Xinping Chen[1,*], and Fusuo Zhang[1]

[1] College of Resources & Environmental Sciences, China Agricultural University, 100094, Beijing, China
[2] Institute of Agricultural Resources & Environment, Hebei Academy of Agriculture and Forestry Sciences, 050051, Shijiazhuang, China
[3] Department of Geography, University of Cologne, 50923, Cologne, Germany
`jiall@cau.edu.cn`

Abstract. The objective of this study was to determine relationship between high resolution satellite image and wheat N status, and develop a methodology to predict wheat N status in the farmers' fields. Field experiment with 5 different N rates was conducted in Huimin County in the North China Plain, and farmers' fields in 3 separated sites were selected as validation plots. The IKONOS image covering all research sites was obtained at shooting stage in 2006. The results showed that single band reflectance of NIR, Red and Green and vegetation indices of NDVI, GNDVI, RVI and OSAVI all well correlated with wheat N status parameters. Field validation results indicated that the prediction models using OSAVI performed well in predicting N uptake in the farmers' fields ($R^2 = 0.735$). We conclude that high resolution satellite images like IKONOS are useful tools in N fertilization management in the North China Plain.

Keywords: IKONOS image, vegetation indices, nitrogen status, winter wheat.

1   Introduction

The North China Plain is one of the most important grain yield production areas in China. Excessive N fertilizer application is considered as a common problem for wheat production in this area and results in low nutrient use efficiency and potentially exerts more pressure on the environment such as nitrate leaching to the ground water[1-4]. In recent years, scientists have developed several in-field soil-plant analysis methods to monitor the N status of various crops to optimize the timing and rate of N application, including the plant total N content and SPAD readings[5], the testing of sap nitrate in the basal stem of wheat at specific growth stage[6-7] and soil Nmin test[8-9]. All these approaches, however, need destructive sampling and laboratory analysis, which limited their applications for wheat N management for large areas in the North China Plain.

---

* Corresponding author.

Considering these problems, some researches now focus on remote sensing methods especially using canopy multispectral reflectance or aerial photography to monitor wheat growth conditions. Aerial photography, which uses color or infrared film, has been used in detection N deficiency, evaluation of plant growth status and prediction of N fertilizer requirements of wheat[10-11] and maize[12-14]. In these contributions, aerial photography provided high spatial resolution images of crop canopy and showed the potential of quantifying crop N status variation within and between fields. However, aerial photography is always restricted by the weather conditions and flying control on the ground. Another shortcoming is that aerial photography can only cover small areas with plane[12] or hellium balloon[11,15]. When it comes to regional scale, such as a county level or provincial level, the acquisition of crop canopy images will be too time-consuming and expensive.

With the development of remote sensing technology, satellites now can provide commercial, public accessible and high spatial resolution image data. For example, IKONOS or Quickbird can provide multi-spectral data of 2-4 meter resolution covering the blue to near infrared spectrum, which may be used for crop N management[16]. Wright et al.[17] found significant correlations between NDVI, extracted from IKONOS image, with N content of wheat flag leaves and pre-season N input. Zhang et al.[18] indicated the IKONOS image successfully estimate the rice leaves N content. Shou et al.[19] found significant correlations between vegetation indices (NDVI, RVI, GNDVI and DVI) and digital values of NIR, Red, Green and Blue bands extracted from a Quickbird image with N status indicators (chlorophyll meter readings, stem sap nitrate concentration, aboveground biomass and shoot N concentration at shooting stage). However, all these researches were conducted in small and controlled experiments. Studies conducted in real farmers' fields, especially in overfertilized fields in the North China Plain, and which transfer the small experimental based studies results to farmers fields N recommendation, are very limited. Therefore, the objective of this research was to use high spatial resolution, multispectral satellite image to predict winter wheat N status in farmers' fields at the shooting stage, which was the critical N sensitive stage to wheat, and evaluate this approach against conventional soil and plant sampling methods.

2   Material and Methods

2.1   The Experiments

2.1.1   Experiment 1: Nitrogen Fertilization Rate Experiment

A winter wheat field experiment was conducted in 2005/2006 in Zijiao town, Huimin County in Shandong province in the North China Plain. This area was representative for soil and crop management in North China Plain. The soil of the experimental field was fine-loam with total nitrogen of 0.79 g/kg, Olsen P of 14.2 mg/kg, exchangeable

K of 201.5 mg/kg and organic matter content of 13.8 g/kg in the 0-30 cm layer. The experiment included 5 pre-plant N levels of 0, 25, 50, 75 and 100 kg/ha with 4 replications. The pre-plant N levels covered the basal fertilization rate on the North China Plain. At shooting stage, each plot was divided into two parts, one received no N topdressing while for the other part N topdressing was applied. The fertilization rate was determined based on an improved Nmin method[8]. The detailed N fertilization management practices in this research are described in Table 1. The area of each plot was 150 m^2 (10 m×15 m). All plots received 120 kg/ha P$_2$O$_5$ as triple super-phosphate and 90 kg/ha K$_2$O as potassium sulfate before wheat sowing in order to avoid soil P and K deficiency. The wheat variety Lumai 23 was sown on 14 October, 2005, and harvested on 10 June, 2006.

**Table 1.** N fertilizer management of the winter wheat experiment

N fertilizer management	1	2	3	4	5
Nmin before sowing (0-60 cm, kg/ha)	49	59	81	98	95
Basal N fertilizer (kg/ha)	0	25	50	75	100
N topdressing rate (kg/ha)	113	107	94	55	54
Total N input (kg/ha)	49	84	131	173	195

2.1.2   Experiment 2: Farmers' Fields

Besides the field experiment, three separate sites of Xujia, Xili and Shizhang were selected as the sub-study areas (Figure 1), which included 10, 12 and 16 farmers' fields, respectively. The sites were selected due to similar wheat varieties were planted in each sites. All the farmers' fields were managed by the farmers themselves, and the basal and topdressing nitrogen fertilization rates for the selected fields were summarized in Table 2.

Table 2. The average basal N fertilization rate for the selected farmers' fields in 2005/2006 wheat season

Sites	Total area (ha)	Number of Fields	Varieties	Planting date	Averaged basal N rate (kg/ha)
Xujia	3.81	12	JM21	14, Oct. 2005	117 (58-180)
Xili	4.00	16	JM23	15, Oct. 2005	149 (49-316)
Shizhang	3.75	10	JM22 and LM23	18, Oct. 2005	78 (34-120)
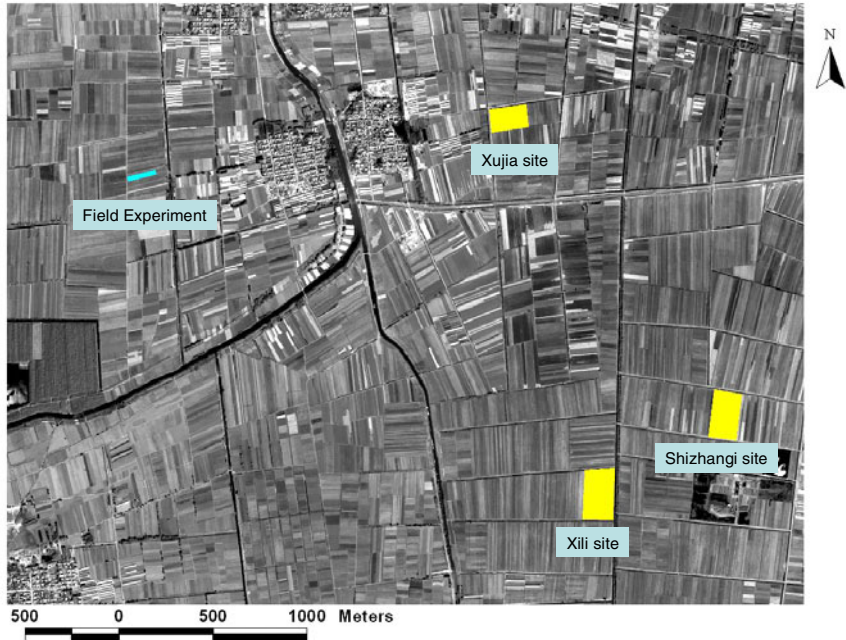
Fig. 1. The distribution of experiment and research sites at 2005/2006 wheat season in Huimin County, Shandong Province

2.2   Field Survey and Agronomic Measurements

Wheat N status indicators were measured at the shooting stage in Apr. 10, 2006. The nitrate concentration of the plant stem sap was determined with a Reflect Meter (Merck Co., Darmstadt, Germany), and SPAD readings were taken on the uppermost fully expanded leaves from 30 randomly selected plants in per plot and the average value was used for further analysis. Five soil cores were collected from each plot and pooled at 0 to 30 cm, 30 to 60 cm, and 60 to 90 cm depth intervals. To analyze the soil Nmin, all samples were sieved, extracted with 0.01 M CaCl2 and analyzed for NH_4^+-N and NO_3^--N by continues flow analyzer TRAACS 2000[20]. The soil Nmin was used to determine the optimum N topdressing rate at shooting stage as described by Chen et al.[8]. Aboveground wheat biomass were harvested (1 m^2 per plot) for the experiment, and dried at 60 °C to constant weight for biomass dry weight, and then analyzed for total N concentration using Kjeldahl digestion method.

For the selected farmers' fields in Xili, Xujia and Shizhang sites, only the aboveground plant biomass and total shoot N concentration were determined at shooting stage.

A "Trimble-AgGPS® 132" GPS receiver was used to georeference the sampling points in the experiment and farmers' fields, and the boundaries of the experiment plots.

2.3   IKONOS Image Processing

The IKONOS images, which included 4-m resolution multispectral bands (Blue band: 0.45-0.53 µm; Green band: 0.52-0.61µm; Red: band 0.64-0.72 µm; and NIR band: 0.77-0.88 µm) and 1 m resolution panchromatic band (0.45-0.90 µm), were acquired on Apr. 20th, 2006, at the shooting stage of winter wheat. Since no temporal comparison or with other imagery was planned, the atmospheric corrections in this research was not performed [21]. The collected DGPS points at the transaction points around the study sites were converted to shapefiles using ArcGIS 8.1 and were imported into ENVI 4.2 as a vector layer. This layer was used to georectify the IKONOS image and to extract the plotted pixels' raw digital spectral reflectance values of NIR, Red, Green and Blue bands. At least 3 pixels in each experiment plots were used to get averaged raw digital spectral reflectance values.

2.4   Data Analysis

The vegetation indices (VIs) of NDVI (normalized vegetation index[22]), RVI (ratio vegetation index[23]), GNDVI (green vegetation index[24]) and OSAVI (optimized soil-adjusted vegetation index[25]) were calculated for the obtained IKONOS satellite image. The detailed expressions of the vegetation indices were listed in Table 3.

Table 3. Broadband vegetation indices used for image analysis

Abbreviation	Name	Vegetation Index	Reference
NDVI	Normalized Difference Vegetation Index	NDVI= （NIR-Red） / （NIR+Red）	Rouse et al., 1973
RVI	Ratio Vegetation Index	RVI=NIR/Red	Jordan, 1969
GNDVI	Green Normalized Difference Vegetation Index	GNDVI=(NIR-Green)/(NIR+Green)	Gitelson and Merzlyak, 1998
OSAVI	Optimized soil-adjusted vegetation index	OSAVI =(NIR–Red)/(NIR+Red+0.16)	Rondeaux et al., 1996

Nitrogen effects were analyzed quantitatively by comparing the means of agronomic parameters of each treatment through LSD test at a probability of 0.05.

The single band reflectance of R, G, B and NIR bands and the vegetation indices of NDVI, GNDVI, RVI and OSAVI were regressed with the chlorophyll meter readings (SPAD), soil N input before sowing (including initial Nmin and basal N), stem sap nitrate concentration, shoot N concentration, aboveground biomass and nitrogen uptake. The regressions of single bands reflectance and VIs with plant N status indicators were made using Statistical Analysis System version 8.1 [26]. Linear or nonlinear models were fitted based on the plot patterns and best-fit R^2 values for the relationship. When linear relationships were not significant, exponential or logarithm models were attempted. Then the best-fit models were used to predict, nitrogen uptake in the farmers' fields. The root mean square error (RMSE) and relative error was calculated to compare the precision of estimation between the measured values (X) and the estimated values (X').

3   Results

3.1   Nitrogen Status of Wheat at Shooting Stage

The N status of winter wheat at shooting stage was significantly affected by the N treatments (Table 4). High N supply (initial soil Nmin + pre-plant N fertilization) led to high SPAD readings, aboveground biomass, shoot N concentration, stem sap nitrate concentration and total N uptake. The significant difference among the 5 pre-N treatments showed that the low N supply treatments of 1 and 2 were in the state of N deficiency, having lower aboveground biomass, SPAD readings, stem sap nitrate concentration and N uptake than the high N treatments of 4 and 5, which were in the state of over fertilization. The treatment 3 seemed to have optimum N status with relatively high SPAD readings and stem sap nitrate concentration.

Table 4. Nitrogen status of winter wheat at shooting stage

Items\Treatments	1	2	3	4	5
Nmin (0-90 cm)	56b	62b	82b	124a	112a
Chlorophyll meter readings (SPAD)	46.2c	48.2b	49.6a	50.1a	50.6a
Upland biomass (kg/ha)	852c	1137c	1273b	1543ab	1703a
Shoot N concentration (g/kg)	32.4c	32.7bc	33.8bc	36.6ab	37.8a
N uptake (kg/ha)	27.5d	36.9cd	43.5bc	57.0ab	63.6a
Stem nitrate Concentration (mg/kg)	1228c	1295c	1672b	1834ab	1955a

*Means with different letter are significant different at 0.05 level.

3.2   Correlations of Vegetation Indices with Wheat N Status at Shooting Stage

Significant correlations were found between NDVI, RVI and GNDVI with SPAD readings, stem sap nitrate concentration, aboveground biomass and nitrogen uptake at shooting stage (R^2 = 0.488-0.739) (Table 5). All of the single band reflectance of

NIR, Red and Green had significant correlations with SPAD, stem sap nitrate concentration, biomass and N uptake (R^2 = 0.240-0.645), which were lower than those with VIs. Neither VIs nor single band reflectance used in this research could significantly explain wheat N concentration variation (R^2 = 0.008 – 0.397), but the high R^2 values with aboveground biomass (R^2 = 0.667-0.708) and nitrogen uptake (R^2 =0.678-0.739) suggested that VIs could still be used to explain N and growth status of wheat at shooting stage.

Table 5. Coefficient of determination (R^2) for wheat N status with single band digital value and vegetation indices derived from IKONOS satellite image at shooting stage

R2	SPAD	Stem sap nitrate	Biomass	N Content	N uptake
NIR	0.424*	0.434*	0.572*	0.291	0.615**
Red	0.645**	0.342*	0.501*	0.046	0.430*
Green	0.483*	0.352*	0.426*	0.008	0.329*
Blue	0.284	0.318	0.297	0.015	0.240
NDVI	0.707**	0.507*	0.704**	0.181	0.678**
GNDVI	0.647**	0.541*	0.708**	0.184	0.683**
RVI	0.686**	0.537*	0.667**	0.199	0.704**
OSAVI	0.678**	0.513*	0.707**	0.397*	0.739**

* *LSD*=0.05, ** *LSD*=0.01, number of observations = 20.

3.3   Nitrogen Uptake Estimation of Winter Wheat for Farmers' Fields

The N uptake of the winter wheat at shooting stage for farmers' fields in Xili, Shizhang, Xujia and Xizhangliu were predicted using the relationships developed with the experimental data (Figure 2). The estimated N uptake values using any of the four vegetation indices were all significantly correlated with measured values. Considering the normally farmers N application rate of 277-441 kg/ha in this region, the RMSE for NDVI and OSAVI were acceptable (26.6-27.4 kg/ha) while the RMSE for RVI and GNDVI were too high (42.2 – 61.0 kg/ha) to be acceptable by the farmers. Compared with the other vegetation indices used in this research, OSAVI was the best for nitrogen uptake prediction (closest to 1:1 line), while N uptake predicted with RVI and GNDVI were either higher or lower than measured values at the state of high N uptake. The results suggested that NDVI and OSAVI derived from IKONOS satellite image could be used for monitoring and estimating of wheat growth and nitrogen status.
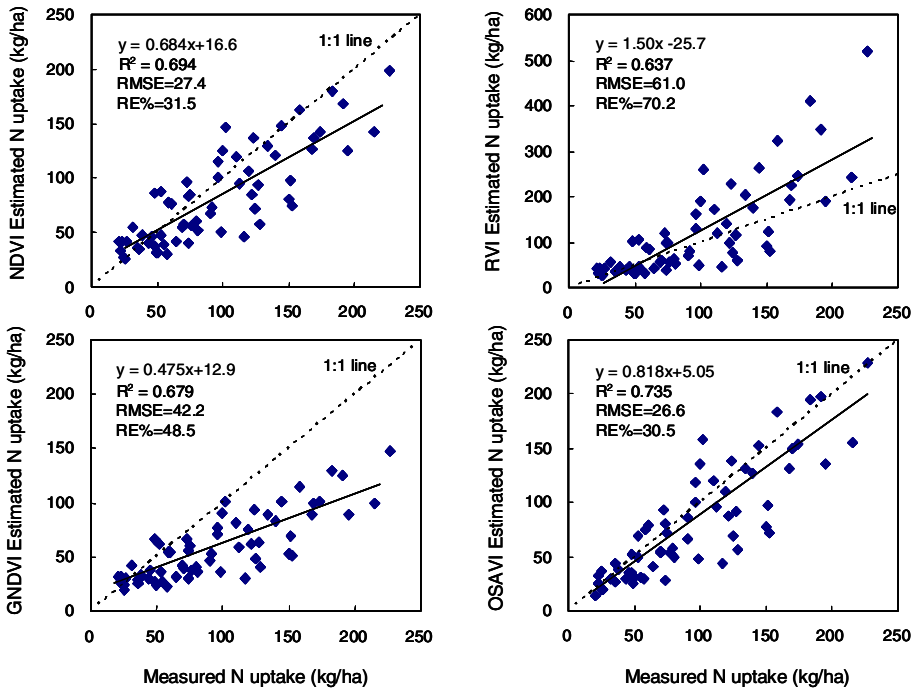
Fig. 2. Comparison of the accuracy of the measured N uptake at shooting stage with predicted N uptake in the farmer's field (RMSE = root mean square error; RE% = relative error; number of observations = 84)

4   Discussion

A very important approach to precision N management is to detect crop N status during the growing season and use this information to guide in-season N management. Traditional soil-plant analysis methods, such as SPAD, biomass, shoot N concent and soil Nmin, are either destructive or too time- and money-intensive for regional investigation. High resolution multispectral satellite imagary provides a promising solution for non-destructive and fast estimation of crops growth and N status. The vegetation indices used in this research all showed significant correlations with wheat N status indicators of SPAD readings, stem sap nitrate concentration, upland biomass and nitrogen uptake, but not with shoot N concentration. These results differ from earlier findings by Shou et al.[19] and Wright et al.[17], which showed that vegetation indices derived from Quickbird satellite image had significant correlations with shoot N concentration. An explanation of the low correlations of IKONOS image with shoot N concentration is that rapid wheat aboveground biomass increase at shooting stage causes unstable changes of the shoot N concentration when field measurements and satellite image were taken. Measurements at later growing

stages and corresponding image acquisition may have a better relationship. More studies are needed to test this hypothesis.

The successful estimation of the nitrogen uptake in overfertilized farmers' fields, rather than in control experiments[27-30], supported the idea suggested by Shou et al.[19] to use high resolution satellite image combined with in-site soil-plant analysis at small reference fields to monitor regional nitrogen status of winter wheat in North China Plain.

The fast development of the satellite technologies have made the acquisition of high spatial and spectral resolution satellite images become cheaper and easier than before. In recent years, some researchers have considered hyperspectral sensors to monitor rice nitrogen status and nutrients disorder[31] and satellites with radar systems (SAR) to study soil and crop moisture[32-33] and forest biomass [34](Austin et al., 2003). Although more sensors have been used in monitoring vegetation growth status than before, satellite images for agriculture uses still face many challenges, such as clouds or dust which interfere the acquisition or quality of the image, low spectral resolution in explaining the canopy and canopy reflectance disorders caused by other factors except nutrients, etc. A possible solution to this problem is multi-source remote sensing data fusion. For example, high spatial resolution IKONOS or Quickbird satellite images combined with high spectral resolution EO-1 Hyperion and soil-plant analysis may provide better explanations for crop canopy nutrient disorders. In such studies, plots with optimum N management should be included to provide reference values. This may allow the development of site-specific fertilizer recommendation maps as described by Scharf and Lory[12]. Further researches are needed to verify this idea and develop a robust system that will allow farmers to access these data to improve their management decisions.

References

1. Zhao, R.F., Chen, X.P., Zhang, F.S., Zhang, H.L., Schroder, J., Roemheld, V.: Fertilization and nitrogen balance in a Wheat-Maize Rotation System in North China. Agronomy Journal 98, 938–945 (2006)
2. Ju, X.T., Kou, C.L., Christile, P., Dou, Z.X., Zhang, F.S.: Changes in the soil environment from excessive application of fertilizers and manures to two contrasting intensive cropping systems on the North China Plain. Environmental Pollution 145, 497–506 (2007)
3. Gao, W., Huang, J., Wu, D., Li, X.: Investigation on nitrate pollution in ground water at intensive agricultural region in Huanghe-huaihe-haihe Plain. Ecological Agriculture Research 7, 41–43 (1999); (in Chinese with English abstract)
4. Zhang, W.L., Tian, Z.X., Zhang, N., Li, X.Q.: Nitrate pollution of groundwater in northern China. Agriculture Ecosystem and Environment 59, 223–231 (1996)

5. Blackmer, T.M., Schepers, J.S.: Analysis of aerial photography for nitrogen. Agronomy Journal 88, 729–733 (1996)
6. Fox, R.H., Roth, G.W., Lversen, K.V., Piekielek, W.P.: Soil and tissue nitrate tests compared for predicting soil nitrogen availability to corn. Agronomy Journal 81, 971–974 (1989)
7. Li, Z.H., Zhang, F.S., Wang, X.R.: Nitrogen nutritional diagnosis and recommendation as topdressing fertilizer N for several crops in North China: diagnosis of nitrate in plant tissue by quick test method. Plant Nutrition and Fertilizer Science 3, 269–274 (1997); (in Chinese with English abstract)
8. Chen, X.P., Zhang, F.S., Roemheld, V., Horlacher, D., Schulz, R., Boening-Zilkens, M., Wang, P., Claupein, W.: Synchronizing N supply from soil and fertilizer and N demand of winter wheat by an improved Nmin method. Nutrient Cycling in Agroecosystems 74, 91–98 (2006)
9. Cui Z.L.: Optimization of the N fertilizer management for a winter wheat – summer maize rotation system in the Northern China Plain – from field to regional scale. Ph.D. Dissertation, China Agricultural University, Beijing, China (in Chinese with English abstract) (2005)
10. Flowers, M., Weisz, R., Heiniger, R.: Remote sensing of winter wheat tiller density for early nitrogen application decisions. Agronomy Journal 93, 783–789 (2001)
11. Jia, L.L., Chen, X.P., Zhang, F.S., Buerkert, A., Roemheld, V.: Low altitude aerial photography for optimum N fertilization of winter wheat on the North China Plain. Field Crops Research 89, 389–395 (2004)
12. Scharf, P.C., Lory, J.A.: Calibrating corn color from aerial photographs to predict sidedress nitrogen need. Agronmy Journal 94, 397–404 (2002)
13. Sripada, R.P., Heiniger, R.W., White, J.G., Weisz, R.: Aerial Color Infrared Photography for Determining Late-Season Nitrogen Requirements in Corn. Agronomy Journal 97, 1443–1451 (2005)
14. Sripada, R.P., Heiniger, R.W., White, J.G., Meijer, A.D.: Aerial Color Infrared Photography for Determining Early In-Season Nitrogen Requirements in Corn. Agronomy Journal 98, 968–977 (2006)
15. Gerard, B., Buerkert, A.: Estimation of spatial variablity in pearl millet growth with non-destructive methods. Expl. Agric. 37, 373–389 (2001)
16. Franizer, P., Lamb, D., Kumar, L.: Remote sensing and satellite image technology: capabilities and implementation (2005),
    `http://www.frfa.asn.au/files/pdf/conference/papers`
17. Wright, D.L., Rasmussen, V.P., Ramsey, R.D., Baker, D.J.: Canopy reflectance estimation of wheat nitrogen content for grain protein management. GIScience and Remote Sensing 41, 287–300 (2004)
18. Zhang, J.H., Wang, K., Bailey, J.S., Wang, R.C.: Predicting nitrogen status of rice using multispectral data at canopy scale. Pedosphere 16(1), 108–117 (2006)
19. Shou, L.N., Jia, L.L., Cui, Z.L., Chen, X.P., Zhang, F.S.: Using high-resolution satellite image to evaluate nitrogen status of winter wheat in the North China Plain. Journal of Plant Nutrition 30(10), 1669–1680 (2007)
20. Bran and Luebbe: Bran+Luebbe Traacs 2000 continuous flow analyzer operation manual. MT9, GB-352-87A and GB-352-87E. Publication No. MT7-50EN-01. Bran+Luebbe GmbH, Norderstedt, Germany (1996)
21. Song, C., Woodcock, C.E., Seto, K.C., Pax-Lenny, M., Macomber, S.A.: Classification and change detection using Landsat TM Data: When and How to Correct Atmospheric Effects? Remote Sensing of Environment 75, 230–244 (2001)

22. Rouse, J.W., Has, R.H., Schell, J.A., Deering, D.W.: Monitoring vegetation systems in the great plains with ERTS. In: Third ERTS Symposium, NASA SP-351, vol. 1, pp. 309–317. NASA, Washington, DC (1973)
23. Jordan, F.: Derivation of leaf area index from quality of light on the forest floor. Ecology 50, 663–666 (1969)
24. Gitelson, A., Merzlyak, N.: Remote sensing of chlorophyll concentration in higher plant leaves. Advances in Space Research 22, 689–692 (1998)
25. Rondeaux, G., Steven, M., Baret, F.: Optimization of soil-adjusted vegetation indices. Remote Sensing and Environment 58, 1–12 (1996)
26. SAS Institute.: SAS/STAT user's guide. Version 8.1. SAS Inst., Cary, NC (1998)
27. Xue, L.H., Cao, W.X., Luo, W.H., Dai, T.B., Zhu, Y.: Monitoring leaf nitrogen status in rice with canopy spectral reflectance. Agronomy Journal 96, 135–142 (2004)
28. Chang, K., Shen, W.Y., Lo, J.-C.: Predicting rice yield using canopy reflectance measured at booting stage. Agronomy Journal 97, 872–878 (2005)
29. Elwadie, M.E., Pierce, F.J., Qi, J.: Remote sensing canopy dynamics and biophysical variables estimation of corn in Michigan. Agronomy Journal 97, 99–105 (2005)
30. Ferrio, J., Villegas, P.D., Zarco, J., Aparicio, N., Araus, J.L., Roco, C.: Assessment durum wheat yield using visible and near-infrared reflectance spectra of canopies. Field Crops Research 94, 126–148 (2005)
31. Chen, J.Y., Tian, Q.J.: Estimating canopy chlorophyll and nitrogen concentration of rice from EO-1 Hyperion data. In: Manfred, O., D'Urso, G., Neale, C.M.U., Gouweleeuw, B.T. (eds.) Proceedings of SPIE Remote Sensing for Agriculture, Ecosystems, and Hydrology VIII, October 3, vol. 6359, p. 63591H (2006)
32. Leconte, R., Brissette, F., Galarneau, M., Rousselle, J.: Mapping near-surface soil moisture with RADARSAT-1 synthetic aperture radar data. Water Resources Research 40, 1029–1038 (2004)
33. Griffiths, G., Wooding, M.: Temporal Monitoring of Soil Moisture Using ERS-1 SAR Data. Hydrological Processes 10, 1127–1138 (1996)
34. Austin, J.M., Mackey, B.G., Van Niel, K.P.: Estimating forest biomass using satellite radar: an exploratory study in a temperate Australian Eucalyptus forest. Forest Ecology and Management 176, 575–583 (2003)

Assessment of Potential Risk in Soil and Early Warning Analysis in Four Counties, Northeast China

Lingling Sang[1], Chao Zhang[1,*], Jianyu Yang[1], Dehai Zhu[1], and Wenju Yun[2]

[1] College of Information and Electrical Engineering,
China Agricultural University, Beijing, 100083, P.R. China
zhangchaobj@cau.edu.cn
[2] Land Consolidation and Rehabilitation Center,
Ministry of Land and Resources, Beijing 100035, P.R. China

Abstract. Heavy metal pollutions in the soil have been an important content of its quality. The objectives of this study were to evaluate the potential risk of heavy metals and early warning analysis of agricultural soil in four counties, Northeast China. One method based on toxicity coefficients was applied to assess the potential risk of these heavy metals in soil. The results showed that average content of Cu, Pb, Zn, Cd, As, Ni, Hg and Cr in agricultural soil was lower than the first level standard of agricultural soil, and Cd led to risk was the most important factor, which 0.21% of the area was in the state of potential risk of elements Cd. And other heavy metals were not harmful to the environment, which potential risk of harm in descending order is E (Cd)> E (Hg)> E (As)> E (Ni)> E (Cu)> E (Pb)> E (Cr)> E (Zn); The other method is that the Cd pollutants can be calculated by the atmospheric dry and wet deposition, the irrigation water, the chemical fertilizer and the crops. There was a point on the assumption that the pollution caused by the atmospheric dry and wet deposition was worse than the pollution from irrigation water and chemical fertilizer. It is concluded that heavy metal pollution of Cd in the southeast of the study area will be getting worse within the next few decades by forecast warning analysis. Overall, some part of study area is in the state of potential risk and the Cd of soil is a major factor to control the retention of the heavy metals in the soil. According to analysis of their characteristics and the level of potential risk, there will not be great impact on quality by developing appropriate control measures timely.

Keywords: agricultural soil, heavy metal, potential risk, early warning.

1 Introduction

The quality of the field is not only the comprehensive measurement for the quality of the soil, environment and health of the overall field, but also the main embodiment for the ability of maintaining the productive power, fighting against natural calamities, and hazard-free treatment of the agricultural products and the high nutrition and

---

* Corresponding author.

quality[1]. In recent years, the evaluation for the agricultural products usually focuses on the productivity and the effectiveness, among which the classification of the agricultural field and the earth chemical evaluation for the quality of the soil are considered as two important branch of the soil quality evaluation [2, 3]. Those two works undertake scientific evaluation to the quality of the soil as well as the eco- environment from different perspective and conform to each other in terms of the rational application of the soil. However, the evaluation upon the agriculture focuses on the effects of the quality of agricultural field and considers the earth chemical evaluation for the quality of the soil as part of the standard for the evaluation of the environment-friendly and healthy elements [4].

In recent years, the intensified industrialization and urbanization as well as the effects of human activity lead to many eco environmental problems, among which the heavy metal pollution of the agriculture field has aroused extensive concern of the overseas and domestic scholars [5]. As the important earth resource, agriculture field has shift from the state of insufficient quantity to the state of premium insufficient quantity, which in essence implies important environment pollution and the food safety and affects negatively upon the human health and environmental safety. Therefore, it is of high significance to undertake risk evaluation to the heavy metal pollution of the soil and the early warning analysis [6].

Taking some counties in Northeast China as the sample region, based on the data of heavy metals, the paper analyzed its potential sources of heavy metals and assessed the potential risks of heavy metals. According to early warning after the study, the protective measures are put forward, providing the scientific suggestion for the farmland protection and the security of land resource.

2   Data and Methods

2.1   Study Area and Data

The southwest of Songnen Plain, which is also the western region of Jilin Province, belongs to the sub humid and sub arid continental and seasonal climate zone with obvious changes of the weather in the year. The annual rainfall is about 400mm, which is $3.5-4.7$ times of the evaporation capacity. Due to the difference of the agricultural climate condition and the earth eco condition, it forms the combined agricultural production state of both the typical agricultural region and the ecotone between agriculture and animal husbandry. The soil types are mainly composed of meadow soil, black earth, saline-alkali soil, and aeolian sandy soil. The loose structure of the soil deduces the capability of preserving moisture and fertility, so that the natural structure and the fertility would be damaged by the drought and flood. Because of human beings' interference, the quality of the soil decreases by a large margin. It is now becoming the potential sand resource for the sandy soil. While the accumulated substance in the plain is sticky and heavy with low penetration power and developing water net so that the water cannot flow out and the soluble salt cannot be drain out, and due to the evaporation, the soil is easy to undertake salinization and

desertification. Because of the long term overlook of the vulnerably field as well as the long-term and large scale agricultural development, the eco environment of this region is deteriorated, the agricultural and economic development is relatively slow.

This research applies BJZ 54 and Gauss-Kruger Projection to pasteurize the blank miles grid, so that the data of the investment sample could be integrated with the attributes of the blank miles grid and construct the foundation of the research work. The main resource of the data derives from the basic material for the chemical evaluation of the earth quality provided by the Ministry of Land and Resources. In accordance with the multi target earth chemical investigation standard, we station according to the earth chemical sampling method. With the facilitation of the fieldwork illustration of 1:50 000 topographic map, we undertake the investigation on the state-province level. The main investigation focuses on 50 elements including the soil, ground water, shallow ground water, atmospheric bulk deposition, and crop's grain etc. The intensity of the sample is 1 point/km^2, among which there are 5179 soil data sampling spots. The sampling station of this time and the sample analysis are completed by China University of Geosciences and Geological Survey Institute of some Province jointly.

2.2   Potential Risk Assessment

The potential risk index method developed by Hakanson [7] was applied in this study. According to this method, the potential risk coefficient (E_r^i) of single element and the potential risk index （RI） of multi-element can be computed via the following equations:

$$RI = \sum E_r^i \tag{1}$$

$$E_r^i = T_r^i \times C_r^i \tag{2}$$

$$C_r^i = C_D^i / C_F^i \tag{3}$$

In these equations: $C_r^i = C_D^i / C_F^i$ is the accumulating coefficient for the element of "i"and is single heavy metal pollution index; C_D^i is actual density of heavy metal quality in soil; C_F^i is background value of heavy metal quality in soil; RI is comprehensive potential risk index of muti-element; E_r^i is the potential risk index of single element; T_r^i is the toxicity coefficient for the heavy metals element of "i", reflects its toxicity levels and the sensitivity to harm the human and the ecosystem. The toxicity coefficients for common heavy metals were Hg （40） > Cd （30） >As （10） > Pb （5） =Ni （5） =Cu （5） >Cr （2） >Zn （1） [7][8]. Table 1 is the Classification standard of potential risk of heavy metals in soil.

Table 1. Classification standard of potential risk of heavy metals in soil

E_r^i	Pollution degree	RI	Pollution degree
$E_r^i < 30$	Low potential risk	$RI < 135$	Low potential risk
$30 \leq E_r^i < 60$	Moderate potential risk	$135 \leq RI < 265$	Moderate potential risk
$60 \leq E_r^i < 120$	Considerable potential risk	$265 \leq RI < 525$	Considerable potential risk
$120 \leq E_r^i < 240$	Very high potential risk	$RI \geq 525$	Very high potential risk
$E_r^i \geq 240$	Serious potential risk		

2.3   Early Warning Analysis

In this paper, the research on the atmospheric dry and wet deposition, the irrigation water and the chemical fertilizer in the Cd were analyzed. Input pathway of Cd in soil is mainly from the atmospheric dry and wet deposition, the irrigation water and the chemical fertilizer, output pathway of Cd in soil is mainly from the crops [9, 11]. According to the concept of potential risks, the total Contents of the Cd pollutants can be calculated by the atmospheric dry and wet deposition, the irrigation water, the chemical fertilizer and the crops via the following equations:

$$DI = (D + I + F) - OP \tag{4}$$

$$M = DI_m \times n + M_0 \tag{5}$$

In these equations: DI is the total input of heavy metal in soil, and the unit is g/(hm^2·a); D, I, F are respectively input of heavy metal in soil by the atmospheric dry and wet deposition, the irrigation water, and the unit is g/(hm^2·a); OP is the total output of heavy metal in soil for the element of "i", and the unit is g/ (hm^2·a); M is the total content of heavy metal in soil for the element of "i" for several years (n) later, and the unit is mg·kg^{-1}; M_0 is the current input of heavy metal in soil, and the unit is mg·kg^{-1}.

3   Results and Discussion

3.1   Characteristic Analysis of the Content of Heavy Metal in Farmland Soil

The analysis results of heavy metal contents of farmland soil in four counties of Northeast China are listed in Figure 1. Statistical analysis of the heavy metals of Cd, Hg, As, Cu, Pb, Cr, Zn and Ni in farmland soil has proved to be normal distribution. In this paper, arithmetic mean value and median value were used to describe the contents of heavy metals in the soil. Compared with Environmental Quality Standard for Soils in China (GB 15618-1995) [10], average content of Cu, Pb, Zn, Cd, As, Ni, Hg and Cr in agricultural soil was lower than the national first class standard of agricultural soil, and the heavy metal contents in the soil was in the natural background level basically in study area. In addition to Cd, the maximum value of the content of each element was lower than the national secondary standard.
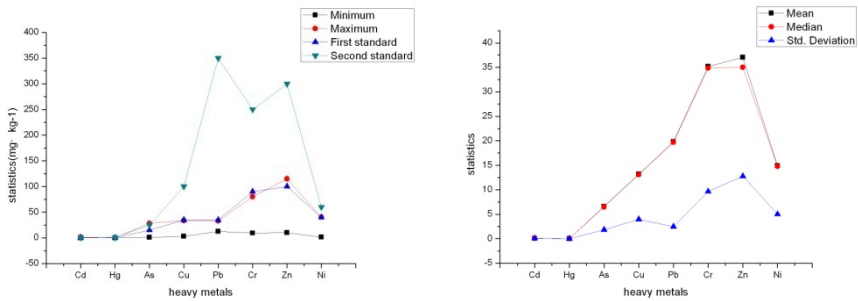
Fig. 1. Statistics of heavy metal concentrations in soil of study area

Based on the monitoring data of soil quality of study area, a quantitative analysis of farmland soil is conducted by applying the method of single and composite pollution exponent Index*e*s. The analysis results showed that there were 11 units in the soil samples pollution of heavy metal Cd and one unit in the soil samples pollution of heavy metal As. Other units in the soil samples were almost non-pollution. Three units in soil samples were light pollution (integration pollution index is from 1.0 to 2.0), and ten units in soil samples were to be at the alert posture (integration pollution index is from 0.7 to 1.0). Other units' synthesis pollution index was smaller than 0.7, belongs to the security rank (integration pollution index is lower than 0.7). 99.79% of the total land area was at safety level, 0.17% affected by light pollution and 0.04% affected by medium pollution.

Several results had been obtained: At the 5179 sampling sites in study area, the heavy metal pollution index descended in the order of Cd(2.30) > As(1.14) > Hg(0.52) > Cr(0.46) > Pb(0.11). The contents of heavy metals of Cd and As in the soil were higher than the contents of other heavy metals, separately accounts for 0.21% and 0.02% in Table 2. So there should pay attention to the increase of the content of soil Cd and bring the potential risks.

Table 2. Heavy metal pollution

heavy metal	index(Minimum)	index(Maximum)	index(mean)	Portion (%)
Cd	0.020	2.297	0.172	0.21%
As	0.040	1.140	0.259	0.02%

## 3.2 Single and Comprehensive Potential Risk Assessment of Heavy Metal in Soil

Refer to the Hakanson assessment method, heavy metal's potential risk index of single element in soil (E_r^i) and comprehensive potential risk index of muti-element (RI) were shown in Table 4. From Table 3, the scopes of the potential risk coefficients of 8 kinds of heavy metals are: E_r^i (Cd) 1.8~206.7, E_r^i（Hg）1.07~69.6, E_r^i (As) 0.67~19, E_r^i（Cu）0.42~4.81, E_r^i (Pb) 1.8~4.73, E_r^i（Cr）0.21~1.78、E_r^i (Zn)

0.1~1.15, E_r^i (Ni) 0.18~5.08。 From the means of the potential risk coefficients of 8 kinds of heavy metals, the potential risk descended in the order of E (Cd) >E (Hg) >E (As) >E (Ni) >E (Cu) >E (Pb) >E (Cr) >E (Zn). According to the calculated accumulating coefficients, Hg, As, Cu, Pb, Cr, Zn and Ni had not achieved the ecology risk level, Cd was the most serious polluting element among these heavy metals, and its mean value of E_r^i was up to 1.8, and Cd was the main influencing factor to causes the risk. Some 0.21% region existed potential risk of Cd element, and other heavy metals were not harmful to the environment. Therefore Cd element was the key element to determine to study.

The scope of the potential risk index (RI) of multi-element is 14.87~228.74 in Table 3. According to its frequency distribution in Figure 2, there is no area in the considerable risk and the serious potential risk. 0.06% of the soil sample points were in a medium risk degree, which the mainly reason was that potential risk coefficient of Cd element was great. The result was shown that agricultural soil in study area would encounter the Moderate risk of Cd element, and 3 soil sampling points (0.06%) were above the level of risk.

Table 3. Statistics of potential risk coefficient (E) and potential risk index（RI）of heavy metals in study area

	potential risk coefficient E_r^i								RI
	Cd	Hg	As	Cu	Pb	Cr	Zn	Ni	
Minmum	1.80	1.07	0.67	0.42	1.80	0.21	0.10	0.18	14.87
Maximum	206.70	69.6	19.00	4.81	4.73	1.78	1.15	5.08	228.74
Mean	14.29	5.02	4.40	1.89	2.84	0.78	0.37	1.86	31.44
Std.Deviation	7.51	3.34	1.21	0.57	0.35	0.21	0.13	0.63	10.02
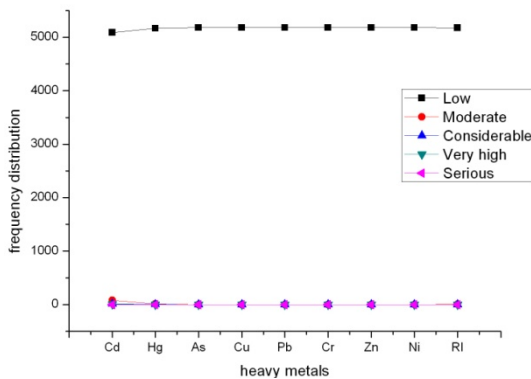
Fig. 2. Frequencies of potential risk coefficient (E) and potential risk index（RI）of heavy metals in study area

3.3   Forecast and Early Warning Analysis

Through the statistical analysis on contents of heavy metals by the atmospheric dry and wet deposition, the ground water and the surface water, the atmospheric dry and wet deposition is the main input way of Cd. Drawing up the maps of Cd distribution by the spatial interpolation, the results showed that the greatest regional distribution affected by the atmospheric dry and wet deposition were located in the central and southwestern part of Da'an, as well as the northern and southern of Zhanlai. Besides, some areas in the eastern of Qianguo were occupied by acid soils which Cd pollution needed to pay more attention. This is because the soil pH is an important factor affecting the content of Cd. The content of Cd in crops will enrich with the increase of pH value [11].

Distribution characteristics of Cd were obtained through connecting the layer properties, re-calculating the content of Cd by the atmospheric dry and wet deposition and restart interpolation. The potential risks were only considered by input of the dry and wet deposition, and the content of Cd in soil of agricultural land was increased at an average annual speed of 0.006 mg in Table 4. Evaluations on the potential risk and forecasting were carried by the formula (5), and calculated the content of Cd in the next 20, 50 and 100 years.

Table 4. Annual input of Cd each grade

grade of Cd	Annual input $(mg \cdot kg^{-1})$
1	0.0005
2	0.001
3	0.0015
4	0.0025
5	0.004
6	0.006

Through computing the content value of Cd in soil in the next 20, 50 and 100 years, and comparing with the standard value of soil environmental quality, the pollution indexes of Cd element in the next 20, 50 and 100 years were calculated and determined whether the risk exists. The areas with heavy warning in the southeast of study area will be lead Cd pollution in the agricultural land area in 20 years in Figure 4(a); the areas with moderate warning in the southeast of study area will be lead Cd pollution in the agricultural land area in 50 years in Figure 4(b), and the areas with light warning in the southeast of study area will be lead Cd pollution in the agricultural land area in 100 years in Figure 4(c).

Through the early warning analysis of Cd element in research area, the area of agricultural land which was polluted by Cd, will be 1 634.79 hectares in 20 years; the area of agricultural land which was polluted by Cd will be 4 050.36 hectares in 50 years; the area of agricultural land which was polluted by Cd will be 10 355.35 hectares in 100 years. 1.5% of the total area was with potential risks, and mainly located in the southeastern of study area, where of agricultural land was high-yield and soil type are mainly based on sand. From Figure 3, the area with warning will continue to expand as time goes by, so some preventive measures and early warning suggestions are put forward on the problems.
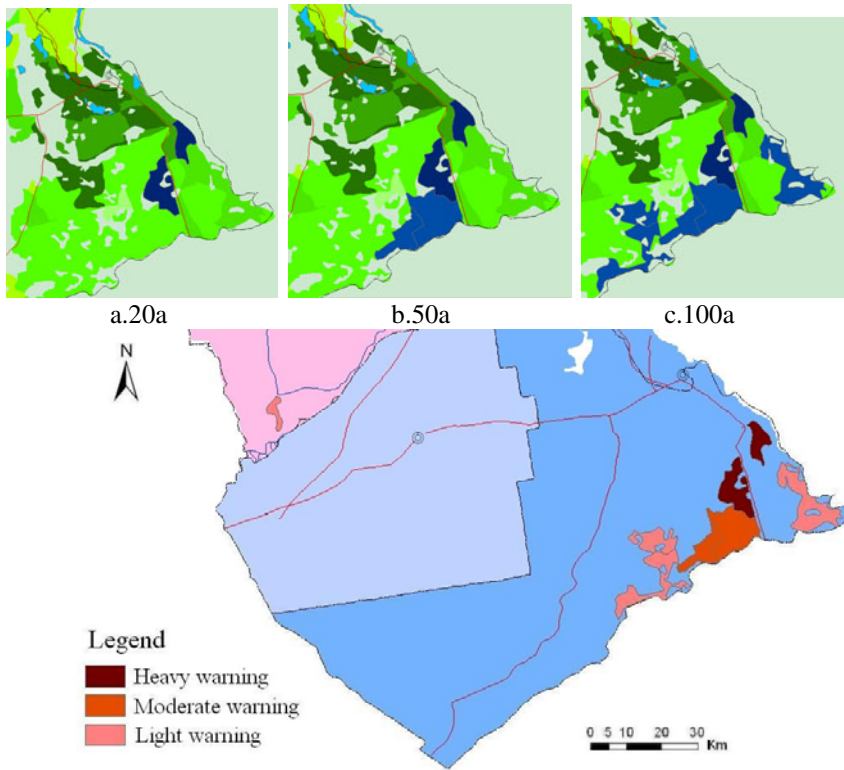
a.20a   b.50a   c.100a

Legend
Heavy warning
Moderate warning
Light warning

Fig. 3. Warning distribution map of heavy metal in soil

4 Conclusions

(1) Among the pollution index in the unit soil of the research area, 11 units of Cd are more than 1, and one of the unit surpasses the As standard, while other elements and unit pollution indexes are all less than 1; In terms of the comprehensive pollution index, there are 3 of the unit belonging to light pollution, 10 of them under the warning state, and other units are blow 0.7, which belong to the safety grade. According to the outcome of the evaluation, 99.79% of the total research area is labeled as the non-polluted area during the quality evaluation of the soil environment. Besides, the 11 polluted grid region accounts for 0.21% of the total research region, among which the number of light polluted area, and 2 middle polluted area.

(2) The damage level of the potential risky heavy metal is as follows: E（Cd）> E（Hg）>E（As）>E（Ni）>E（Cu）>E（Pb）>E（Cr）>E（Zn）. Hg, As, Cu, Pb, Cr, Zn and Ni has not achieved the risky level of the eco pollution, among which Cd element is the main reason for the influence of the eco system. 0.21% of the sampling regions are under the potential risk of Cd pollution, and other heavy metals place little or no damage to the environment. Therefore, Cd element has been taken as the important research target.

(3) Through the warning analysis of Cd element in the sampling region, we can draw the conclusion that the Cd pollution would reach 1 634.79 hm2 after 20 years, 4 050.36 hm2 after 50 years and 10 355.35 hm2 after 100 years. The total area of the agricultural field under potential risk accounts for 1.5% of the total area, mainly lying in the southeast region of the research area.

According to the risk evaluation analysis, we could estimate the potential risk confronted by agriculture field, which would set the quantitative basis for not only the administration of the soil environment and the environmental protection, but also the construction of the green agriculture product and the establishment of the heavy metals pollution as well as the proper evaluation to the agriculture field and the environment. Therefore, it is of vital importance to the rational application of the soil resource. Aiming at the outcome of the evaluation, we can draw the conclusion that part of the area is in the potential risk, and we would analyze the reason through further intensification of the sampling station; Because of the lack of investigation data, we only take the atmospheric bulk deposition into consideration in the net quantity of heavy metal Cd in the potential risk evaluation method. For the lack of the elements like fertilizing, watering and the crops, further study concerning this dynamic mechanism is in need.

References

1. Gao, X., Ma, R.: Research advances of gradation and evaluation of agricultural land in China. Transactions of the CASE 18(1), 165–168 (2002) (in Chinese)
2. Huang, X., Hu, J., Li, C., et al.: Heavy-metal pollution and potential ecological risk assessment of sediments from Baihua Lake, Guizhou, P.R. China. International Journal of Environmental Health Research 19(6), 405–419 (2009) (in Chinese)
3. Ahnstrom, Z.S., Parker, D.R.: Development and assessment of a sequential extraction procedure for the fractionation of soil cadmium. Soil Sci. Soc. Am. J. 63(6), 1650–1658 (1999)
4. Fernandes, H.M.: Heavy metals distribution in sediments and ecological risk assessment: The role of diagenetic processes in reducing metal toxicity in bottom sediments. Environ. Pollut. 97(3), 317–325 (1997)
5. Donald, L.S.: Toxic metals in the environment: the role of surfaces. Elements 1(4), 193–197 (2005)
6. Casper, S.T., Mehra, A., Farago, M.E., Gill, R.A.: Contamination of surface soils, river water and sediments by trace metals from copper processing industry in the Churnet River Valley, Staffordshire, UK. Environ. Geochem. Health 26(1), 59–67 (2004)
7. Hakanson, L.: An ecological risk index for aquatic pollution control: a sedimentological approach. Water Research 14(8), 975–1001 (1980)
8. Sznopek, J.L., Goonan, T.G.: The materials flow of mercury in the economies of the United States and the world, USA Geological Survey Circular 1197, USA Geological Survey (2000)

9. USEPA, Review of ecological risk assessment methods EPA/230/1088/041 (1988)
10. GB15618-1995, Environmental Quality Standard for Soils. Standards Press of China, Beijing (1997) (in Chinese)
11. Qifeng, T.: A study of cadmium ecological security forecasting and early warning in agroecosystem of the Chengdu Economic Region. Sichuan. China University of Geosciences (2007) (in Chinese)

The Comparative Analysis of Spatial Structure of Ji Wheat 22 Yield Based on Different Stochastic Samplings

Yujian Yang[*] and Xueqin Tong

S & T Information Engineering Technology Center of Shandong Academy of Agricultural Science, Information center of agronomy College of Shandong University Jinan 250100, P.R. China
yyjtshkh@gmail.com

Abstract. The spatial variation of crop yield was mainly caused by the complicated factors, such as soil and land factors. Based on sound theoretical reasoning, the paper performed an exploratory investigation of the relationship between the spatial distribution of yield measurement about 849 samplings on Thiessen polygon. Reasonable sample size or sampling density is an assurance of yield estimation surface. To investigate the spatial structure of Ji wheat 22 yield, we developed the comparison analysis of surface interpolation precision by stochastic samplings of 568, 425, 212 and 144 from a total of 849 samplings in zoning area, the other emphasis of the study was explored the influence of the distribution change of yield with the threshold of samplings numbers for the higher estimation precision. We first detected the presence of global spatial autocorrelation characteristics about the distribution of the Thiessen polygon region generated yield measurement samplings, then with the support of IDW interpolation, the higher precision surface was analyzed and obtained by the comparison of 5 samplings mode. High yield Thiessen polygon region of Ji wheat 22 tended to be clustered close to the high yield polygon region, the low yield Thiessen polygon region tended to be clustered close to the low yield polygon region. It is important to estimate the yield distribution for sampling density, samplings method, interpolation methods and sample size, we suggested that a suitable sample size of about 200 samplings should be used for Ji wheat 22 yield developed from the sampling zoning to acquire the higher surface estimation precision, beyond or the samplings number, the estimation precision presented the fluctuation characteristics.

Keywords: Ji wheat 22, Thiessen polygon, Autocorrelation, IDW, Yield.

1   Introduction

Many studies have been conducted in recent years about all kinds of regional variance. The sampling intensity and the position of sample points are critical to obtaining good predictions and for constructing maps from sample data, provide the context for the development of geostatistics, at the same time, Webster also pointed out that it

---

is essential to develop the consistence study about the sampling density and the scale of spatial variation for spatial prediction map[1]. For three elements of spatial sampling scale, including sampling extent, sampling spacing and sampling support, different sampling numbers create different influence of estimation precision. About the analogy studies, in Nanjing soil institute of CAS, correlated researcher developed the effects of sample size on spatial characterization of soil fertility properties in an agricultural area the Yangtze River Delta Region, results showed that sample size of 250 is suitable for soils to develop the spatial variability of soil fertilizer indices on country level. In China Agriculture University, Zhang fusuo also explored how the change of sampling numbers have impact on the spatial variability of soil nutrients, the case of Wuhu city, in Anhui province[2,3].

Under soil-crop system framework, crop yield has the more complicated characteristics than the soil characteristics. The spatial variability of yield on the regional scale is the hotspot of soil-crop system study and integration demonstration. Though RS technology became the better data of crop yield estimation, application of a kind of method, theoretically, interpolation methods are also essential for the surface formation of crop yield, correlated study about crop growth information and crop yield of samplings is rarely according to the study literatures using interpolation methods. Crop information is a close relationship to the climatic conditions, soil characteristics and agriculture cultivation status. Soil characteristics has the spatial structure, so crop information including crop yield has the spatial structure. An important objective was developed to explore the spatial structure of wheat yield in the paper, taking the example of Ji wheat 22 cultivar in demonstration base of precision agriculture in Shandong province. Obviously, the spatial variation of Ji wheat 22 yield was caused by complicated factors, the spatial variability of soil and land factors are only a kind of parts, other factors, the spatial variability of weather factors and pests and diseases, the temporal variability, the prediction variation, and so on. A detailed description of crop yield can be found in the relevant literatures(Blackmore,1996)[4,5].

Wollenhaupt explored the comparative study of soil characteristics with the support of Kriging and Inverse Distance Weighted method, the study results that the interval of 31.8m with the gridding samplings in the region is suitable, the higher interpolation precision for IDW than Kriging. We developed the spatial estimation comparison study of yield was developed on the precision scale, with the support of Kriging and Inverse Distance Weighted (IDW) method, referred to in document[4], the study indicated that IDW method is more accurate on basis of the sampling sites data than Kriging method, under less than 40m interval between sites for Ji wheat 22 yield, though Kriging method has merits, it not only provides a means of predicting values on a surface given sample data, but also confidence intervals can be attached to each prediction using the prediction standard errors, the other advantage of Kriging is estimating weights for locations based not only on the distance between separated locations but also on the global semivariogram model fitted to the collected data. The results also showed that the estimation precision depended on the threshold of sampling numbers, if the numbers are less than the threshold, the sampling density is in contact with the spatial estimation precision, if the numbers are more than the threshold, the density of the sampling numbers is improved, but the estimation results are not proportional to the density, the detailed contents referred to document [6].

In the study, we discussed the spatial structure of Ji wheat 22 yield based on IDW, firstly, we developed the spatial autocorrelation characteristics based on 849 yield measurement samplings on Thiessen polygon region, then we analyzed the spatial structure of Ji wheat 22 yield in detail, based on different samplings methods, stochastic samplings of 568, 425, 212 and 144 from a total of 849 samplings, which not only the study extension of the spatial structure of Ji wheat 22 yield, but also the typical cases of spatial clustering of the crop yield on the regional scale. Exploring the reasonable samplings numbers on precision scale, and inducing the threshold of sampling density and spatial variability scale, the paper ultimately developed the interpolated map precision of different samplings under the random samplings mode, the optimization samplings numbers for the higher continuous surface.

2   Material and Methodology

2.1   Study Area and Ji Wheat 22 Cultivar

Winter wheat (*Triticum aestivum L.*) is one of the most important crops in China, is planted in about in north-China plain, and it has the highest yield and the largest consumption, its cultivation area is about 22% of the total crop area, and its yield is 20% of the total crop production. Especially, the 65.4% of the total wheat cultivation area, the 75.5% of the crop yield, respectively. Ji wheat 22 was intensively planted in Lingxian country, which is located in the north china, is characterized by a warm temperate monsoon climate, the precipitation and temperature presents the space-time variability, an average annual precipitation of 900mm in the south of sampling area and 480mm in the north of sampling area, the 80% of annual precipitation distributed from June to September and the soil is formed from the sediments of the Yellow River, so the irrigation is the crucial factor for agriculture development. Typically, agriculture in the area is intensified by a double cropping system with a high-yield cultivar and high fertilizer and water inputs. Winter wheat, Ji wheat 22 and summer maize is a typical double cropping system a year in the study area, and the ratio of the total yield of two kinds of crops is about 1/5 in China.

As super high yield wheat cultivar by cross breeding, Ji wheat 22 cultivated by Shandong Academy of Agricultural Science was absolutely planted predominance in the study zoning area, located in Lingxian country, precision demonstration base and the commodity demonstration base of China.

2.2   Experimental Sites and Data Acquisition

The Ji wheat 22 yield obtained from the actual yield of the observation area, with the support of DGPS, the area covered the zoning area about 4 kilometers×2 kilometers, 1 square meter area was selected around DGPS location site, the wheat was bounded and weighted using the suitable balance, yield information involved in plant height, the total spike numbers of Ji wheat 22, spike numbers per 666.7 m^2, grain number per spike, 1000-grain weight, yield per 666.7 m^2 and biological yield, the above contents were measured and described in the yield measurement process, the corresponding

per 666.7 m^2 yield of Ji wheat 22 was transformed in yield per hectare, and a total of 849 yield measurement sampling sites of Ji wheat 22 were obtained.

Sampling plans has a large effect on the continuous surface formation, stochastic sampling does not ensure even coverage of the area to be sampled, so gridding sampling combined with stochastic samplings in the study, a total of 849 sampling sites of Ji wheat 22 yield measurement, stochastic samplings of 568, 425, 212 and 144, according to the ratio of 2/3,1/2,1/4,1/6 from a total of 849 samplings, as illustrated in Fig. 1.
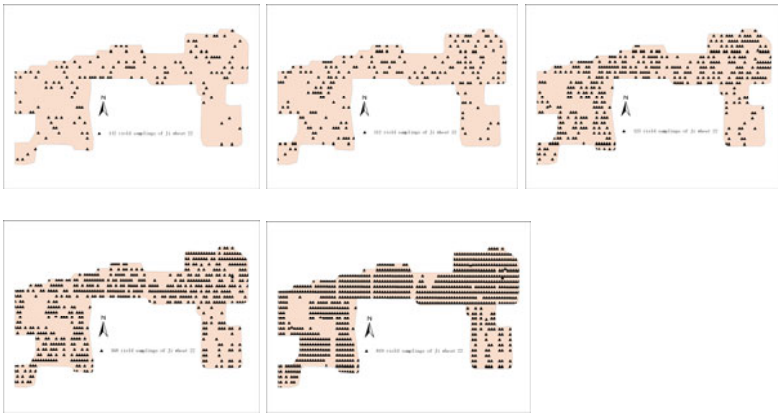
Fig. 1. Distribution of yield measurement sites on precision scale on 5 kinds of sample modes

The statistical characteristics was analyzed, results showed the mean value of wheat yield is 6716.3kg/hectare, the maximum value is 10364.1kg/hectare, the minimum value is 1828.0kg/hectare, the range of yield was 1828.0kg/hectare-10364.1kg/hectare. The above description explained the characteristics of wheat yield rather than the regional characteristics of the whole sampling area. Coefficient of variation of was 0.2 which suggested wheat yield belonged to the medium variation. Consequently, it is necessary to apply spatial interpolation methods together with GIS to solve this problem and reveal the spatial structural characteristics.

2.3   Inverse Distance Weighting Method

Inverse Distance Weighting (IDW) are interpolation techniques in which interpolated estimates are made based on values at nearby locations weighted only by distance from the interpolation location. IDW make assumptions about spatial relationships except the basic assumption that nearby points ought to be more closely related than distant points to the value at the interpolate location. The formula used for Inverse Distance Weighting is:

$$Z_j = \frac{\sum \left[z_i \Big/ (h_{ij} + s)^p \right]}{\sum \left[1/(h_{ij} + s)^p \right]} \tag{1}$$

Where Z_j : estimated value for location j ; z_i :measured sample value at point i ; h_{ij} :distance between Z_j and z_i ; s :smoothing factor; p :weighting power. The weighting power p defines the rate at which weights fall off with h_{ij}, the distance between the interpolated and sample locations. A value of 1-5 is typical, the smoothing factor s reduces the likelihood that any one sample value will overly influence an estimated value for a given interpolation location. IDW is an exact interpolator, so where an interpolation location j coincides with a sample location i, $Z_j = z_i$ and a sharp "peak" or "valley" may result; setting $s > 0$ reduces this peaking effect when it occurs[7].

IDW emphasized that attribute values close in distance terms tends to be similar but that the similarity weakens as distance separation increases, which is consistent for the Tobler Law by estimating unknown measurements as weighted averages over the known measurements, "Everything is related to everything else, but near things are more related than distant things"[8]. IDW give differential weights to observations based on their proximity to the missing value, it is the nearest sites that should be given most weight in any imputation, so the method estimators depend only on geometric relationships between the sites and the estimates they yield are dependent on neighboring size and how the weights are specified and the method do not provide an estimate of the possible error associated with the imputation.

3   Results and Analysis

3.1   Spatial Autocorrelation Characteristics

The presence of spatial dependence means that values for the same attribute measured at locations that are near to one another tend to be similar, and tend to be more similar than values separated by larger distances. We construct the Thiessen polygons to develop spatial autocorrelation characteristics of 849 yield measurement points. Thiessen polygons can be used to apportion a point spatial file into regions known as Thiessen or Voronoi polygons, and each region contains only one input spatial file point, all points are triangulated into a triangulated irregular network (TIN) that meets the Delaunay criterion. From the principle, the perpendicular bisectors for each triangle edge are generated, forming the edges of the Thiessen polygons, the locations at which the bisectors intersect determine the locations of the Thiessen polygon vertices. On the basis of Thiessen polygons construction process, each region has the unique property that any location within a region is closer to the region's point than to the point of any other region. The Thiessen polygon of 849 yield measurement sites generated was displayed in Fig. 2.

The spatial autocorrelation function, examines the correlation of a random process with itself in space, not only showing the correlation between a series of points or a map and itself for different shifts in space or time, but visualizing the spatial variability of the phenomena under study. In general, large numbers of pairs of points that are close to

each other on average have a lower variance than pairs of points at larger separation, the autocorrelation quantifies this relationship and allows gaining insight into the spatial behavior of the phenomenon under study clarify the spatial interactions of model, so we need to understand the spatial connectivity between each country unit.

The spatial weight matrix is the fundamental tool used to represent the spatial connectivity between regions units, based on constructing a *K*-nearest neighbouring weights file. For the spatial weight matrix W(k) which we use in this study is based on the *K*-nearest neighbouring function rules, *K*-function is a second-moment measure as it is closely related to the second-order intensity of a stationary isotropic point process, it captures the spatial dependence between different country regions of the point process, in a short word, the general form of the k-nearest neighbouring weight matrix W(k) is defined as the formular (2):

$$w_{ij}(k) = 0 \, if \, i = j$$
$$w_{ij}(k) = 1 \, if \, d_{ij} \le d_i(k) \, and \, w_{ij}(k) = w_{ij}(k) / \sum_j w_{ij}(k) \ . \quad (2)$$
$$w_{ij}(k) = 0 \, if \, d_{ij} > d_i(k)$$

Where $d_i(k)$ is a critical cut-off distance defined for each region i. More precisely, $d_i(k)$ is the kth order smallest distance between regions i and j such that each region i has exactly k neighbors. We use the *K*-nearest neighbor weight matrix and we check for robustness in the study[9].
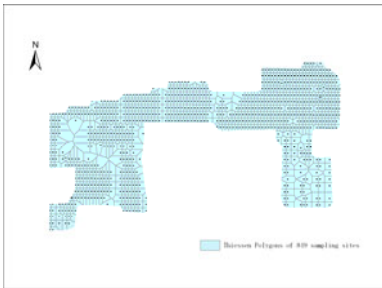
Fig. 2. Thiessen polygon of 849 yield measurement sites (left)
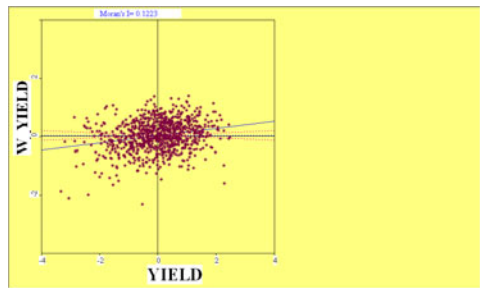
Fig. 3. Moran scatterplot of the yield measurement in 2010 (right)

The Moran's measurements which indicated spatial association of the Ji wheat 22 yield was positively spatially autocorrelated (Moran's I value is equal to 0.1223, mean value in theory is equal to 0.0004, the experience mean value is equal to -0.0012, the experience standard deviation value is equal to 0.0239) with 499 times Monte-Carlo tests, and the statistics are significant with p=0.002 for 849 sampling sites. Moran's I coefficient referred to Fig. 3. Results indicated that yield of Ji wheat 22 cultivar presented the positive autocorrelation, high yield Thiessen polygon region

of the cultivar tended to be clustered close to the high yield polygon region, the low yield Thiessen polygon region tended to be clustered close to the low yield polygon region. According to the Thiessen polygon of four quadrants of Moran scatter plots, from univariate local spatial autocorrelation results of local Moran`I computerization, 98 Thiessen polygon regions which presented the autocorrelation is significant characteristics in confidence interval of 95%.

3.2   Spatial Structure Analysis of Wheat Yield

Spatial autocorrelation of yield measurement mapped the spatial clustering, to detect spatial structures of distributional yield samplings, the continuous surface results should select interpolation method to solve. Kriging is estimating weights for locations based not only on the distance between separated locations but also on the global semivariogram model fitted to the data. But Estimation precision of IDW is higher estimation precision than Kriging method in author`s other paper study[6]. IDW are affected by clusters in the data points, the autocorrelation characteristics of yield polygon analyzed the outlier tendency, satisfied the outlier stationary condition for IDW interpolated map. We obtained and developed the comparison analysis of surface interpolation precision by stochastic samplings of 568, 425, 212 and 144 from a total of 849 samplings, referred to Fig.4. On the whole, with the samplings increasing, from 144, 212 to 425,568,849, the interpolation surface presented the corresponding changes, the interpolation surface of wheat yield showed that the lower yield located in the east of the middle part, the higher yield located in the south of the middle part, the total distribution of wheat yield, the groovy regions, wheat yield is higher in the west of the region than the east of the region. IDW estimation of 144 samplings of yield measurement presented the fairly thick distribution. However, the IDW estimation of 212 yield measurement has the relatively thinner distribution and the better optimization pattern, the different block shape, which explained the crucial points of estimation precision caused by the samplings number change. From the interpolation surface of 144 samplings, of the estimation distribution, which reflected the high yield or low yield in different areas, indicated that the change of samplings numbers has the key impact on the estimation precision. Maps generally provide some interesting hotspot, the obvious high value area of block shape located in the north of the west part in the zoning area, centered in the high yield area, the continued low value distributed extending the south part. For the interpolation surface of 212 samplings, the high value area still is the same area of surface of 144 samplings, with a difference, the block scope is smaller than the surface of 144 samplings, the secondary high value area is clear. The new content was described as the unobvious crossing area of "high yield and low yield" for 212 samplings estimation, or no sudden changes, but the two high yield area of 144 samplings surface, and the gradient changes from the high yield area, the low yield area is located in the south-east part of estimation area, the relatively moderate estimation distribution was presented, which showed the nonsensitive estimation of IDW method for the high and low observed points, for continuous surface based on different random samplings(144,212,425,568), there existed in relatively moderate distribution mode, no sudden changes situation[10].
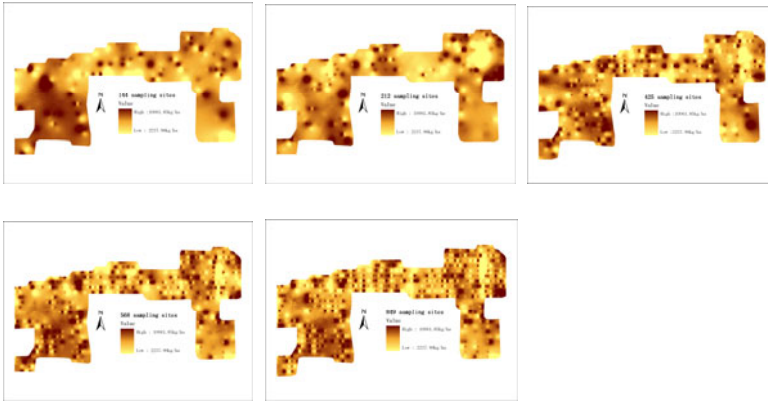
Fig. 4. Spatial distribution map of yield measurement sites based on 5 kind of modes

The cross-validation method which mainly described the sampling estimation required from the other samplings interpolation results was used to the comparative study of the prediction value and the actual value, the computerization of MSE was created by the way in the study, the smaller MSE value, the better prediction of the samplings estimation, with a difference, The correlation of prediction value and observed value is an important index, the larger, the better[3]. Based on the IDW interpolation method, the MSE of different samplings (144,212,425,568) interpolation results has the small difference. Moreover, the analysis of the Pearson`s correlation coefficient between predicted value and observed value of different samplings was computerized by the R software tool in the study, the results indicated that the value is equal to -0.051, 0.211, 0.01 and 0.039 for 144 measurement samplings, 212 measurement samplings, 425 measurement samplings and 568 measurement samplings of Ji wheat 22 yield in confidence interval of 99%, respectively. According to this analysis, 0.211 was regarded as the better correlation characteristics in the study, which is regarded as the better prediction precision of 212 samplings, or the number of samplings is probably the threshold of the low prediction value and the high prediction value, means that the better significance characteristics over the threshold. But the value of Pearson`s correlation coefficient is equal to 0.113 of the prediction value and observed value on basis of 849 samplings measurement of Ji wheat 22 in the study, showed the fluctuation was embodied by the estimation precision with the measurement yield samplings increasing. The results of correlation coefficient of prediction value and observed value about the different measurement numbers illustrated that the non-linear relationship is between the estimation precision and samplings numbers, and the threshold of sampling numbers is also the best estimation precision for the prediction areas, the more samplings cannot indicate the higher estimation for the special sampling regions, the suitable, optimization sampling plans and numbers is the only law to the best precision of the estimation precision.

Theoretically, the more samplings number, the better estimation precision, the samplings cost constrained the sampling numbers, so the optimization sampling numbers attracted the attention of researchers. Different random samplings

(144,212,425,568,849) predicted the yield surface distribution, with the increasing sampling numbers, the larger of the correlation of prediction value and observation value, but the threshold existence between sampling numbers and estimation precision, beyond threshold of sampling numbers, the increasing scope decreased, below threshold of sampling numbers, the correlation change scope increased. As the above mentioned, IDW interpolation results indicated that if the sampling numbers are less than the threshold, the sampling density is in contact with the spatial estimation precision. If the numbers are more than the threshold, the density of the sampling numbers is improved, but the estimation results are not proportional to the density. Considering the estimation precision and cost at the fields, the suitable yield measurement sampling numbers should be 200, or is regarded as the threshold, which has the higher estimation precision of yield measurement samplings in zoning area[1,10].

4   Conclusions and Discussion

The study revealed the presence of legitimate structures in the yield data in zoning area, which involved in the discrete valued how extended the continuous surface. The principal focus of this work was the comparative analysis and threshold of samplings data interpolated continuous surface for yield measurement, and the paper did not deal with otherwise important issues. During the interpolation, the method which is "best" depends on the yardstick we choose, using all the nearby sites and accounting for the possibility of outlier is the better selection, IDW estimators depend only on geometric relationships between the sites and the estimates yield dependent on neighboring size and how the weights are specified and the method, which improved the outliers[10].

The sample data between the sample data points and the locations ignored the spatial correlation or dependence meaning the large prediction errors[11]. Spatial autocorrelation of Ji wheat 22 yield solved the problem, results showed yield of Ji wheat 22 cultivar presented the positive autocorrelation, high yield thiessen polygon region of the cultivar tended to be clustered close to the high yield polygon region, the low yield thiessen polygon region tended to be clustered close to the low yield polygon region. 98 thiessen polygon regions which presented the autocorrelation are significant aggregation characteristics in confidence interval of 95%.

From the theory and method, the interpolation surface was discussed from different samplings(144,212,425,568,849) in the study, considering the estimation precision and cost at the fields, the higher precision of IDW interpolation comparison results of Ji wheat 22 yield, the suitable samplings number of 200 in the zoning area about 4 kilometers×2 kilometers, or the threshold of sampling numbers constrained the estimation precision, over 200, the higher estimation precision is not presented in the interpolation surface, the fluctuation precision changes, which probably is a close relation of requirements any method for estimating or predicting values on a spatially continuous surface, such as data structure, sampling density, experimental area shape.

Choosing models and fitting them to data remain among the most controversial topics in creating the continuous surface, Bayesian model should be probably better solved the issues, hierarchical bayesian model and bayesian maximum entropy

(BME). Bayesian method makes a decision according to the posterior probability distributions combination with the probability density function and prior probability distributions, the method has merits which taken into account the error of sample estimation and prior probability and knowledge. The estimation of Bayesian maximum entropy (BME) has a set of criteria and the relative theoretical merits, spatial-temporal covariance models available was taken into account, such as covariance variogram and semivariogram, ordinary and generalized, separable and non-separable, at the same time, the spatiotemporal structure of the yield distribution combination with the temporal variation was adequately represented[13,14].

References

1. Webster, R., Oliver, M.A.: Geostatistics for Environmental Scientists. John Wiley, Chichester (2001)
2. Wang, Z.-G., Zhao, Y.-C., Biao, H., Jeremy, D., Sun, W.-X.: Effects of Sample Size on Spatial Characterization of Soil Fertility Properties in an Agricultural Area of the Yangtze River Delta Region, China, Soils. 42(3), 421–428 (2010)
3. Sun, Y., Wu, C., Zhu, K., Cui, Z., Chen, X., Zhang, F.: Influence of interpolation m ethod and sampling number on spatial prediction precision of soil P. Chinese Journal of Applied Ecology 20(3), 673–678 (2009)
4. Blackmore, S.: Precision farming, an overview. Agriculture Engineering, 85–87 (Autumn, 1997)
5. Blackmore, S., Godwin, R.J., Taylor, J.C., et al.: Understanding variability in Four Fields in the United Kingdom. In: Processing of the 4th International Conference on Precision Agriculture, Paul, July 19-22, pp. 3–18 (1998)
6. Yang, Y., Liu, S., Feng, W., Shang, M.: Comparison Study on the Spatial Estimation of Ji Wheat 22 Yield on the Precision Scale. In: 2011 3rd International Conference on E-business and Information System Security Proceedings, IEEE Catalog Number:CFP1129F-PRT, Special Session Forestry&Agriculture Engineering and Computer Science, Produced, pp. 576–579. IEEE eXpress Conference Publishing (2011)
7. Schabenberger, Oliver, Gotway, Carol, A.: Statistical Methods for Spatial Data Analysis Texts in Statistical Science. CRC Press (2005)
8. Tobler, W.R.: A computer movie simulating urban growth in the Detroit region. Economic Geography 46(2), 234–244 (1970)
9. Russo, G., Reggiani, A., Nijkamp, P.: Spatial activity and labour market patterns: A connectivity analysis of commuting flows in Germany. Ann. Reg. Sci. 41, 789–811 (2007)

10. Haining, R.: Spatial data analysis: theory and practice. Cambridge University Press (2004)
11. Yang, Y., Yang, J., Li, S., et al.: Comparison of spatial interpolation methods for maize growth period. Transactions of the CSAE 25(9), 163–167 (2009) (in Chinese)
12. Yang, Y., Zhu, J., Zhao, C., Liu, S., Tong, X.: The spatial continuity study of NDVI based on Kriging and BPNN algorithm. Mathematical and Computer Modelling 54, 1138–1144 (2011)
13. Yu, H.-L., Kolovos, A., Christakos, G.: Steve War-merdam and Boris Dev. Interactive spatiotemporal modelling of health systems: the SEKS–GUI framework. Stoch Environ Res Risk Assess (2007)
14. Wang, et al.: Hand, foot and mouth disease: spatiotemporal transmission and climate. International Journal of Health Geographics 10, 25 (2011)

Research on Throwing Soil Regular Pattern of Reversal Cultivated Land and Fertilization Seeder

Yongliang Zhang, Jianping Hu, Chunjian Zhou, and Chuantong Lu

Key Laboratory of Modern Agricultural Equipment and Technology,
Ministry of Education & Jiangsu Province, Jiangsu University,
High-tech Key Laboratory of Agricultural
Equipment & Intelligentization of Jiangsu Province,
Zhenjiang 212013, China
yy19860922@126.com

Abstract. The distribution of soil particles and planting tube space position affects reversal cultivated land and fertilization seeder seeding performance. Through the establishment of rotary tiller system coordinate system; Firstly giving the analysis of soil grain not heel cover shell collision and the first meeting with cover shell collisions motion situation, through Matlab found the first meeting with cover shell collisions will collision with spin plow knife again, and calculated separately spin plow knife in a horizontal position, tangent 45 ° direction and in vertical direction of three position of the collision curve. Findly, fitting total soil particle flow distribution curve. According to sow requirements, for agriculture combined machine seeding tube decorate location provides an important basis.

Keywords: Soil particle flow, Distribution curve, Second collision, Sowing tube, Matlab.

1   Introduction

Working principle of reversal cultivated land and fertilization seeder as shown in figure 1, is 1GHB-175 agriculture combined machine using up-cut rotary way, no open ditch the way the homework, throw the soil particles into the sowing fertilizer tube of behind to cover seeds and fertilizers. According to the soil particle distribution laws, the reasonable layout of sowing pipe line, in order to realize the deep fertilization, shallow the purpose of sowing.
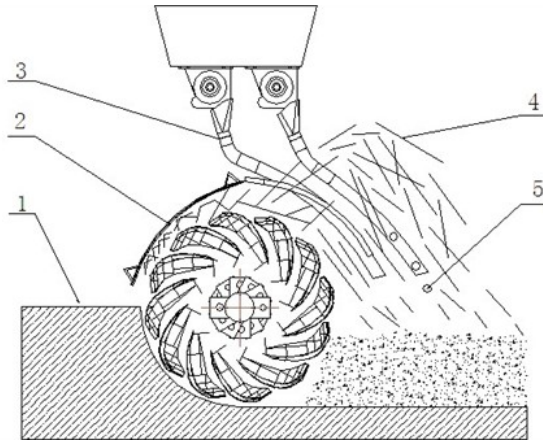
1. Fallow land 2. Cover shell 3. Row fertilizer tube 4. Being thrown around the soil particles
5. Seeds

Fig. 1. Agriculture combined machine working principle diagram

2   Pattern of Reversal Cultivated Land Soil Grain Distribution Calculation

Knife shaft speed for 270r/min, cover shell radius is 356mm, angle is 70. From the macro analysis, will spin tillage soil into forward throwing and backward throwing two circumstances.using calculation of back-throw ratio of soil for up-cut rotary cultivation cast soilprocess simulation program, 86.32% soil grains not with cover shell collision and then thrown back, the rest of the soil grains collision with cover shell and throw in front of the machine. In order to guarantee operation quality, must be theory discussion and analysis for the two parts are soil particles. Draw the total cast soil trajectory fitting curve, then determines the resettlement of sowing pipe line position.

2.1   No Collision Soil Particle Movement Analysis

soil for up-cut rotay cultivating research, when regardless of the air resistance, was thrown soil particles motion equation is:

$$\begin{cases} x = x_o + v_x t \\ y = y_o + v_y t - \dfrac{1}{2} g t^2 \end{cases} \tag{2.1}$$

According to the condition of soil particle threw back :

$$\psi > \arctan \frac{v_y}{v_x} > \pi / 2$$

Can get curve equation of soil particle flow which was not collision with the cover of shell:

$$y_2 = -\frac{(x-40)^2}{28.5714} + 56 \tag{2.2}$$

In the formulas

x_o　y_o——The soil was thrown the initial position coordinates;

v_x　v_y——The soil was thrown initial velocity;

g——Gravity acceleration;

ψ——The velocity v direction angle

2.2　The First Meeting with Cover Shell Collisions Motion Situation Analysis

According to the actual size of this model, establish spin plow knife static coordinate system xoy, make the shaft center of spin plow knife as coordinates xoy origin, same as the positive direction of x axis and machine driving direction. Establish the polar coordinate of cover shell $o_1 x_1$, radius is $r = 356mm$, the center of the circle at o_1, extremely angle is $a = 8°$, cover shell angle is $\beta = 70°$, o_1 and o_2 position parameter is l_x and l_y. Its space arrangement as shown in figure 2.
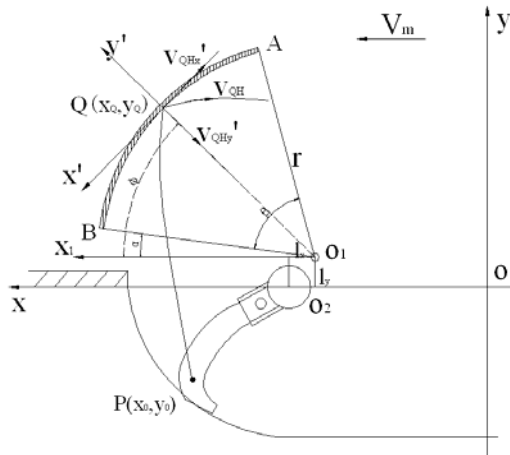

Fig. 2. Curved shell with knife shaft rotary center position relations and collision motion analysis

Assumptions are thrown soil particles point $S(r,\phi)$ and Q superposition, the coordinates in static coordinate system is S(x,y); collision point is Q, the normal and tangent pass the collision point Q and forming the Dynamic coordinate system. According to dynamic simulation of the collision of throwing soil with camber cover of rotary plow research : known cover shell parameters and soil particles being thrown around the initial velocity and direction angle, get after collision trajectory:

$$S' = \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{cases} (1-f)[(v_{ox}+v_m)\sin\phi - \sqrt{v_{oy}^2 - 2g(y_Q - y_0)}\cos\phi]\sin\phi \\ -(1-f)[(v_{ox}+v_m)\sin\phi - \sqrt{v_{oy}^2 - 2g(y_Q - y_0)}\cos\phi]\cos\phi \end{cases}$$

$$\left. \begin{matrix} -k[(v_{ox}+v_m)\cos\phi + \sqrt{v_{oy}^2 + 2g(y_Q - y_0)}\sin\phi]\cos\phi \\ -k[(v_{ox}+v_m)\cos\phi + \sqrt{v_{oy}^2 + 2g(y_Q - y_0)}\sin\phi]\sin\phi \end{matrix} \right\} t' + \begin{bmatrix} x_Q \\ -\dfrac{1}{2}g(t')^2 + y_Q \end{bmatrix} \quad (2.3)$$

2.3   Determine Trajectory of Soil Particle Flow Which Collision with Spin Plow Knife Again

All parameters in generation going into the type 1.3, use Matlab calculation the curve track when collision extremely sngle in 10 °, 20 °, 30, 40 °, 60 °, 50 °, ss shown in figure 3.
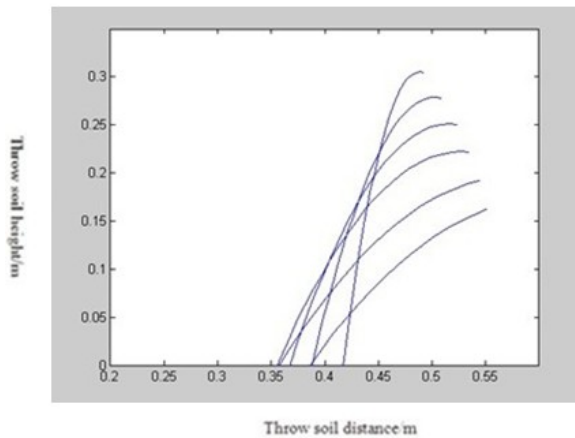

Fig. 3. Soil grain after collision with cover shell trajectory

We know from figure 3 soil grain fell on 0.245m-0.490m between will collisions again with spin plow knife and thrown by high speed. according to the soil particle collisions cover shell movement analysis, similarly establish mathamatics grains and spin the plow knife collision coordinate system x_2Qy_2, as shown in figure 4 shows spin plow knife is 45°scene, the main analysis the soil particles which was thrown back by spin plow knife.

Side blade of Spin plow knife curve equation is $R = R_o(1 + K\theta)$

Among them

R_o —Helical starting radius;
R —The radius of turn angle θ;
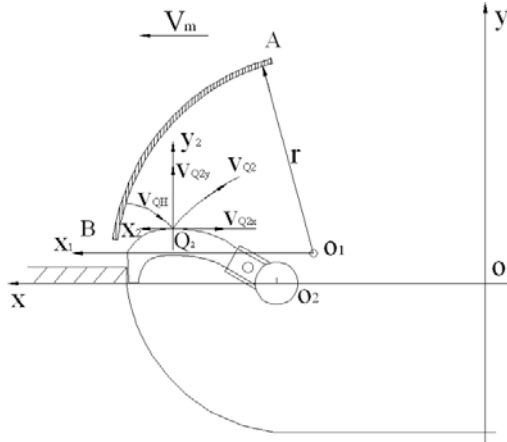K —coefficient $K = 2$

Fig. 4. Collision analysis between soil grains and Spin plow knife

In this paper, using spin plow knife rotation radius is 245mm, side blade of spin plow knife spirals radius $R_O = 120mm$, therefore, calculate the points Q_2 location on the static coordinates xoy for (476.67 63.46), namely $x_{Q2} = 476.67$, $y_{Q2} = 63.46$.

After collision, tangential velocity that along side blade of spin plow knife point Q must consider the influence of friction coefficient. namely:

$$\begin{cases} v_{Q2x} = (1 - f)v_{QHx} \\ v_{Q2y} = -kv_{QHy} \end{cases} \tag{2.4}$$

In the formula
f —Collision friction coefficient 0.4
k —Collision recovery coefficient 0.4

After soil particle S collisions spin plow knife in static coordinate system xoy will starting with Q_2 (x_{Q2}, y_{Q2}), according to v_{Q2} for initial velocity continue to do parabola, trajectory equation for

$$S'' = \begin{bmatrix} x'' \\ y'' \end{bmatrix} = \begin{bmatrix} v_{Q2x}t'' \\ v_{Q2y}t' - \frac{1}{2}gt''^2 \end{bmatrix} + \begin{bmatrix} x_{Q2} \\ y_{Q2} \end{bmatrix} \tag{2.5}$$

Generation into the parameters, Matlab calculation spin plow knife end tangent in a horizontal position, tangent 45 ° direction and vertical direction when the soil whereabouts of three position curves. As shown in figure 5, 6, 7
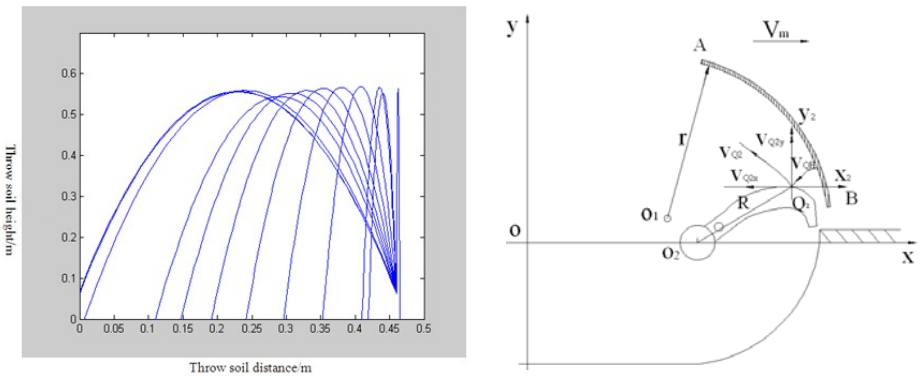
Fig. 5. Horizontal position soil distribution curve and spin plow knife position

That from figure 5 we know that most of the soil distributed in spin plow knife rotating scope inside will collision with spin plow knife again.
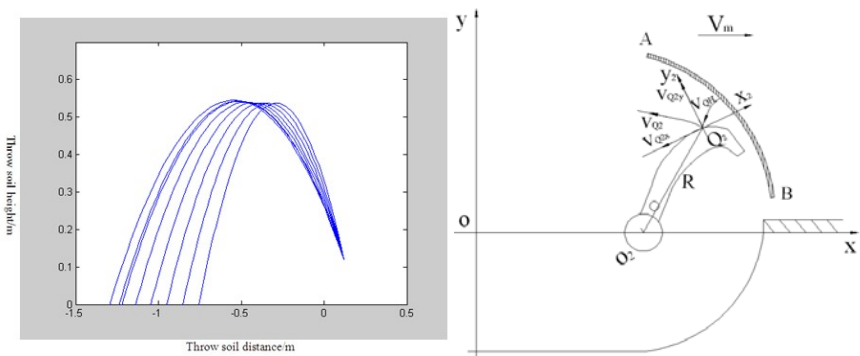
Fig. 6. Tangent 45 ° direction soil distribution curve and spin plow knife position

That from figure 6 we know that soil particles is mainly distributed in between 0.75m and 1.3m.
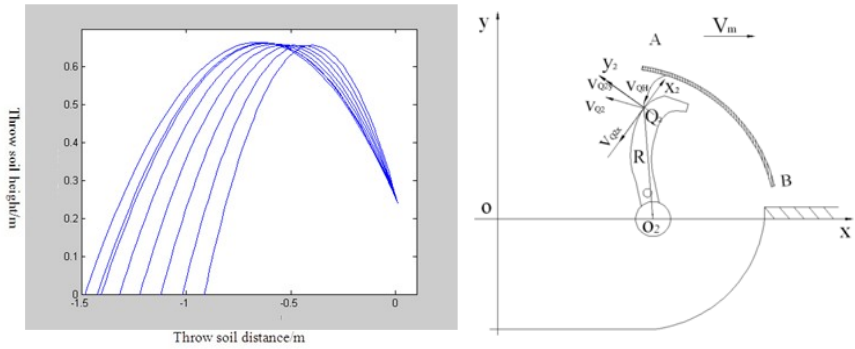

Fig. 7. Vertical direction soil distribution curve and spin plow knife position

That from figure 7 we know that soil particles is mainly distributed in between 0.85m and 1.5m   According to type 2.2, assuming soil grain evenly distributed, its distribution in between 0 and 0.8m. combining figure 5, figure 6 and figure 7 knowable, the collision happened twice almost entirely throw at the soil particles that no collided rear.

3　　　Decorate Sowing Tub

From the analysis of soil grain not heel cover shell collision and the first meeting with cover shell collisions motion situation fitting two parts soil particle flow distribution, its distribution curve shown as shown in figure 8
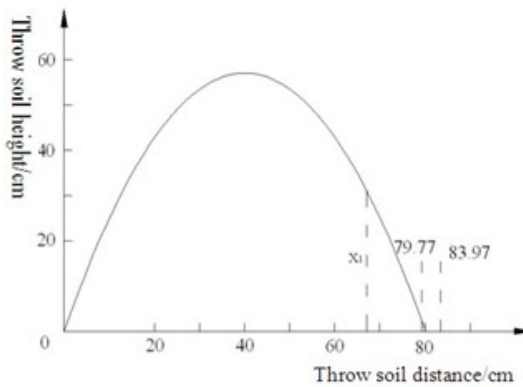

Fig. 8. Throw the soil trajectory curve, sowing decorate position and throw seeds range

The machine's spin plow depth is 15cm, planting depth is 1cm-3cm, so 6.7% to 20% soil cover seeds. Assuming thrown soil particles evenly distributed, according to the figure 8 can calculate the most distant soil particles is 6.32%, assuming critical point 6.32% coordinates is x1, through solving equations:

$$\int_{x_1}^{80}[-\frac{(x-40)^2}{28.5714}+56]dx = 2986.667\times6.32\% \tag{3.1}$$

Can get $x_1 = 67.746$, figure 8 shows, if the pipe line at the bottom port layout online x_1. The seed will fall in between 79.77cm and 83.97c, namely seeds in the soil amount covered 1.5cm and 2.0cm, achieving the desired between planting depth.

4   Conclusion

1. No collision soil particle movement analysis, get curve of soil particle flow which was not collision with the cover of shell.
2. The first meeting with cover shell collisions motion situation analysis, give trajectory and description.
3. Use Matlab calculation the first meeting with cover shell collisions motion situation, through Matlab found the first meeting with cover shell collisions will collision with spin plow knife again. Found the collision happened twice almost entirely throw at the soil particles that no collided rear.
4. From the analysis of soil grain not heel cover shell collision and the first meeting with cover shell collisions motion situation fitting two parts soil particle flow distribution, use of distribution curve, rational arrangement of pipe line

References

1. Yan, J., Hu, J., Fan, G.: Design of 1 GHB-175 Type of Teamwork Machine of Rototilling and Stubble-breaking and Fertilization Seeder. Farm Machinery Research (2), 74–77 (2010)
2. Chen, C., Shi, Y.: Calculation of back-throw ratio of soil for up-cut rotary calculation. Journal of Agricultural Machinery 30(3), 25–29 (1999)
3. Li, B., Chen, C.: Dynamic Simulation of the Collision of Throwing Soil with Camber Cover of Rotary Plow. Jiangsu Polytechnic University Journal (2000)

4. Gao, J., Zhou, P.: Development and test of high speed soil-cutting simulation system based on smooth particle hydrodynamics. Transactions of the Chinese Society of Agricultural Engineering, 20–26 (August 2007)
5. Scale Conservation Tillage. Conservation Tillage & Sustainable Farming, pp. 407–415. China Agricultural Science and Technology Press, Beijing (2004)
6. Molin, J.P., Basford, L.L., Von Bargen, K., Leviticus, L.I.: Design and Evaluation of a Punch Planter for No-till Systems. Transaction of the ASAE 41(2), 307–314 (1998)
7. Zhao, M., Zhao, S., Wang, C., Hao, J., Hou, H., Shi, D., Liu, H.: The Performance Test and Extension of The 2BM Model Series NO-Till Seeder. In: Conservation Tillage & Sustainable Farming, pp. 149–152. China Agricultural Science And Technology Press, Beijing (2004)
8. Sharma, R.K., Gupta, R.K., Chhokar, R.S., Mongia, A.D., Shoran, J.: Rotary Disc Drill-A Machine for Conservation Agriculture on Small Farms. In: Conservation Tillage & Sustainable Farming, pp. 247–251. China Agricultural Science and Technology Press, Beijing (2004)

A Research about the Application of Information Technology in the Precision Agriculture: Taking the Operating System of Shanghai Agriculture Economy as an Example

Hong Yao[1,2] and Yong-xing Wu[1,*]

[1] College of Resources and Environmental Science, East China Normal University, Shanghai, China
[2] College of Ecology and Tourism, Shanghai Business School, Shanghai, China
juddy_yao@163.com

Abstract. In the background of the digital agriculture advanced step-by-step, Shanghai government is also accelerating the pace of agricultural informationization. "Precision Agriculture" has become a pointcut of "Digital Shanghai". Building the Operating System of Shanghai Agricultural Economy is a meaningful practice in Shanghai digital agriculture engineering. Its powerful function of spatial data analysis and inquiry provides the support system for the realization of "precision agriculture". It can provide the decision support and the information service for the government.

Keywords: precision agriculture, GIS, the Operating System of Shanghai Agricultural Economy.

1   Introduction

With the farm produce's large scale marketing and circulation coming into being, the future sustainable development of agriculture in Shanghai will, to a great extent, be decided by the restructuring and optimization of agriculture. On this background, Shanghai's agriculture has quickened the transformation from "Quantity Agriculture" to "Quality Agriculture", from planning typed into market-oriented, from suburban into modern urban agriculture. The government has accordingly adjusted its management, further stressing the macro adjustment and the function of service. The utilization of modern information technology enables the dynamic monitoring and timely appraisal of achievement and outcome of main information of Shanghai suburban agricultural economy operation, thus providing support for relevant government departments in their macro-control and management of agriculture in a timely, scientific and efficient

---

way. Also farmers can obtain effective guiding information in their optimization of their agriculture investment, reasonable restructuring, and gaining higher yield. Obviously the application of information technology in precision agriculture is one of important measures of Shanghai Municipal Government in realizing "Precision Agriculture" for the sustainable development of agriculture.

2   The Definition of "Precision Agriculture"

The concept of Precision Agriculture or Precision Farming was put forward by some U.S and Canadian agriculture research institutions in the 1990s. Precision Farming is a modern farming operation and management system, supported by the information technology, and exerting a series of modern farming techniques and managing expertise according to variations of locations and operated in a fixture of location, timing and quantity. The basic content of precision farming is to adjust the investment of crops according to the condition of the soil, in which farmers, after confirming the soil condition and the mutation of the space of productivity, decide the objective of the farming. And in this way, by "Systematic diagnosis, optimized recipe, technical installation and scientific management", farmers can fully utilize the fertility of the soil, thus obtaining higher yield with the least and most economical investment without damaging the environment and obtaining economic and environmental benefit by efficiently utilizing various agricultural resources. Ten systems are needed for the technical support in the realization of the precision farming which include the GPS, Information Collection System for Farmland, Farmland Remote Sensing and Monitoring System, Farm GIS, Agricultural Expert System, Smart Farming Equipment System, Environment Monitoring System, Integration System, Network Management System and Training System. The core of the system is the establishment of a complete farmland GIS. Precision Farming can be said as a brand-new farming with the complete integration of information technology and agricultural production, the goal of which is not only for achieving economic efficiency but also for social and ecologic benefits.

3   The Present Situation and Problems in Precision Farming in Shanghai

3.1   The Present Situation

Since the beginning of the 21st century, the agriculture in Shanghai has entered a new phase of development and the traditional suburban farming, which only provides farm produce, is no longer suitable for the status of Shanghai as internationalized metropolis. At present the agriculture of Shanghai is undertaking the transformation from being quantitative to qualitative, from suburban-style to metropolis, from extensive farming to integrated farming, from production-styled to service-oriented farming. The changes of farming functioning and ways of farming require urgently the change of mode of agricultural production and management and service system as well. Precision farm has

become an inevitable option for Shanghai to develop its metropolitan agriculture because the information resources for agriculture in Shanghai have become multi-factored, with the characteristic of large quantity, effectiveness of timing and broadness of respects concerned. Further developing digital management of agriculture in Shanghai will have a profound sense for pushing forward precision agriculture in Shanghai with the advantages of Shanghai's human resources of GIS, software research and development and complete infrastructure of information technology.

3.2   Advantages and Disadvantages

Shanghai has the superior technical conditions and basis in realizing the precision agriculture. With speeding up construction of "digital city" and "digital Shanghai", Shanghai Municipal Agricultural Commission (SMAC) and Shanghai Municipal Agricultural and Forestry Bureau (MAFB) accelerate the construction of digital agriculture. They developed many practical works around the digital agriculture in recent years. First, the network publishing system of agricultural information has basically completed, including building the interior local area network platform of each agriculture department, developing the intranet information management system about the vegetable planting, animal husbandry and aquaculture, and including publishing all kinds of the agricultural information on web. Second, the basic research and development on sharing the basic data and technical standard has been launched. It includes the developer guide for the application system of Shanghai agricultural GIS, the agricultural information classification and coding standard in Shanghai agricultural GIS, working out the classification standard about agriculture industry in thematic mapping and the symbols norms. Third, some application projects have been carried out successively, including the dynamic monitoring system of non-point pollution control of farmland fertilizers and pesticides in Shanghai, the planning management information system of forestry resources in Shanghai, the GIS application system of animal husbandry management, etc. This series of practices on digital agriculture laid the foundation of the technology for achieving precision agriculture. In addition, there are some disadvantages in the construction of Shanghai precision farming. For example, the construction of network infrastructure is relatively inadequate in Shanghai suburbs. There are great differences on the Information infrastructure in different areas, lack of integrated talent in the agricultural information field. And all kinds of application systems of agricultural informationization also lack sharing mechanism or can't be integrated effectively. These problems in the development need to be considered urgently and should be solved gradually in the future process of development of Shanghai precision agriculture.

4   The Relationship between "the Operating System of Shanghai Agricultural Economy" and "Precision Agriculture"

The Operating System of shanghai Agricultural Economy is a practical project of Shanghai digital agriculture. It provides the support of technology and information for

achieving precision agriculture. The system bases on GIS, database and the network technology, and has established the farming application systems facing the management, the decision-making, the relevant departments and specific users. Because the ordinary farmers are the major producers in suburban agriculture, the system takes advantage of progressive network GIS technology and regards the retail farmers as monitoring objects. In each plot the farmers should keep steady production. Among the plots there are objective differences. Thus, the system must follow these principles and realize digitalization in all aspects such as the agricultural production and circulation of the monitoring plots. This system constructed a dynamic professional database for the planting and economic operation. Through the comprehensive treatment and analysis, it can provide analysis and evaluation of the thematic information about the planting input-output and the dynamic monitoring data in Shanghai suburbs. It can also evaluate the operating performance of summer crops, autumn crops and crops for rotation under the different patterns, and express in space. The GIS expressive functions of spatial analysis, query and visualization not only show overall situation and spatial distribution pattern about the planting input, output and operating performance, but also show the production state of every monitoring point, various crops and every crop in different suburban, the structure of input and output, the operating performance, and the performance evaluation information in the conditions of different operating scale, different product certification and different market flow direction. All these provide scientific basis for the decision making of the agricultural department. The Operating System of Shanghai Agricultural Economy is a kind of management model of digital agriculture, which is line with the development foundation of crop production in Shanghai. After five years of operation, this system has already completed the construction and management of the dynamic monitoring network, remote data entry and the construction of professional database on the planting industry. It has exerted the certain function and guided the practical planting work. Thus, the Operating System of Shanghai Agricultural Economy is a meaningful practice in the huge digital agriculture projects of Shanghai, and provides the necessary conditions for realizing Shanghai precision agriculture.

5   Technical Framework and Workflow of "the Operating System of Shanghai Agricultural Economy"

5.1   Technical Framework of the System

The implementation of this system adopts the B/S and C/S architecture (as shown in Figure 1). It's comprised of the client (browsers), application service layer and the database server. All operation and data are deposited in server. Through the general Web browser the client can propose request of query and analysis. And the server returns the results. The client can realize the functions of data acquisition and editing, GIS inquire, statistical analysis and so on. The application service layer is the core of the system, is also a bridge between the client and the underlying database. The underlying database server includes the central database and model library.
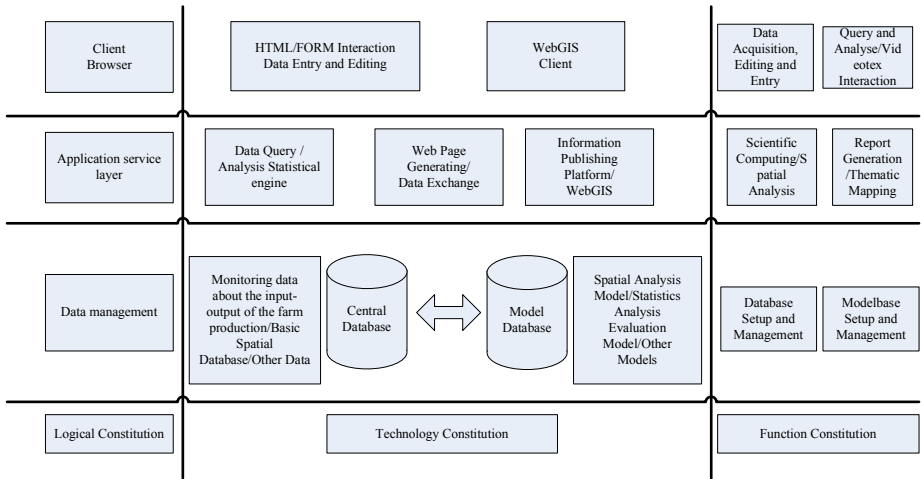
Client Browser	HTML/FORM Interaction Data Entry and Editing	WebGIS Client	Data Acquisition, Editing and Entry	Query and Analyse/Videotex Interaction
Application service layer	Data Query / Analysis Statistical engine	Web Page Generating/ Data Exchange	Information Publishing Platform/ WebGIS	Scientific Computing/Spatial Analysis	Report Generation /Thematic Mapping
Data management	Monitoring data about the input-output of the farm production/Basic Spatial Database/Other Data	Central Database	Model Database	Spatial Analysis Model/Statistics Analysis Evaluation Model/Other Models	Database Setup and Management	Modelbase Setup and Management
Logical Constitution		Technology Constitution		Function Constitution

Fig. 1. The technical framework about "the Operating System of Shanghai Agricultural Economy"
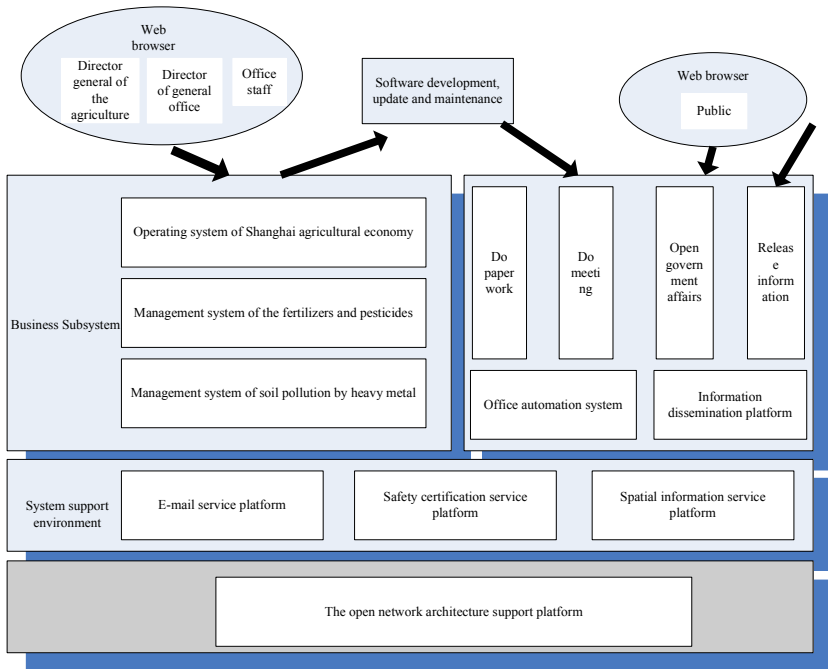
Fig. 2. The schematic diagram of "Shanghai Agricultural Comprehensive Management Information Platform"

On the business logic, this system is an important component of comprehensive management information platform for Shanghai agriculture. The design of whole system based on an open system, fully considering the data statistics within the agriculture department and the data exchange with other relevant departments, and strived to achieve standardization.

GIS platform software of this system uses the technical route combining the independent development of GIS components with commercial GIS platform. The whole framework system conforms to the standard of TC211/OpenGIS, and can carry out interoperability and data exchange with other GIS software (such as ArcGIS, MapInfo, GeoMedia) that fits this standard.

5.2   The Workflow of the System

The research of whole project contains several tasks, including the system design, the questionnaire design, the data acquisition and database design, the data input, the model design, programming and debugging, drawing up the achievements, and documentation, etc.(as shown in Figure 3)
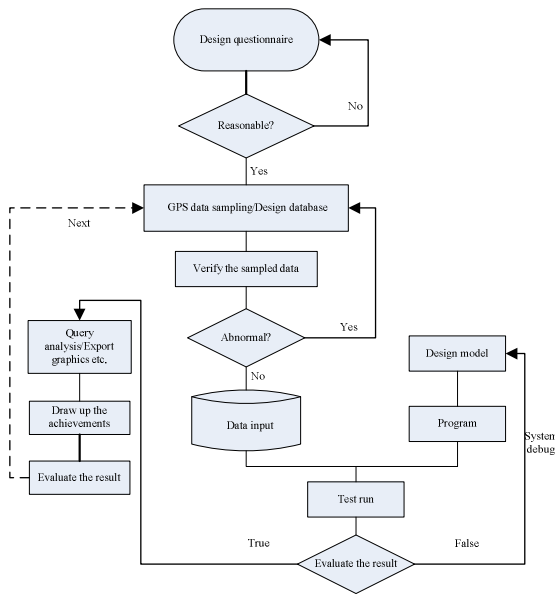

Fig. 3. The workflow of "the Operating System of Shanghai Agricultural Economy"

6   Database Design of "the Operating System of Shanghai Agricultural Economy"

The data of this system includes the attribute data, the spatial data and special graphics library data. The data storage and access use the technical framework of MSADO, and

database system software adopts ORACLE, in order to make full use of the existing database system's functions in multi-user management, query and analysis, and safety and affairs management.

6.1   Basic Space Database

The spatial data accuracy is based on "the Shanghai Digital Agriculture Platform for Basic Data". And the administrative division level reaches the township. The coordinate system combined the coordinates system of WGS84 (the coordinates system of GPS) with Shanghai local coordinate system. The system supports two coordinate systems transformation and the real-time updating of the dynamic GPS monitoring points.

6.2   Service Database

According to the crop type, the crops for rotation and the administrative division, the agricultural specialized data and the related attribute data are classified and graded. It has formed a database system facing the dynamic monitoring and management of agricultural economy. The table design of the database and the relationship among the tables are shown in figure 4. The sampled data and tabulate data, in different years and quarters, are in a list. So this design is beneficial to the year to year comparisons.
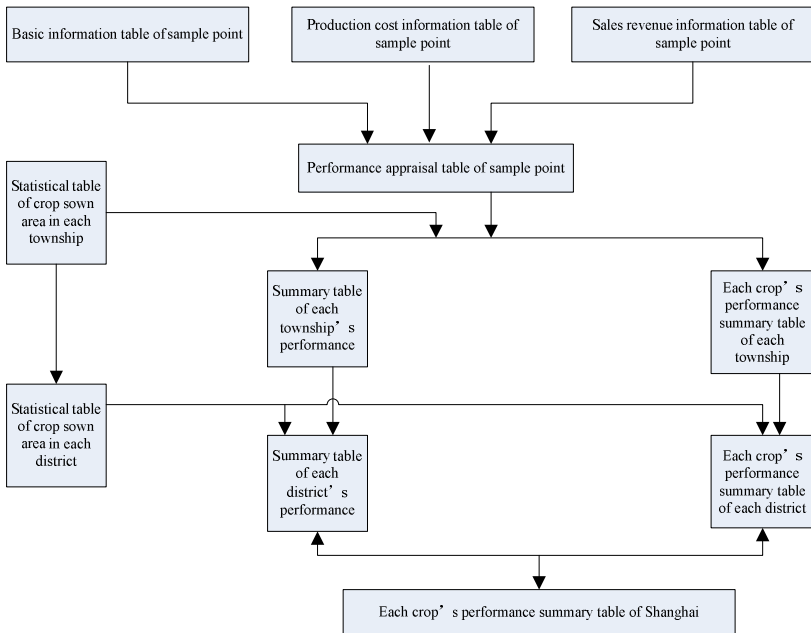

Fig. 4. Table design of the database and the relation among tables

6.3   Data Description

(1) Static data

The static data in this system mainly includes basic geographical spatial database provided by the SMAC. It contains the, the road network and the channel network.

The data of the administrative division includes some basic layers of each district and township. It's used as the base map for the query, analysis and thematic mapping.

The data of the road network includes the layers about the main roads and streets in Shanghai. It can show the road network distribution in Shanghai suburban.

The data of the channel network includes the layers of the main rivers which are under municipal management, or under district management, or township management, etc. It's used for display the water system distribution in the suburban of Shanghai.

(2) Dynamic data

According to the farming production structure and layout features in Shanghai suburban, the expected dynamic trend, and the development characteristics of the planting industry in each county, we reasonably assigned and established the crops type and the crops for rotation of the dynamic monitoring points in the initial period. The system has formed the dynamic monitoring network which covered 121 townships of Shanghai, the main crops types, and the main crops for rotation types. Each township established 10 monitoring points. There are a total of 1210 monitoring points in Shanghai suburb. In the future, if the administrative division of the townships is adjusted, the monitoring points won't be changed in principle. If a monitoring point is revoked as a result of land acquisition and other reasons, we will add a monitoring point in the same township at the beginning of next year to maintain 1210 monitoring points. We have set up the field archive for every monitoring farmer (enterprise), timely recording the farming activities. So through the agricultural service system and agricultural office system of Shanghai, the archive has formed the standardized data acquisition system.

According to the farmers` demand and the system design, the questionnaire data of "the Operating System of Shanghai Agricultural Economy" mainly includes the following respects: the total cost of production, the output, the data of fertilizer and pesticides utilization, and the other relevant data. The total cost of production includes the material costs, the service costs, the labor costs, the other costs. The output includes the total quantity of production, the total of sale, the total value of production, and the total sales.

This system has completed the design of four basic attribute databases. One is the basic survey information database of 1210 monitoring points. The second database contains two basic sub databases which are the operating system of the summer crops and the operating system of the autumn crops. The third is the annual database characterized by the crops for rotation. The last is the grading summary database of the total sown area in Shanghai suburb. The attribute database takes the monitoring point numbers for the keyword to number, and carries out a targeted data survey of the economy operation for the monitoring farmers in the whole year.

7   The Goal of "the Operating System of Shanghai Agricultural Economy"

(1) Taking use of GIS technology and based on "the Shanghai Digital Agriculture Platform for Basic Data", the system will complete establishing a dynamic professional database facing the planting production and economy operation.

(2) Through the system functions of the comprehensive processing and analysis, the system will provide the dynamic monitoring data of the planting input-output in Shanghai suburb, the process analysis and evaluation of the thematic information, the performance appraisal under the different conditions, and the spatial analysis and expression.

(3) Through the functions of GIS, such as the spatial analysis inquiry and visualization, the system could analyze and evaluate the situation, and show the spatial distribution characteristics about the input-output and performance of Shanghai planting production. Aiming at each monitoring point, each crop, or each district, it can also provide the performance evaluation information about the production situation, the constitution of the input-output, and the performance under the conditions of the different production scale, the product certification and the market flow direction. The system can provide the inquiry function of the related information.

(4) Through the GIS spatial analysis and inquiry of the system, we can know the application conditions of fertilizers and pesticides in every plot. So we can timely take measures to increase the utilizing efficiency of them in order to reduce the environmental pollution.

(5) The system has the ability to generate the relevant reports, the report forms, the statistical figure, and the thematic maps. And it supports the real-time inquires query and printout, the data exchange among the systems. These functions can provide the powerful decision support and information services for the macro-control and managing the planting production in the suburb of Shanghai.

In short, "the Operating System of Shanghai Agricultural Economy", by means of monitoring the huge data dynamically, gets the relevant information of the planting timely, and evaluates the real-time conditions of the planting. It is good for advancing the scientific and precision level of the crop production management, improving the economic benefit and social benefits and environmental benefits. And the system provides the technical support for achieving precision agriculture in Shanghai.

References

1. Wu, Y.-x., Li, W.-j., You, L.: On the construction of metropolitan agriculture in Shanghai. Geography And Territorial Research 16(3), 27–31 (2000) (in Chinese)
2. Han, J., Cai, Y., Cercone, N.: Data-driven discovery of quantitative rules in relational databases. IEEE Trans. Knowl. Data Eng. 5(1), 29–40 (1993)
3. Juan, W., Zha, L.-S.: Application of the Integration of GIS and SDM in Agriculture. Journal of Anhui Agricultural Sciences 36(29), 12974–12975, 12985 (2008) (in Chinese)

Research on the Data Conversion Specification for Chinese Agricultural Product Quantity Safety

Kaimeng Sun

Laboratory of Digital Agricultural Early-warning Technology of Ministry of Agriculture of
China, Institute of Agricultural Information, CAAS,
100081 Beijing, China
kmsun@mail.caas.net.cn

Abstract. This paper summarizes the current main methods of heterogeneous data conversion, focuses on the necessity of establishing the data conversion specification for China's agricultural product quantity safety, as well as specifies the main designs and solutions for data conversion structure, data conversion process and data conversion format in research. Establishing a data conversion specification is a pioneering work of applying modern information technology management methods in the research areas of agricultural macroeconomic policy-making system, which plays a positive role in promoting the study of modern agricultural information system. The establishment of data conversion specification can not only improve the level of agricultural information technology research, but can also create a better research and development environment for the study of agricultural information system.

Keywords: specification, agricultural product, data conversion.

1   Problem Statement

Chinese agricultural product quantity safety management system is an information management, analysis and decision-making system established for the study of agricultural product quantity safety. It involves large amounts of data related to agricultural product quantity safety, including the agricultural economic base data, field collection dynamic weather data, soil and crop growth data, agricultural product price data from internet, agricultural product quantity safety historical data for statistical analysis, etc., with different data sources and types. For overall system needs, to establish a unified and standardized database that can serve the agricultural system is the base for agricultural product quantity safety research.

The key indicators involved in agricultural product quantity safety management system are divided into two categories. The first category is production indicators: including the arable land area, acreage, production, chemical fertilizer and pesticide use, etc. of crop production. As the agricultural product quantity safety in our study includes various consumer goods needed by human life, here the agricultural product is a broad concept, and therefore the indicators also include the breeding area,

quantity, meat, milk, etc. of livestock production; the forest area of forestry production, all kinds of fruit production, etc.; production environment such as soil conditions, texture, moisture, fertility, weather factors, natural disaster indicators and others; the second category is consumer indicators: including demographic, agricultural economy, all kinds of income, number of expenditures, various related indices and so on. Overall, the indicators of system show a large quantity, a wide variety of sources, complex formats and other features.

Taking agricultural economic base data as an example, data are from the Ministry of Agriculture, national statistical offices and various types of information and internet dynamic snapshots, due to the differences in database platform and type of various departments, there will be heterogeneous database problems. These databases include SQL Server, Oracle, Access, etc. As different databases have different data types, storage and data retrieval methods, how to convert and integrate data so as to access to information from the huge data resources, has become a hot topic in the current database research field, especially the data conversion of heterogeneous database has become a pressing issue now.

The study of agricultural product quantity safety management system is based on the establishment of agricultural product quantity safety-related databases, while the most important issue for establishing agricultural product quantity safety databases is the conversion of various types of data, which means in accordance with the unified format requirements of agricultural product quantity safety database, to convert different types of data into agricultural product quantity safety databases in the format required. Therefore, how to convert various sources of data into a unified data structure to enter the agricultural product quantity safety database, requires creating a data conversion specification to uniform data formats, types, and conversion methods.

Specification is defined as standards of conduct established by groups, the content of specification document is binding and regulating people's behavior in nature. To establish the data conversion specification in agricultural product quantity safety-related research, is to define data formats and standards of conduct for the relevant data conversion. The meaning of establishing specification is to unify the conversion method, format and process of agricultural product quantity safety data, that is, to keep access to the data of agricultural product quantity safety management system under a unified agreement, so as to ensure various types of data convert into system use in accordance with system requirements and unified steps , in order to protect the quality of the data.

2   The Current Main Methods of Data Conversion

Data conversion means the process of converting data structure, form of expression and file format from one into another. Taking agricultural product quantity safety database as an example, data are accessed through a variety of sources, the database structure and storage form of data are not the same, as the amount of data increaseing, the original unreasonable data structure cannot meet various requirements. The replacement of database and data structure requires the conversion of data themselves.

Only by data conversion to generate data and document exchange of heterogeneous system, can data structure of database be unified in accordance with the requirements of agricultural product quantity safety research, and data be used for research to play their real role. Currently there are a variety of ways of data conversion for different information systems.

2.1   Conversion Based on the Timeliness of Data

Based on timeliness, data conversion consists of real-time and non-real-time (delayed treatment). Real-time data conversion is a system as a data source providing data call service for another system. Receiving service conducts real-time conversion and processing of the data accessed to provide data services for the system. This real-time data conversion is usually found in the network. For example, the agricultural product quantity safety management system calling the content of a particular website that provides data services, or the system calling some of the external programs, are all real-time conversion, which requires filtering, extracting, computing, converting and sorting the data or program called from the website by definition, then adding the converted data in accordance with certain requirements of form to the database of agricultural product quantity safety management system. This processing is required to be completed within a very short time, that is, the timeliness of data access.

Non-real-time data conversion refers to data that are difficult to convert and do not need to immediately convert after data extraction between two information systems, such as databases, office software, etc. Through some technical means, such as the application of some intermediary format, the data has been converted into a common data format, and after data conversion, it can be directly applied to different information systems. Or by programming, specially write a program for the two data systems, to convert data extracted from the two systems so that it can be applied to another system. According to the conversion process, it can be divided into intermediate format conversion and direct conversion between source data and target data.

Comparing the two methods, we can find direct conversion between two different data formats is more efficient and of less negative effect (noise brought about, a small amount of data forced to discard, the degree of structure forced to reduce), which means the mapping from original format of data to target format of data will be maintained better. However, the possibility of direct conversion with ready-made means is not big, very often requiring special programming for converting from the source format to the target format.

2.2   Conversion Based on Different Data Structure

Through studying heterogeneous databases and making efforts to develop solutions to convert heterogeneous database data, we can achieve centralized management and uniform use of data in a unified environment. Heterogeneous database conversion is relatively complex in techniques, and currently it is still difficult to use a generic DBMS (Database Management System) to solve this problem. The common language

specification XML (Extensible Markup Language) defines data structures in an open self-described manner, which can highlight the description of the structure while describing the data content, so as to reflect the relationship between the data, therefore is the most widely used normative language for data conversion. XML can achieve the separation of data representation with data content, allowing seamless integration of data from different sources and different processing of the same data. XML language uses XML Schema to describe the structure of the document, meanwhile XML Schema has rich data types and powerful ability to express constraints, making it feasible and convenient to take XML schema as the common data model to establish a data conversion tool between unified pattern of relationships. With the development of XML technology and XML'sadvantages in data applications such as easy to express, cross-platform, semi-structured data description, good scalability, separation between content and form, etc., favorable conditions have been provided for heterogeneous database data conversion. The heterogeneous database data conversion based on XML is not only beneficial for data sharing, making savings in software development investment, but also conducive to better management and use of data, laying the foundation for government and enterprise applications of information.

The importance of information system lies in the information exchange between different information systems. The information stored in separate information systems is not complete and single, only by exchange of information can information be enriched and completed. However, with the ubiquitous presence of information systems, their development is towards diversification and specialization. Therefore, information sharing and data and document conversion between information systems has become a subject to be solved now. Only with exchange of information, as well as timely and accurate information conversion, information is useful. A data conversion specification for agricultural product quantity safety should be established, so that various types of data conversion is in accordance with the requirements of system design, exchange of different types of information is smooth, information resources are fully utilized, in order to ensure the research work of agricultural product quantity safety be carried out smoothly.

3   The Functional Structure of Agricultural Product Quantity Safety Data Conversion

3.1   The Functional Structure of Data Conversion of Agricultural Product Quantity Safety Management System

Based on the data sources and structure of agricultural product safety management system, through analyzing and researching, the functional structure of data conversion is established, consisting of database operation module, database source extraction module, data structure conversion module and data key-in module according to functions. The functional structure chart of data conversion is as shown in Figure 1:
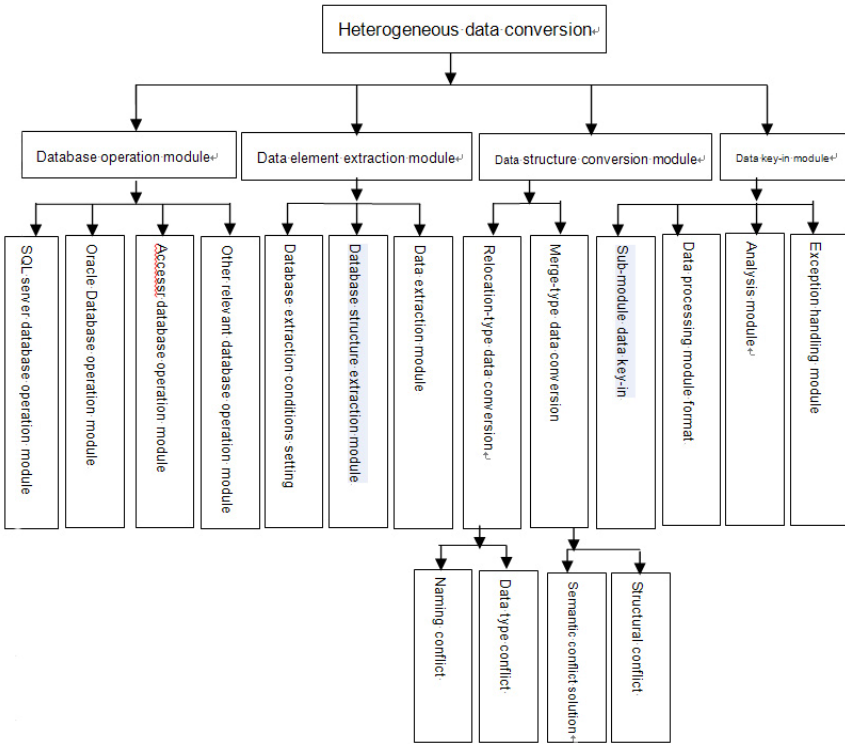
Fig. 1. The functional structure chart of data conversion of agricultural product safety management system

3.2   The Main Functional Module Features of Data Conversion

3.2.1   Database Operation Module

This module creates different classes respectively for different databases to achieve connection between a variety of databases, data query, insert, delete, modify and other operations. When you need to add a database type, you can just add the corresponding class to improve the scalability of the tool.

3.2.2   Data Extraction Module

This module sets data extraction conditions according to the needs of data conversion to extractthe source database data as XML data files and XML structure files.

3.2.3   Data Structure Conversion Module

This module implements the data structure conversion between source database and target database, including two data conversion solutions, the relocation type and the merge type, to solve the existing naming conflict, data type conflict, structural conflict and semantic conflict in two data conversion solutions, achieve multi-table conversion under complex conditions, and finally generate data structure conversion files for controlling the key-in of XML data from the source database to the target database.

3.2.4   Data Key-in Module

According to the corresponding data insert statement generated by the data structure conversion files generated by data structure conversion module, this module inserts and updates the data of XML data files generated by data extraction module from the source database to the target database. In the data key-in process, sub-data key-in can be selected according to the data file size; when sub-data key-in is selected, XML data file will first be cut into several smaller XML data files, and then each XML data file will be circularly processed to complete the key-in of XML data to the target database. Meanwhile, the data key-in module can also complete data format adjustment based on the actual situation.

4   The Basic Data Conversion Process of Agricultural Product Quantity Safety

The data conversion process divides data conversion into two cases: relocation-type data conversion and merge-type data conversion. The relocation-type data conversion refers that the target database is empty, just needing to completely relocate the data to be converted from the source database to the target database, which approach requires to establish the target database in full accordance with the source database structure; the merge-type data conversion means the target database already exists and there is data, needing to establish the corresponding relationship between source database structure and target database structure, inserting and updating the source database data into the target database according to the structural correspondence setting.

The conversion process chart from the source database to the target database is as shown in Figure 2.

First, data extraction, which is the starting point of data conversion, is responsible for accessing information from the source database to match the target database, specifically including:

- calling the database operation module
- setting the conversion data according to needs
- calling the data extraction module, to extract the setting conversion data as XML data files, and extract the structural information of conversion data as XML structure files.
- extracting the target database structure as XML structure files to prepare for the conversion between the source database and target database data structure.

Then comes data structure conversion, the data structure conversion between source database and target database is the core of data conversion process, which sets the corresponding relationship between source database and target database structure based on the XML structure files of source database and target database, to resolve the semantic conflict, naming conflict and type conflict in heterogeneous database data conversion, and generate data structure files to control the updating and inserting of source database data into the target database.

Finally, data key-in is achieved, according to the corresponding data insert statement generated by data structure conversion files, updating and inserting the source database into the target database to complete the data conversion.
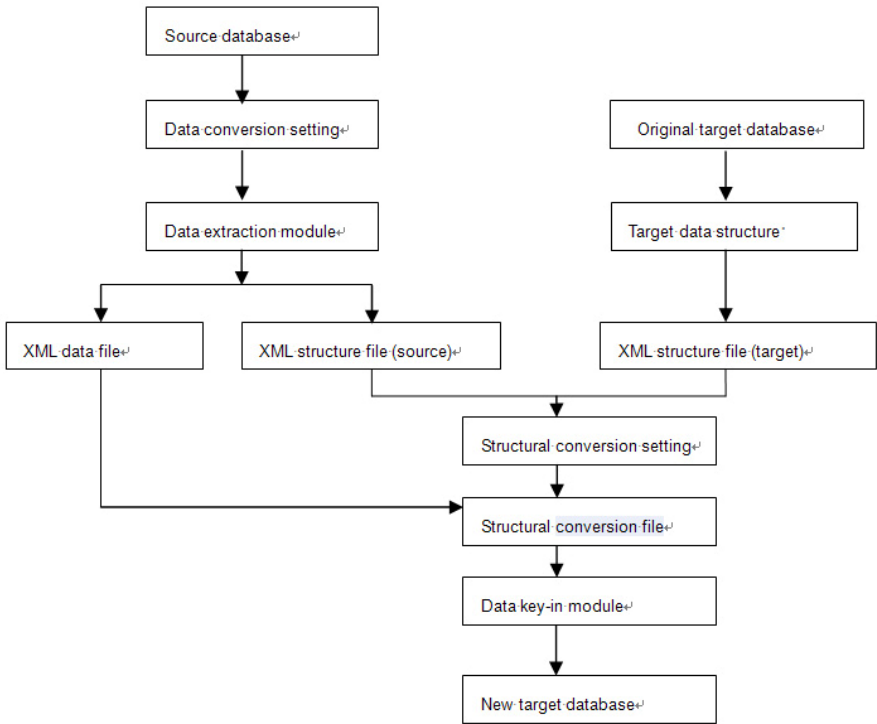

Fig. 2. Data conversion process chart

5   The Exchange Format Design for Basic Data and Shared Data of Agricultural Product Quantity Safety

5.1   Extraction and Standardization of Data Elements

5.1.1   Data Source Extraction

Using the logical analysis of natural language, the basic data objects of agricultural product quantity safety that need exchange format design are analyzed in terms of role, location, thing, event and time, of which the thing and event, each also contains two aspects: identification and characterization. Detailed steps are as follows:

- Make sure all relevant details of participants;
- Determine the identity elements, such as: ID numbers and names fall into this category;
- Make sure all the details related to the event. Event must have a date / time and a type. In general, there should be a location / position where event takes place;

- Define the relevant participants. Party A and Party B are participants related to the registered event;
- Define the location / position. There are three basic types of location: geographical address, physical address (latitude and longitude) and mailing address.

5.1.2   Standardization of Data Elements

As for the data elements extracted, experts in the related field of agricultural product quantity safety basic data sharing will standardize the data elements in accordance with the basic principles and methods and other relevant standards of data element standardization, so as to form reusable data elements that can directly compose of syntax-independent data exchange format.

5.2   Data Classification and Coding

Based on the data elements formed, code processing is needed for the data elements which require coding in accordance with the methods described in the basic principles and methods of data classification and coding and other relevant standards, so as to form the basic data classification and coding for agricultural product quantity safety, the code mapping of code-type data elements is the attribute in XML Schema.

After combing the business process, relevant information model is formed, how to convert these information models into syntax-independent, data element-composed data exchange format, can be illustrated in the following two cases:

5.2.1   Relevant Data Elements Are Readily Available

If a data element directory has been established in specific areas of agricultural product quantity safety basic data sharing, related data elements in the directory should be used with a priority, for the standardized processing of data in the information model, so as to form data element-based, syntax-independent data exchange format with mapping to any syntax corresponding to the information model.

5.2.2   Relevant Data Elements Are Not Readily Available

If there is no established data element directory, or there are no reusable data elements in the established data element directory to meet the needs of data exchange, you need to extract data elements based on the information model, in order to develop a data element directory in the field. On the basis of the directory, as described in 5.2.1, further form data element-based, syntax-independent data exchange format with mapping to any syntax corresponding to the information model.

In accordance with the mapping rules provided by this specification, the data element-based, syntax-independent data exchange format is mapped to XML Schema, and finally will form XML Schema that can achieve unambiguous exchange between homogeneous or heterogeneous systems.

5.3   The Composition of Data Exchange Format

Data exchange format consists of three parts, including head of document, body of document and end of document.

- Head of document includes document identification information, such as: document number, document date, role information and so on.
- Body of document includes the specific business information of data exchange format, such as: the registered data content.
- End of document is some descriptive information, which can be omitted sometimes. Due to space limitations, the XML mapping rules of data exchange format are described in detail in the article.

6   Summary

To ensure the research of agricultural product quantity safety management system, the data accuracy, timeliness and effectiveness of database system must be ensured, for data quality determines the success or system. As already mentioned, due to large volumes of data and a wide range of data sources, data conversion is the basic work. In order to ensure the quality of the data, it is necessary to establish a data conversion specification in line with the requirements of agricultural product quantity safety management system. This specification includes the definition of data formats, data conversion functions and data conversion processes, is one of the most important work of the system database. The same can be said that the data specification will directly affect the statistical analysis, model operations and decision-making results of the system. The data conversion specification research is a useful attempt to use modern information technology management methods in the research field of agricultural macroeconomic policy-making system, and is also an inevitable trend of the continuous development of modern agricultural information system R&D work.

The development of data conversion specification itself is an important part of the research of agricultural product quantity safety management system, which is basic research. In the previous study of agricultural information technology, very few norms and standards were involved; therefore, the study of norms and standards is a new attempt. In addition to data conversion specifications, we have also studied and established data access and data processing specifications, the establishment of which not only improved the level of agricultural information technology research, but also created a better development environment for the study of agricultural information system, so that agriculture information technology research was carried out in strict norms, promoting the standardization, industrialization and modernization of agricultural information system research. It can be said that the establishment of various norms have injected new content into the development of agricultural information technology research, with such norms and standards, agricultural information technology will be able to better serve macroeconomic policy analysis, serve agricultural production and life of the peasants, as well as serve agricultural modernization in China.

References

1. Zhang, C.: Study of the XML-based Data Exchange Technology of Heterogeneous Databases. Computer Knowledge and Technology 3(8) (2008)
2. Geng, X.: XML-based Tutorial. Tsinghua University Press, Beijing (2006)
3. Du, S.: The Design and Implementation of XML-based Data Conversion Tool for Heterogeneous Databases. Master's degree thesis of Shenyang Polytechnic University (2009)

Small-Scale Evaluation of Tobacco Planting Suitability Based on Spatial Information Technology

Fengrui Chen[1], Guangxiong Peng[2,*], Wei Su[3], Yaochen Qin[1], and Xi Li[4]

[1] The College of Environment and Planning, Henan University, Kaifeng, China
[2] School of Geosciences and Info-Physics, Key Laboratory of Metallogenic Prediction of Nonferrous Metals, Ministry of Education, Central South University, Chansha, China
[3] College of Information and Electric Engineering, China Agriculture University, Beijing, China
[4] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, WuhanUniversity, Wuhan, China
fruich@gmail.com, pgxcsu@csu.edu.cn

Abstract. Crop planting suitability evaluation is important for agricultural production. In this study, a comprehensive evaluation framework which combined geographical information system, remote sensing, global position system and geostatistics with other models was presented to evaluate the tobacco planting suitability on a small scale. Ten indicators from three respects of climate, terrain and soil nutrients were selected to build evaluation indicator system. The results show that 70% of study area is suitable for tobacco planting. However, the spatial distribution of current tobacco fields is some degree of irrationality, and 20% of fields are located in moderate or not suitable area. Through scale analysis of the arable land, the suitable fields are mainly distributed in the middle and southern parts of the study area, which is 6324.5 ha.

Keywords: Spatial Information Technology, Planting Suitability Evaluation, Tobacco, Scale Analysis.

1   Introduction

Developing precision agriculture is an inevitable direction in the world [1-4]. The key of actualizing precision agriculture is spatial information technology, which involves spatial data acquisition, processing and decision–making [5-7]. Tobacco is a special agriculture crop which products excellent tobacco leaves only in regional cultivated land. In order to improve the tobacco quality and have a high market competitiveness, it is meaningful to evaluate tobacco planting suitability based on spatial information technology, and on this basis, tobacco production is redistributed rationally.

Geographic information system (GIS) which provides powerful spatial data management and analysis functions has been widely used for manage spatial resource

---

* Corresponding author.

[8-11]. In recent years, tobacco planting suitability evaluation based on GIS has been presented [12-14]. Although they can guide tobacco production overall, the evaluation results can not meet the requirements of precision agriculture on small scale. For example, there is always a meteorological station per country[15] and several soil sampling points standing the whole area, which will be very rough for assessment. However, few studies have focused on small-scale tobacco planting suitability evaluation such as on the county and its below scale. Furthermore, modern tobacco agriculture demands that tobacco fields not only are distributed in suitable area, but also have a certain scale. All these pose challenges to the traditional evaluation method.

The increasing needs of modern tobacco agriculture and the shortage of current evaluation method stimulate a need for sophisticated methods of suitability evaluation. In this study, according to the physiological characteristic of tobacco [16] and the study area condition, ten evaluation indicators from three factors of climatic, terrain and soil have been chosen to build the tobacco planting suitability evaluation indicator system on small scale. Then, in order to attain accuracy evaluation indicators, remote sensing data was used to retrieve surface parameters, and soil sampling data based on global position system (GPS) was used to attain soil nutrient information by geostatistical analysis. On the basis of analysis above, GIS combined with analytic hierarchy process (AHP) and membership function model were applied to comprehensive evaluation of tobacco planting suitability, which could provide scientific bases for sustainable development of tobacco production.

2   Materials and Method

2.1   Study Area

The research was conducted in one part of Huili County (E 102°13′-102°24′ and N 26°33′-26°48′), located in the southwest of Sichuan province, China (Fig.1). It is one of the main tobacco production regions of Sichuan province. The highest elevation of the study area is 2390 m and the lowest is 1680 m. There are many kinds of landform types such as mountain, hill, plain and basin, which results into redistribution of light, heat and water resources, arable land mainly distribute in plain and basin area. The local climate is classified as temperate continental with annual average rainfall of 1150 mm, annual mean temperature of 15.1℃, annual sunshine of 2398 h and annual frost-free days of 240. There is big different in temperature between day and night here, and is rich in water, heat and good natural conditions, therefore, it is very suitable for high-quality tobacco production.
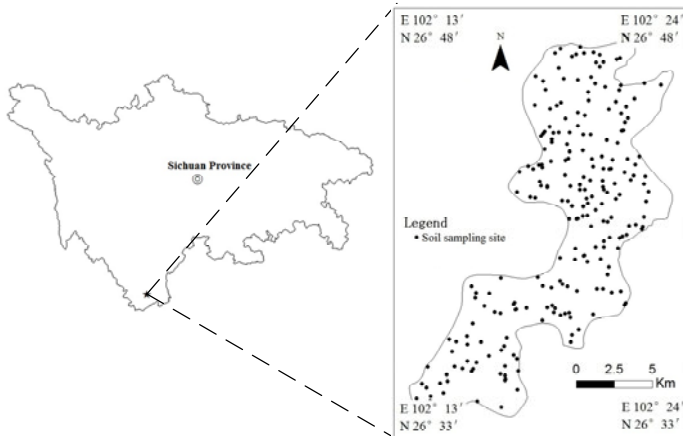
Fig. 1. The location of study area and the spatial distribution of soil sampling

2.2   Data Source

Solar radiation is an important impact factor for crop growth. In order to attain high accuracy spatial distribution of solar radiation, a series of models and methods have been established and developed[17, 18]. In this paper, the model proposed by Fu and Rich[19] was used to calculate the solar radiation of the study area, which has been applied widely. The tobacco growing season is from May to July, so the mean solar radiation of this period was calculated (Fig.2).

Owing to the sparse distribution of meteorological station, about a station per country, so it is difficult to obtain high accuracy climate data from observation site at country or its below scale. Remote sensing is an effective method, which can acquire surface parameters with high spatial resolution in any region, and compensate for the loss caused by the absence of observation stations[20]. Moderate resolution imaging spectroradiometer (MODIS) which has high temporal and spectral resolution, moderate spatial resolution and abundant products was used to retrieve land surface parameters. MODIS products in May, June and July from the year of 2005 to 2010 were adopted, and these data were downloaded from the website of national aeronautics and space administration (NASA) of the United States.

Land surface temperature (LST) is obtained from MOD11 products of MODIS with the spatial resolution of 1 km. Its average error is smaller than 0.5 K [21, 22]. Fig.2 shows the mean LST from May to July in the study area.

Soil moisture is derived from temperature vegetation dryness index (TVDI) based on normalized difference vegetation index (NDVI) products and land surface temperature (LST, MOD11) products of MODIS [23]. It can effectively reflect the spatial distribution of relative value of soil moisture, and has high correlation with the

soil moisture. The value of TVDI is from 0 to 1. The soil moisture increases with the decrease of TVDI. The spatial distribution of TVDI is shown in Fig.2.
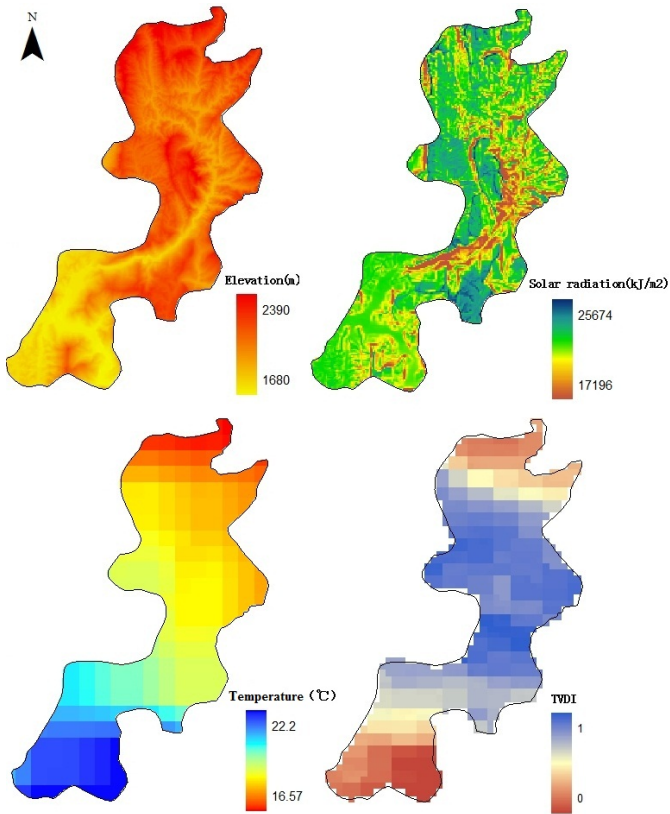
Fig. 2. The evaluation indicators of elevation, solar radiation, temperature and TVDI

Soil sampling was carried out in April 2009, immediately after harvest of wheat and cole. In total 268 soil samples at the depth of 20 cm were taken and their precise locations were recorded by a GPS receiver (Fig.1). Before analysis, all soil samples were air-dried and passed through a 2-mm sieve prior to analysis. Soil pH value was determined by the method of electrical potential; organic matter (OM) content was measured by dilution heat method with potassium bichromate; and available nitrogen (AN), available phosphorus (AP) and available potassium (AK) were measured by alkaline hydrolysis diffusion, molybdenum stibium anti-color method and flare photometer, respectively.

In addition, elevation and slope indicators are derived from DEM data with the spatial resolution of 20 m. Arable land is derived from the digital of thematic map with the scale of 1:50000.

2.3   Geostatistics Method

Geostatistics was used to estimate and map soil nutrients in unsampled areas [24, 25]. Among the geostatistical techniques, kriging is a linear interpolation procedure that provides a best linear unbiased estimation for quantities, which varies in space. Kriging estimates are calculated as weighted sums of the adjacent sampled concentrations. That is, if data appear to be highly continuous in space, the points closer to those estimated receive higher weights than those farther away[26].

In geostatistical analysis, the semivariogram was calculated for each soil variable as follows:

$$r(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left[z(x_i) - z(x_i + h) \right]^2 \tag{1}$$

where $z(x_i)$ is the value of the variable z at the sampled location x_i, h is the distance lag in meters and $N(h)$ is the number of pairs of sample points separated by h. For irregular sampling, it is rare for the distance between the sample pairs to be exactly equal to h. Therefore, h is often represented by a distance interval. For the distance lag h, the semivariance is $r(h)$.

The experimental semivariogram was then fitted with a suitable theoretical model: spherical, exponential or gaussian. The models provide information about the spatial structure as well as the input parameters for kriging interpolation. Kriging is considered as the optimal spatial interpolation for making best linear unbiased estimates of regionalized variables at unknown locations. The spatial prediction of the value of a soil variable z at an unknown point x_0 is calculated as a weighted average:

$$z^*(x_0) = \sum_{i=1}^{n} \lambda_i z(x_i) \tag{2}$$

where $z^*(x_0)$ is the value to be estimated at the location x_0; $z(x_i)$ is the known value at the sampling site x_i and λ_i is the weight. There are n sites within the search neighbourhood around x_0 used for estimation, and the magnitude of n will depend on the size of the moving search window and on user's definition. Kriging differs from other methods, such as inverse distance weighted, in that the weight function λ_i is no longer arbitrary, and is calculated from the parameters of the fitted variogram model with unbiasedness and minimized estimation variance for interpolation.

2.4   Comprehensive Suitability Evaluation

According to the principles of locality, similarity, diversity, integration and practicality, an evaluation indicator system based on three factors including ten indicators was built. The flow chart and processes method of this study is as Fig.3 .
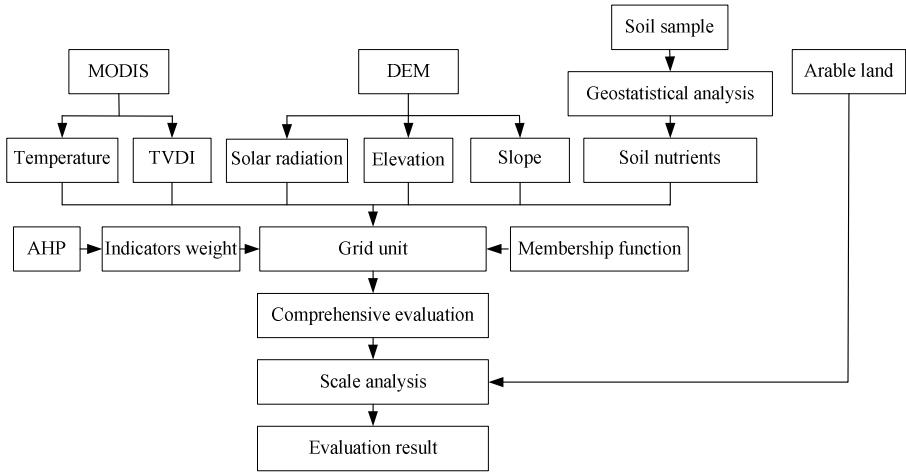
Fig. 3. The flow chart and processes method of this study

2.4.1   Indicator Scoring

Table 1. Standard scoring functions and parameters for quantitative tobacco planting suitability evaluation indicators

	FT	L	L1	U1	U	MF
Temperature(℃)	S-type	17	--	--	19	
SR(kJ/m²)	S-type	15000	--	--	23000	$f(x)=\begin{cases}1 & x\geq U\\ 0.9\times\frac{x-L}{U-L}+0.1 & U>x\geq L\\ 0.1 & x<L\end{cases}$
AK(mg/kg)	S-type	80	--	--	180	
TVDI	Z-type	0	--	--	1	$f(x)=\begin{cases}1 & x\geq U\\ 0.9\times\frac{x-L}{U-L}+0.1 & U>x\geq L\\ 0.1 & x<L\end{cases}$
Elevation(m)	parabola-type	800	1400	1800	2400	
Slope(°)	parabola-type	0	6	15	25	$f(x)=\begin{cases}1 & L1\leq x<U1\\ 0.9\times\frac{x-L}{L1-L}+0.1 & L<x<L1\\ 0.9\times\frac{x-U1}{U-U1}+0.1 & U1<x<U\\ 0.1 & x\leq L, x\geq U\end{cases}$
pH	parabola-type	4.5	5.5	6.5	8.0	
OM(g/kg)	parabola-type	10	22	28	45	
AN(mg/kg)	parabola-type	40	80	100	150	
AP(mg/kg)	parabola-type	5	20	30	45	

FT means membership function type; L means lower limit; U means upper limit; L1 means lower optimal value; U1 means upper optimal value; MF means membership function.

Because of different indicator units, a standard scoring function was used to score suitability indicators. The scoring of evaluation indicators on growth suitability of tobacco can be described with the fuzzy methods. The interaction between growth and quality of tobacco was different. Referring to the existed research results [16, 27, 28] and the actual situation of the study area, different types of membership functions, including parabola-type, Z-type and S-type, were chosen to express their relations

(Table 1). The membership degree which is between 0.1 and 1.0 can be calculated according to these membership functions, which reflects the degrees of membership. The upper limit 1.0 is referred to as the best situation of indicators which is suitable for the growth of tobacco crops, while the lower limit 0.1 is referred to as such that there is a serious deficiency in that indicator.

2.4.2   Weight Assignment

The indicator weight is the contribution degree of the effects of evaluation indicators to tobacco growing. There are many methods to assign indicator weight for tobacco planting suitability, such as the Delphi, AHP, regression analytical method and principal components analysis. In order to weaken subjective influence, the application of the AHP method, developed by Saaty [29] for environmental assessment, has been used extensively[30, 31] and was used to derive the weights of indicators, which are shown in Fig.3.
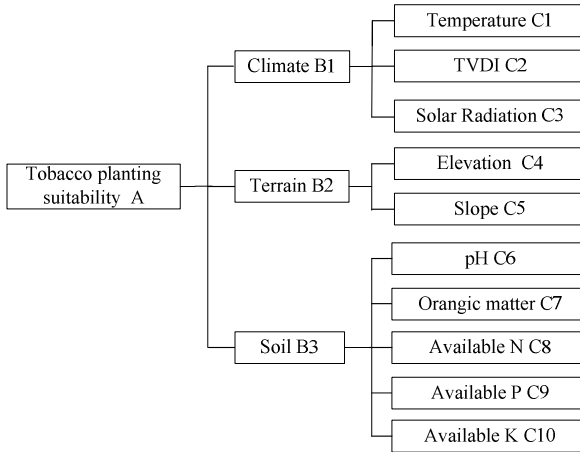
Fig. 4. Hierarchical structure for the indicator weight assignments

Table 2. The weight of evaluation indicator for tobacco planting suitability

C	B1 0.49	B2 0.20	B3 0.31	Weights
Temperature C1	0.54	0	0	0.26
TVDI C2	0.16	0	0	0.08
Solar radiation C3	0.30	0	0	0.14
Elevation C4	0	0.75	0	0.15
Slope C5	0	0.25	0	0.05
pH C6	0	0	0.44	0.14
Organic matter C7	0	0	0.29	0.09
Available N C8	0	0	0.13	0.04
Available P C9	0	0	0.05	0.02
Available K C10	0	0	0.09	0.03

The results of AHP analysis are presented in Table 2. The consistency test for single and general hierarchy storing was conducted by calculating the average random consistency index (RI). The results indicated that all RIs for single and general hierarchy storing were lower than 0.1, which means all the matrixes in the hierarchy A, B and C were logically constructed. In addition, the maximum of the eigen value and its eigen vector should be calculated and verified by the coherence index-coherence ratio (CR). If the CR value is less than 0.1, it deems that the variance should be allowed. Or else, the matrix must be adjusted once again. In this study, CR value was 0.04, which was less than 0.1. Therefore, the analysis results were acceptable.

2.4.3   Calculation of Comprehensive Evaluation Values

According to the character of study area, the size of evaluation unit for tobacco planting suitability is 30 * 30 m. All the indicators should be resampled into grid unit with the spatial resolution of 30 m before comprehensive evaluation. After all indicators were scored and weighted, the integrated quality index (IQI)[32] model was used to calculate the tobacco planting suitability level in study area:

$$IQI = \sum_{i=1}^{n} W_i N_i \tag{3}$$

Where W_i is the assigned weight of evaluation indicator, N_i is its score, and n is the number of evaluation indicators. The higher the IQI value is, the more suitable the unit is for tobacco planting.

3   Results

3.1   Descriptive Analysis of Soil Nutrients

Table 3. Descriptive statistics of soil nutrients in the study area

	Min	Max	Mean	Std	CV	Skewness
pH	4.9	8.1	6.3	0.72	0.11	0.19
OM(g/kg)	6.0	62.2	24.7	10.5	0.43	0.84
AN(mg/kg)	11.4	212.2	81.4	39.1	0.48	0.64
AP(mg/kg)	2.3	51.8	23.3	11.7	0.5	0.45
AK(mg/kg)	58.4	490.9	198.6	102.9	0.52	1.05

The descriptive statistics of soil pH, organic matter, AN, AP and AK for study area are shown in Table 3. In general, the soil was suitable for tobacco planting, and the means of pH, OM, AN and AP were in their optimal value ranges (Table 1). The CVs of OM, AN, AP and AK were similar, all exceeding 0.4, which indicated considerable spatial variability. However, the CV of pH was much smaller, only 0.11, which indicated significant spatial constant. The skewness of AK was more than 1, so it was necessary to transform its non-normally distributed raw data into data with an

approximately normal distribution. We found that the log transformation of AK significantly improved their normality, therefore, log(AK) were used for the geostatistical analysis.

3.2   The Spatial Analysis of Soil Nutrients

The best-fit semivariogram models and parameters for pH, OM, AN, AP and log(AK) are listed in Table 4. We used variable distance lags to compute the experimental semivariogram to see how the semivariograms behave near the origin and when the distance lag increases. As listed in Table 4, the exponential model gave the best fit to the pH, AN, AP and log(AK) experimental semivariograms, except OM. For OM semivariogram, a gaussian model (C_0=29.1, C=29.57 and range = 1720 m) produced the best fit (Table 4).

Table 4 also indicates a distinctly different degree of spatially structured variations of the five nutrients. Spatial continuity for a regionalized variable can be characterized by its range. The longer the range is, the better spatial continuity becomes. Among the five nutrients, the range of log (AK) (2970 m) was the longest, so it had the best spatial continuity. In addition, spatial structured variation for a regionalized variable can be divided into three categories: strong, moderate and weak spatial dependence, corresponding to a nugget-to-sill ratio $C_0/(C+C_0)$ of < 0.25, 0.25~0.75 and > 0.75, respectively. In this study, all five variables demonstrated moderate spatial dependence, and AP had the strongest spatial dependence due to its lowest nugget-to-sill ratio (0.39). Although pH and log (AK) had similar nugget to sill ratios (0.5 and 0.52), their spatially structured variations appeared at different spatial scales as their ranges were significant different (1820 m for pH and 2970 m for log(AK)).

Table 4. Semivariogram models and their parameters for soil nutrients

	Model	C_0	C	$C_0/(C+C_0)$	Range (m)
pH	Exponential	0.18	0.23	0.44	1820
OM	Gaussian	29.1	29.57	0.5	1720
AN	Exponential	721	624	0.54	2850
AP	Exponential	55	86	0.39	2474
log(AK)	Exponential	0.12	0.11	0.52	2970

Fig.5 shows the spatial patterns of the five soil nutrients generated from ordinary kriging analysis based on their semivariogram parameters (Table 4). Owing to the smooth effect of kriging, all predicted value ranges of five nutrients were smaller than the corresponding sampling data. Cross validation was carried out to evaluate the accuracy of the ordinary kriging interpolation. As a result several indices, mean prediction error (ME), mean absolute error (MAE), root mean squared error (RMSE) and correlation (*r*) between observed and predict value were adopted. Table 5 summarizes the cross validation of ordinary kriging interpolations for pH, OM, AN, AP and AK. The results indicted that AN and AK were slightly overestimated and OM and AP slightly underestimated. According to the correlation coefficients *r*, pH

was predicted best ($r = 0.77$), then AP($r = 0.72$), and the worst AK ($r = 0.55$). The relatively low r for AK may be due to its predicted value which was transformed from the predicted value of log (AK).
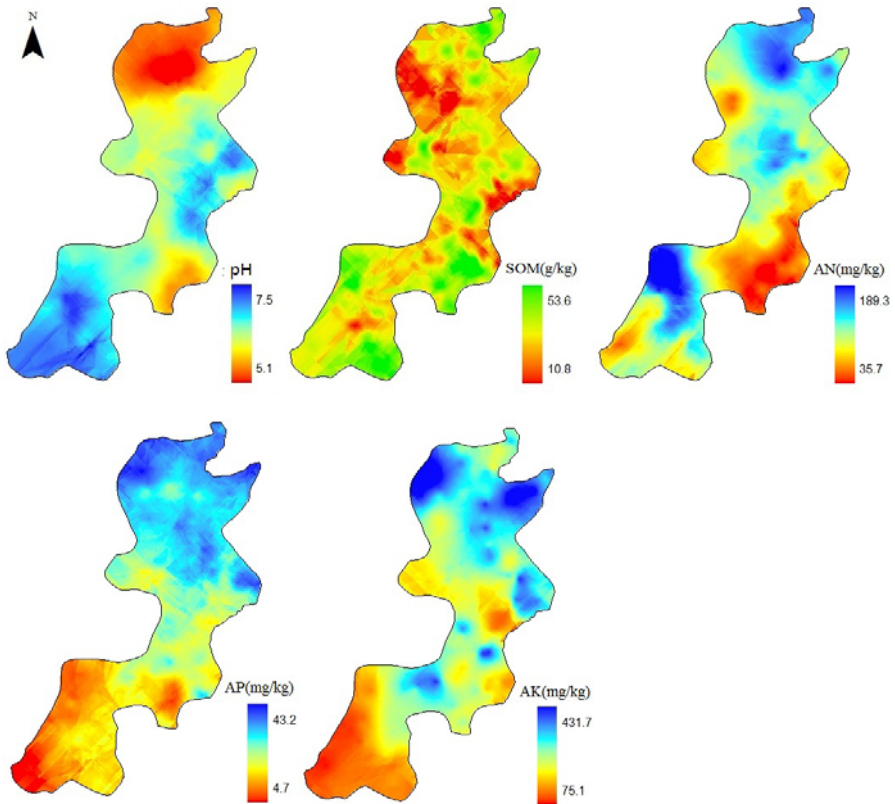

**Fig. 5.** The spatial distribution of soil nutrients

Table 5. Summary of cross validation of ordinary kriging interpolation for soil nutrients

	pH	OM	AN	AP	AK
ME	0.0	-0.03	0.01	-0.02	0.05
MAE	0.13	3.61	3.63	3.13	5.3
RMSE	0.45	4.77	4.79	4.23	7.5
r	0.77	0.68	0.72	0.64	0.55

The spatial distribution of pH was characterized by an increasing trend from south to north, pH in most parts of the north is less than 6.0 while is more than 6.5 in most parts of the south. AP and AK have similar spatial distribution, in contrast with pH, they show decreasing trend from south to north. In addition, it is difficult to discern a trend in OM and AN. OM content in the north border is higher, and AN content is higher in the centre of study area.

3.3   Evaluation Result

Based on the method of AHP and fuzzy mathematics, the tobacco planting suitability of study area is classified into four grades. The high suitability of IQI ranges from 3.4 to 4, while the suitability of IQI ranges from 2.9 to 3.4, the moderate suitability of IQI ranges from 2.4 to 2.9, the not suitability of IQI ranges is less than 2.4. As shown in Fig.6, the high suitable tobacco planting parts are mainly distributed in the southern part of the study area, accounting for 26.1% of the whole investigated areas. The suitable parts are 53% of the whole areas, mainly distributed in the northern and middle parts of study area. In addition, the moderate and not suitable parts are mainly distributed in the northeast of the study area, which are of 16.4% and 4.5% of the whole area, respectively.
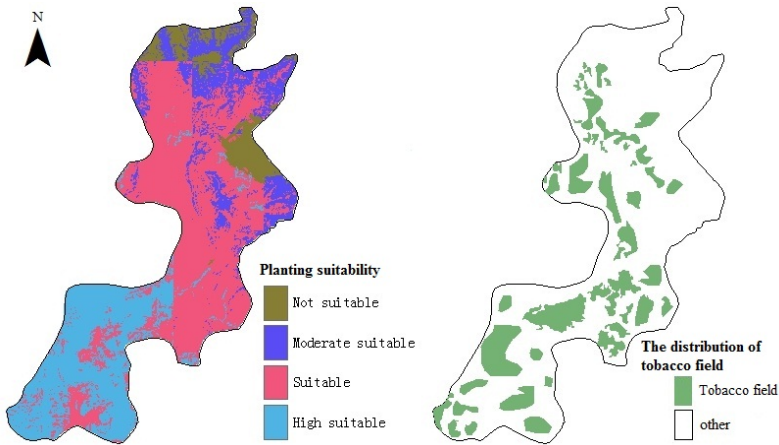
Fig. 6. The evaluation result of tobacco planting suitability and scale analysis

Then, we analyzed the suitable condition of the current tobacco fields on the basis of the evaluation results above. In summary, most of the current tobacco fields were suitable for tobacco planting. 79.8% of which were located in suitable and higher suitable areas. However, there were more than 20% of tobacco fields which were not suitable for tobacco planting; these would result into the decline of tobacco quality. Beyond that, the size of 22.4% of tobacco fields were less than 6.7 ha, which didn't meet the demand of the large scale in modern tobacco agriculture. Therefore, it is necessary to reset the spatial distribution of tobacco planting in study area.

In modern tobacco agriculture, the tobacco fields not only are suitable for tobacco planting, but also have a certain scale, which can improve the quality and standardization of tobacco leaves. Considering the tobacco planting condition and arable land in study area, 6.7 ha was selected as the scale criterion of tobacco fields, the analysis chart and its result are shown in Fig.2 and Fig.6, respectively. There were 6324.5 ha of tobacco fields meeting the requirement of modern tobacco agriculture, which were mainly distributed in the middle and southern parts of the study area.

There are two main reasons for this. 1) The current layout of tobacco planting is based on the planning of the 1980s. Subject to conditions at the time, the suitability

evaluation was performed at larger small, and the indicators selected were simple and parts of these were qualitative description. In addition, owing to the rapid development of China's economic, the environment conditions have changed a lot in past 30 years. 2) The tobacco planting is mainly based on family unit in this country, the family can planted tobacco according to the distribution of their field optionally; therefore, some tobacco was planted in a little field without considering the demand of scale production.

4   Discussion and Conclusion

The integration of GIS, RS and GPS provided a powerful tool for this study in terms of tobacco planting suitability evaluation. Ten indicators from three respects of climate, terrain and soil nutrients were selected to build evaluation indicator system. In order to measure the relation between indictors and tobacco planting accurately, the membership function and AHP method were adopted to determine the suitability of indicators and their weights. Then, tobacco planting suitability evaluation of study area was calculated by IQI model. The results show that most parts of study area were suitable for tobacco planting.

  General, traditional suitability evaluation is on provincial level or state level, counties or townships are often used as the basic evaluation units. In China, most of suitable areas for tobacco planting are located in mountainous region, where the environment variables such as air temperature, precipitation, sunshine, soil nutrients and so on have significant difference in micro-region. Large-scale evaluation is very difficult to reflect these differences and not meet requirements of modern precision agriculture. While small-scale evaluation can solve these problems, it can reflect the differences of crop growth environment micro-region and makes the evaluation results more accurate.

  The distribution density of meteorological stations is about 50*50 km per station in China, so it is difficult to attain the accuracy spatial distribution of meteorological data by observation station. Remote sensing can acquire surface parameters with high spatial resolution in any region, which can compensate for the loss caused by the absence of observation station.

  In order to enhance market competitiveness of tobacco, it should be improve the quality of tobacco leaves. The current tobacco production is mainly based on family unit, and each family adopted his planting mode, which is difficult to attain this target. The scale planting is an effective way to deal with this problem. It can be conducive to mechanized production, improve farm management level, and save labor and reduce production costs, which will help to achieve standardization of tobacco production. However, the current evaluation method which only evaluates the suitable level of each location is powerless to do these. In this study, through integrating suitability evaluation and arable land, the scale analysis was developed by spatial analysis of GIS.

  After analyzing the current tobacco fields, it was found that more than 20% of tobacco was planted in moderate and not suitable area, therefore, the spatial

distribution of tobacco planting needed to be adjusted. Based on scale analysis for arable land, we found that the tobacco fields which were suitable for tobacco planting and had a certain scale mainly distributed the middle and southern parts of study area, accounting for 75.3% of arable land.

References

1. Lund, E., Christy, C., Drummond, P.: Practical applications of soil electrical conductivity mapping. Precision Agriculture 99, 771–779 (1999)
2. McBratney, A., Whelan, B., Ancev, T., Bouma, J.: Future directions of precision agriculture. Precision Agriculture 6, 7–23 (2005)
3. Qi, Y., Darilek, J.L., Huang, B., Zhao, Y., Sun, W., Gu, Z.: Evaluating soil quality indices in an agricultural region of Jiangsu Province, China. Geoderma 149, 325–334 (2009)
4. Zhang, Q., Yang, Z., Li, Y., Chen, D., Zhang, J., Chen, M.: Spatial variability of soil nutrients and GIS-based nutrient management in Yongji County, China. International Journal of Geographical Information Science 24, 965–981 (2010)
5. Mesev, V.: Integration of GIS and remote sensing. John Wiley & Sons Inc., New York (2007)
6. Burrough, P.A., Frank, A.U.: Concepts and paradigms in spatial information: are current geographical information systems truly generic? International Journal of Geographical Information Science 9, 101–116 (1995)
7. Peng, G., Deng, L., Cui, W., Ming, T., Shen, W.: Remote sensing monitoring of tobacco field based on phenological characteristics and time series image—A case study of Chengjiang County, Yunnan Province, China. Chinese Geographical Science 19, 186–193 (2009)
8. Perotto-Baldiviezo, H., Thurow, T., Smith, C., Fisher, R., Wu, X.: GIS-based spatial analysis and modeling for landslide hazard assessment in steeplands, southern Honduras. Agriculture, Ecosystems & Environment 103, 165–176 (2004)
9. Kamusoko, C., Aniya, M.: Land use/cover change and landscape fragmentation analysis in the Bindura District, Zimbabwe. Land Degradation & Development 18, 221–233 (2007)
10. Lane, C.R., Brown, M.T.: Diatoms as indicators of isolated herbaceous wetland condition in Florida, USA. Ecological Indicators 7, 521–540 (2007)
11. Figueroa, F., Sanchez-cordero, V., Meave, J.A., Trejo, I.: Socioeconomic context of land use and land cover change in Mexican biosphere reserves. Environmental Conservation 36, 180–191 (2009)
12. Jingxun, L., Xiaohui, L., Changqing, C.: Analysis and Comprehensive Evaluation of Soil Fertility Status for Tobacco-Growing Areas in the North-west of Hu'nan province. Chinese Agricultural Science Bulletin 72, 34–41 (2009) (in Chinese)
13. Chen, H.S., Liu, G.S., Yang, Y.F., Ye, X.F., Shi, Z.: Comprehensive Evaluation of Tobacco Ecological Suitability of Henan Province Based on GIS. Agricultural Sciences in China 9, 583–592 (2010) (in Chinese)
14. Zhang, J., Zhang, J., Liu, C., Zhao, B.: Comprehensive Assessment on Ecological Adaptability of Flue-cured Tobacco in Shandong Province. Chinese Tobacco Science 29, 11–17 (2008) (in Chinese)

15. Kui-dong, L., Huang, W., Han-qian, X., Chao, Z., Bai-cheng, X.: Application of climatic factors' gridding technology in tobacco-planting regionalization in Hunan Province. Chinese Journal of Ecology 2, 235–242 (2008) (in Chinese)
16. Liu, G.: Tobacco Cultivation. China Agriculture Press, Beijing (2003) (in Chinese)
17. Javier, G.C.: Vectorial algebra algorithms fox calculating terrain parameters from DEM and solar radiation modeling in mountain terrain. Geographical Information Science 17, 1–23 (2003)
18. Lifen, Z., Yongzhong, T., Tianxiang, Y., Zemeng, F., Shengnan, M., Ying'an, W.: Simulation of solar radiation on ground surfaces based on 1 km grid-cells. Transactions of The Chinese Society of Agricultural Engineering 5, 45–53 (2005) (in Chinese)
19. Fu, P., Rich, P.M.: A geometric solar radiation model with applications in agriculture and forestry. Computers and Electronics in Agriculture 37, 25–35 (2002)
20. Campbell, J.B.: Introduction to remote sensing. The Guilford Press, New York (2007)
21. Zhengming, W., Dozier, J.: Land-surface temperature measurement from space: Physical principles and inverse modeling. IEEE Transactions on Geoscience and Remote Sensing 27, 268–278 (1989)
22. Li, Z.L., Becker, F.: Feasibility of land surface temperature and emissivity determination from AVHRR data. Remote Sensing of Environment 43, 67–85 (1993)
23. Sandholt, I., Rasmussen, K., Andersen, J.: A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. Remote Sensing of Environment 79, 213–224 (2002)
24. Goovaerts, P.: Geostatistics for natural resources evaluation. Oxford University Press, New York (1997)
25. Isaaks, E., Srivastava, R.: An introduction to applied geostatistics. Oxford University Press, New York (1989)
26. Cressie, N.: Statistics for spatial data. John Wiley & Sons, New York (1993)
27. Hu, G., Li, Z., Wang, Z., Zhao, Q., Zheng, W.: Nutrition Principle of Flue-cured Tobacco. Science Press, Beijing (2000) (in Chinese)
28. Xi, Z.: Potassium content of flue-cured tobacco from different tobacco production areas. Chinese Tobacco Science 13, 13–16 (2002) (in Chinese)
29. Saaty, T.L.: The analytic hierarchy process: planning, priority setting, resource allocation. McGraw-Hill International Book Co., New York (1980)
30. Lai, V.S., Wong, B.K., Cheung, W.: Group decision making in a multiple criteria environment: A case using the AHP in software selection. European Journal of Operational Research 137, 134–144 (2002)
31. Rezaei-Moghaddam, K., Karami, E.: A multiple criteria evaluation of sustainable agricultural development models using AHP. Environment, Development and Sustainability 10, 407–426 (2008)
32. Doran, J., Parkin, T.: Defining and assessing soil quality. Precision Agriculture 35, 3–21 (1994)

Study on Spatial Model and Service Radius of Rural Areas and Agriculture Information Level in Yellow-River Delta

Yujian Yang[1,*], Guangming Liu[2], Xueqin Tong[1], and Zhicheng Wang[1]

[1] S & T Information Engineering Technology Center of Shandong Academy of Agricultural Science, Information center of agronomy College of Shandong University Jinan 250100, P.R. China
yyjtshkh@gmail.com

[2] Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, P.R. China

Abstract. Based on the evaluation methods and systems of information measurement level, and according to the principles of agriculture information subject, the study constructed 13 indices system for the measurement of the rural areas and agriculture information level in Yellow-river Delta in 2007. Spatial autocorrelation model of rural areas and agriculture information of 19 country units showed that the comprehensive information level of Hanting, Shouguang, Guangrao, Bincheng, Huimin, Wudi and Yangxin country unit is very significant, has the obvious spatial agglomeration and homogeneity characteristics, but information level agglomeration of Kenli country and Zouping city has the significant heterogeneity, and information level agglomeration characteristics of other 10 country units is not significant. The radius surface of the complicated information level from radial basis function model indicated that rural areas and agriculture information service has a certain service radius, the distance of service radius in theory is 30Km, the gradient and hierarchy is obvious. According to it, the comprehensive service node should be established in Bincheng district and the secondary service node should be set up in Wudi country for improving the service efficiency. Combining with GIS grid technology, the profile line results of rural areas and agriculture information service from centroid coordinates based on 19 country units illustrated the fluctuation characteristics of the comprehensive information level. For yellow-river Delta, the high-efficiency ecological zone construction should consider the gradients of information service, continuity of information service and spatial agglomeration characteristics of 19 country units. Correspondingly, the regional policy which reflected the regional difference and association characteristics was carried out to achieve the leap development.

Keywords: Rural areas and agriculture information level, Service Radius, Measurement indices, Radial basis function, Yellow-river Delta.

---

[*] Corresponding author.

1   Introduction

Rural areas and agriculture Information is not only the inevitable trend of the modern agricultural development, but also an important component of the national economy information (Zang Chunrong et al., 2004). Rural areas and agriculture information assessment is to evaluate the level and effect of rural information construction and development level by adopting mathematical statistics(such as AHP method, fuzzy theory study), operation research principle and specific indicator system (Liu Shihong et al., 2007; Du Jing et al., 2010). Evaluating rural information level comprehensively, measuring and grasping the information development level accurately will help to find problems and shortcomings in the process of rural information. Rural information assessment is a huge system engineering that synthesizes technological and institutional innovations. Its implementation involves all aspects of rural business and information technology (Qin Xiangyang et al., 2008)[1,2,3,4,5,6,7]. Rural areas and agriculture information constructions reflected the science and technology progress and the main promotion power of agriculture economy(Chen Xiwen, 2010), the present study showed that the correlation domains involved in the evaluation indices system of rural areas and agriculture information, but the measurement evaluation of it should be strengthened on the regional scale to intensify the agriculture resources and high-efficiency utilization of rural areas and agriculture information. The regional measurement of rural areas and agriculture information expanded the open degree and propelled the regional economy development, the service scope and service radius study of information level is rarely on the regional scale from the correlation documents[8].

There is crucial to select the service node of rural areas and agriculture information level, whether in theory, or in practice for the comprehensive service system of information level. The study selected the regional economy body—Yellow river delta, as demonstration site, and explored the regional effects of rural areas and agriculture information level. Furthermore, the gradient and hierarchy of rural areas and agriculture information service was developed in the study, with the support of spatial analysis technology, the important service node of rural area and agriculture information level where the suitable location is selected was analyzed to improve the service efficiency in Yellow river delta. And other emphasis study is the fluctuation characteristics of the comprehensive information level of 19 country units in the paper.

2   Material and Methodology

2.1   Study Area and Indices System Establishment

Yellow river delta was formed on the basis of the yellow river alluvial plain and the adjacent marine regions in the north parts of Shandong province, extension the round regions, including Dongying city, Binzhou city, Hanting, Shouguang city, Changyi city in Weifang city, Leling and Qingyun in Dezhou city, Gaoqing in Zibo city and Laizhou in Yantai city. The total area is 26500Km^2, the 1/6 of the total province. Yellow river Delta is the high-efficiency ecological zone, which has the important strategic significance, it located in the north emphasis of the whole strategy in Shandong

province and is the important part of "blue economic zone of peninsula", the comprehensive development of rural areas and agriculture information level in Shandong province is crucial part of the national demonstration province of rural areas and agriculture information on 11th and 12th five year development planning of China.

Indices system establishment of rural areas and agriculture information referred to the new countryside planning of socialism construction from development and reform commission of Shandong province "developed production, affluent life, civilized rural atmosphere, clean and tidy village and democratic administration." Combing with well-off indices from Wu Dianting (2006), and considering the measurement system rule, study data were effectively collected and the selected 13 indices was applied in the study[9].The 6 indices reflected material life (Enger coefficient, per house area of household, tap water countryside, household car, hospital numbers, urbanization). The 3 indices reflected spiritual life (telephone communication countryside, cable television countryside, patents per year). The income and distribution situation, included 2 indices（per capita net income and per capita GDP. Population quality considered 2 indices (students in middle school and per capita expenditure of education, science-technology and culture). In circumstances, Data acquisition, elimination dimension, the determination of indices weight, model establishment, the comprehensive indices computerization and comparison analysis, a series of steps were carried out to deal with the correlation data. As mentioned, measurement indices system was used to explore the autocorrelation characteristics and service radius of rural areas and agriculture information level[10,11].

2.2   Moran's I Coefficient and Radial Basis Function

Spatial autocorrelation can be defined as the coincidence of value similarity with site similarity. As the name suggests, autocorrelation is the correlation of a variable with itself, the autocorrelation coefficient (Moran's I) was used as description about the spatial dependency and was selected to describe the correlation degree of the location attribution value. The Moran's I coefficient is given by:

$$I = \frac{n \sum \sum (x_i - \bar{x})}{W \sum (x_i - \bar{x})^2} \tag{1}$$

Where x_i is the data value at location i, h is the distance between locations i and j, w_{ij} takes 1 if the pair (i, j) pertains to distance class h (the one for which the coefficient is computed), otherwise 0, W is the sum of w_{ij}. Where: $I \in [-1,1]$, $I = 0$ stands for the independent variables, $I > 0$ stands for the positive correlation, $I < 0$ stands for the negative correlation[12,13,14]. Moran's I coefficient reflecting spatial dependency embodied the similarity of nearby observations. If high values of an attribute tend to cluster together in some parts of country unit and low values tend to cluster together in other parts, the attribute is said to exhibit positive spatial autocorrelation. Conversely, if high values tend to be found in close proximity to low values and vice versa, the attribute is said to exhibit negative spatial autocorrelation[15].

Radial basis function (RBF) enables you to create a surface that captures global trends and picks up the local variation, which helps in cases where fitting a plane to the sample values will not accurately represent the surface. To create the surface, suppose you have the ability to bend and stretch the predicted surface so that it passes through all of the measured values. The study uses a set of n basis functions, one for each data location. The predictor is a linear combination of the basis functions.

$$\hat{Z}(s_0) = \sum_{i=1}^{n} \omega_i \phi(\|s_i - s_0\|) + \omega_{n+1} \qquad (2)$$

Where $\phi(r)$ is a radial basis function, $r = \|s_i - s_0\|$ is Euclidean distance between the prediction location s_0 and each data location s_i and $\{\omega_i : i = 1,2,3\cdots,n+1\}$ are weights to be estimated, ω_{n+1} is a bias parameter[16].

RBF methods are a series of exact interpolation techniques, that is, the surface must go through each measured sample value. RBF are conceptually similar to fitting a rubber membrane through the measured sample values while minimizing the total curvature of the surface. The selected basis function determines how the rubber membrane will fit between the values. RBF can predict values above the maximum and below the minimum measured values as in the cross section below. The optimal parameter is determined using cross-validation in a similar manner. The RBF are also used for calculating smooth surfaces from data points. The functions produce good results for gently varying surfaces[17].

3   Results and Analysis

3.1   Moran's I Coefficient Analysis

The heterogeneity and dependence of rural areas and agriculture information level was analyzed by Moran`s I computerization, Moran`s I value is 0.2424, the positive spatial autocorrelation, Monte-Carlo tests showed the significant correlation of information level in confidence interval of 99%, the rural areas and agriculture information measurement on basis of 19 country units has the strong agglomeration characteristics in Yellow river delta in 2007. As illustrated in Fig. 1.

Hanting, Shouguang and Guangrao(High-High), which displays the country units with a high information level (above the average) surrounded by country units with high information level (above the average) Bincheng, Huimin, Wudi and Yangxin(Low-Low), which presented the country units with low information level surrounded by country units with low information level. In the study, information level of the 7 country units presented the spatial clustering and homogeneity, information level of Kenli (Low-High) country showed the country units with low value surrounded by country units with high values, the evident difference existed in them. Zouping (High-Low) shows the country units with high value of information indices surrounded by country units with low value of information indices, the two country units presented the heterogeneity. The other information indices of 10 country units has not obvious significance.
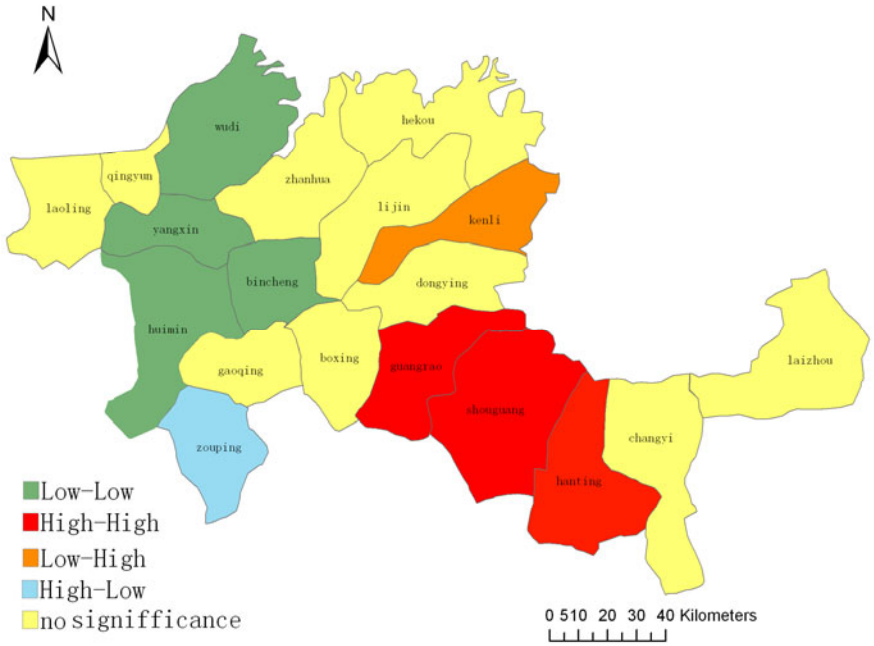
Fig. 1. Scatterplot map of measurement indices of rural areas and agriculture information level

The regional spatial pattern of rural areas and agricultural information undertook the historical consistence, the better economic foundation, relative location advantage and the regional development policy, these factors have a large effect on the spatial pattern of rural areas and agricultural information level. Combining with Moran`s I computerization results, showed that Hanting, Shouguang and Guangrao has the high level of information level, promoting the information coordination development of the surrounding country units, such as Zouping city, Kenli country. For the low information level of Bincheng district, Huimin country, Wudi country and Yangxin country, should mainly depend on the sparking sites of rural areas and agricultural information to promote service efficiency. Especially, the policy power has a large effect on shortening the difference.

3.2   Service Radius Surface

Clustering characteristics represents the spatial effect of adjacent regions, clustering characteristics of rural areas and agriculture information level provide the foundation of information service radius, that is to say, high information level region tend to be clustered close to the high information level region, the low information level region tend to be clustered close to the low information level region. The gradients and continuity of rural areas and agriculture information level need explore the sound analysis by spatial model. In the study, the centroid coordinates extracted from 19

country units on basis of spatial analysis were regarded as the data points of rural areas and agriculture information level. Owing to the gently varying of information level, the precision of interpolation surface was improved with the support of RBF, as showed in Fig.2.

On the whole, rural areas and agriculture information level in Yellow-river Delta has spatial variability, there is obvious gradients of information level from the weight value(-7.597) to the weight value(13.657) depending on RBF analysis of the comprehensive level in 19 country units. As shown in Figure 2, Bincheng district, the "sink" area of information level, the "peak" area around Dongying, the service radius of rural areas and agriculture information level by GIS calculations, about 30Km. In order to upgrade the whole pattern of information service nodes and improve the service efficiency of information level, combining with the service radius, The comprehensive service node should be established in Bincheng district and the secondary service node should be set up in Wudi country, the kind of the spatial pattern had not only important decision significance, but also provided the foundation for constructing the better hierarchy service system of rural areas and agriculture information level in Yellow-river delta.
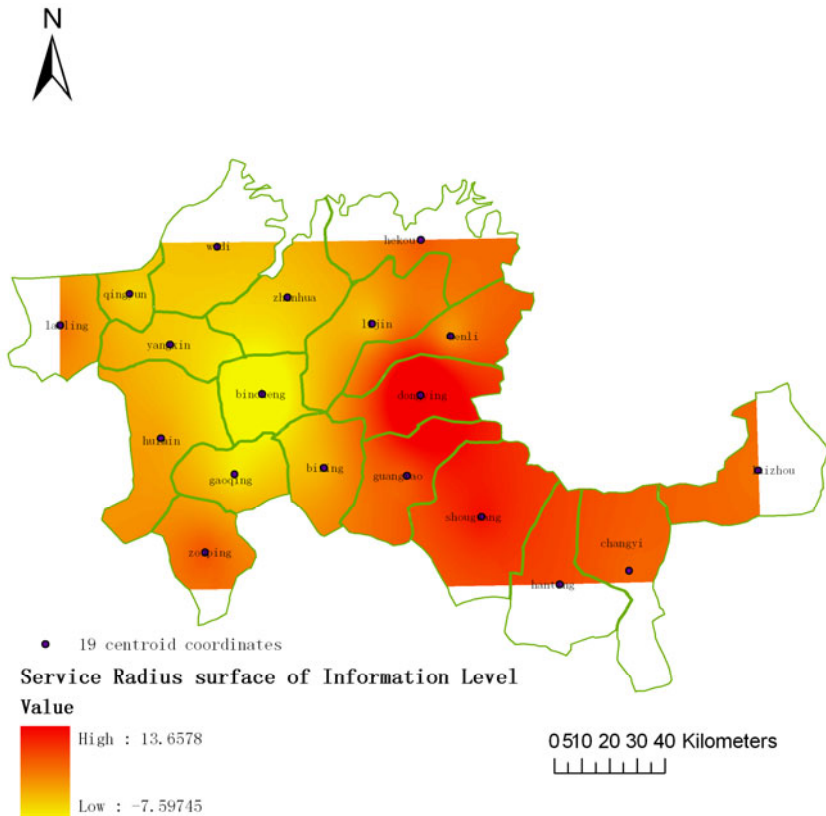
Fig. 2. The Service radius surface of rural areas and agriculture information level

Based on GIS grid technology[18], as the foundation of centroid coordinates of 19 country units, according to the following sequence, Laizhou, Changyi, Hanting, Shouguang, Guangrao, Dongying, Kenli, Hekou, Lijin, Zhanhua, Wudi, Qingyun, Leling, Yangxin, Bincheng, Gaoqing, Boxing, Zouping and Hui min, the profile line of rural areas and agriculture information level was drawn, the results referred to Fig. 3. Results showed that there is the fluctuation characteristics of rural areas and agriculture information level in Yellow river delta. The "peak" area is in Dongying and Shouguang, etc., the "sink" area is Bincheng district, the quantity description and sound fluctuation characteristics of information level is as the following contents, the weight value of the comprehensive information level referred to the brackets values, and represented the service level in different country units. Laizhou(0.707), Changyi(1.354), Hanting(2.456), Shouguang(4.592), Guangrao(1.947) Dongying(13.642), Kenli(-0.576), Hekou(1.678), Lijin(-2.249),Zhanhua(-2.879),   Wudi(-2.358),   Qingyun(-3.174),   Leling(0.553), Yangxin(-2.806), Bincheng(-7.593), Gaoqing(-3.981), Boxing(-1.840), Zouping(1.791) and Hui min(-1.263).The profile line also indicated the information level of other regions excluded by 19 country units and there is the continuous distributions of information level, which is consistent with the service surface results of rural areas and agriculture information level from RBF surface results.
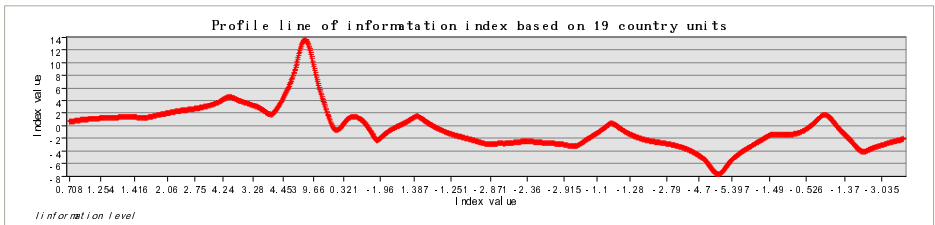
Fig. 3. The profile line of rural areas and agriculture information level in 19 country units

4   Conclusions and Discussion

The paper constructed 13 indices measured the rural areas and agriculture information level in Yellow-river Delta in 2007. Moran's I Coefficient analysis of rural areas and agriculture information level of 19 country units showed that the comprehensive information of 7 country units is significant, has the obvious spatial agglomeration and homogeneity characteristics, information level of Kenli and Zouping has the significant heterogeneity. The continued spatial surface from radial basis function method indicated that rural areas and agriculture information service in theory is 30Km, the gradient and hierarchy is obvious. The comprehensive service node should be established in Bincheng district and promote the development of information level, the secondary service node should be set up in Wudi country for improving the service efficiency according to the study. Combining with GIS grid technology, the profile line results of rural areas and agriculture information service of centroid coordinates of 19 country units illustrated the fluctuation characteristics of the comprehensive information level.

The service of rural and agricultural information is the inevitable trend of the world modern agricultural development, we should emphasize the comprehensive platform construction for rural areas and agriculture information services, support platform making full use of existing low-cost, mature means of communication and platform, multiple platform integrated organic body. In essence, the service construction of information platform of rural areas and agriculture information in Yellow-river delta is the branch platform and sub-nodes of the rural areas and agriculture information platform in Shandong province, so we should perfect the sound measurement system and the evaluation system of rural areas and agriculture information level, and especially measurement system on the regional scale, correspondingly, improve and strengthen the quality of rural areas and agriculture information service in demonstration site, promote the whole upgrading of the information level in Shandong province.

References

1. Zang, C., Chen, X., Liu, S., et al.: Present situation analysis of China rural informatization development. Journal of Hunan Agricultural University(Social Sciences) 4(5), 10–13 (2004) (in Chinese)
2. Liu, S.: Study on the indicator system for measuring the rural area information level in China. Library And Information Service 51(9), 33–35 (2007)
3. Mei, F.: Choice of development model of low-cost rural informationization in China. Journal of China Information Times (5), 12–13 (2007) (in Chinese)
4. Jing, D., Li, D., Li, H., Zhang, Y.: Research on assessment method for rural informatization level based on AHP. In: Li, D., Yang, S.X. (eds.) Computer and Computing Technologies in Agriculture II, vol. 24, pp. 125–134 (2010)
5. Lu, L.: Construction of index system for IT application in agriculture. Journal of Library and Information Sciences in Agriculture 19(4), 178–183 (2007)
6. Qin, X., Zhang, X., Zhang, X.: Study on assessment index system for rural informatization in Beijing. Journal of Beijing Agricultural Vocation College 22(1), 42–46 (2008)
7. Tan, G., He, W.: A study on contents, obstacles and countermeasures of rural informatization in China. Journal of Jiangxi Agricultural University 6(2), 86–88 (2007)
8. http://news.xinhuanet.com/politics/2010-01/31/content_12908364.htm
9. Wu, D.: Regional Analysis and Planning Turtorials. Education Press, Beijing (2004)
10. Statistical Bureau of Shandong Province, China county (city) Socio-economic Statistical Yearbook. China Statistics Press, Beijing (2008) (in Chinese)
11. Statistical Bureau of Shandong Province, Shandong Statistical Yearbook. China Statistics Press, Beijing (2008) (in Chinese)
12. Pu, Y., Ge, Y., Ma, R., Huang, X., Ma, X.: Analyzing regional economic disparities based on ESDA. Geographical Research 24(6), 965–974 (2005) (in Chinese)
13. Wu, Y., Li, J.: A spatial econometric analysis of industrial total factor productivity in China's provincial regions. Scientia Geographica Sinica 26(4), 385–391 (2006) (in Chinese)
14. Anselin, L., Florax, R.J.G.M., Rey, S.J.: Advances in spatial econometrics: methodology, tools and applications. Springer, Berlin (2004)

15. Stewart Fotheringham, A., Brunsdon, C., Charton, M.: Geographically weighted regression the analysis of spatially varying relationships
16. Haining, R.P.: Spatial data analysis: theory and practice. Cambridge University Press, Cambridge (2003)
17. Using ArcGIS Geostatistical Analyst. GIS by ESRI (2008)
18. Yang, Y., Zhu, J., Zhao, C., Liu, S., Tong, X.: The spatial continuity study of NDVI based on Kriging and BPNN algorithm. Mathematical and Computer Modelling (54), 1138–1144 (2011)

The Effect of Intercropping of Maize and Soybean on Microclimate[*]

Hanming He, Lei Yang, Liming Fan, Lihua Zhao, Han Wu,
Jing Yang, and Chengyun Li[**]

Key Lab. of the Ministry of Education for Agro-Biodiversity and Pest Control,
Yunnan Agricultural University, Yunnan, Kunming 650201, China
`licheng_yun@163.com`

Abstract. Intercropping induces the diseases decreasing, and yield increasing, may partly due to the improvement of microclimate in fields. In order to understand the mechanism and efficiency of resource utilization in intercropping of maize (Zea mays) and soybean (Glycine max), a field experiment was conducted as factorial on the bases of randomized complete block design of three patterns with three replications. Three cropping patterns were maize monocropping (A), 2 rows maize and 2 rows soybean intercropping (C) and 2 rows maize and 4 rows soybean intercropping (D). Our studies showed that compared with monocropping, the temperature in intercropping was a little higher in the daytime, but in the nighttime, the contrary results were observed; the relative humidity in intercropping was lower in the daytime, but in the nighttime, the contrary results were observed; the light intensity in intercropping was markedly higher than that in monocropping. The yield components of maize in intercropping, including thousand kernel weight, yield per plant and leaf area were increased than that in monocropping. These results imply that microclimate variation of intercropping probably play important role to maize yield increasing.

Keywords: Intercropping, Temperature, Relative humidity, Light intensity, Biological characters.

Introduction

Compared with the monoculture, there may be more effectively utilize the temporal-spatial remainder of intercropping crops growth and development to bring into play the production potential of limited agricultural resources including radiation, fertilizer, water, gas and heat[1-3]. Therefore intercropping plays an important role in agricultural productions.

The microclimate including temperature, relative humidity (RH) and light intensity in farmland is an important factor in the growth and production of crops. Previous studies showed that the relative humidity, which played vital roles in disease injuries of

---

crops, was found to present downtrend and reducing the number of hours per day with relative humidity ≥92% in intercropping[3-4]. And the intercropping can increase the amount of light intercepted of crops in unit planting area so that improve the crop dry matter production, yield and radiation use efficiency[5-10]. So we explore the microclimate change regulation in multi-culture patterns of maize and soybean and compared with monocropping so to discover the mechanism of intercropping.

1   Materials and Methods

Field experiments were conducted at the farm of Yunnan Agricultural University, Yunnan province, China, 25°01′N, 103°E during 2009/05/06 to 2009/10/22. the crops used in the experiment were   maize (Yunrui 88) and soybean (Nandou 12), and were sown by north-south rows at the same time. The experiments was made up of three cropping modes, one of which was maize monocropping (A, row distance 40cm), two of which were maize intercropping, viz., 2 maize rows (row distance 50cm) and 2 soybean rows (row distance 50cm) (2:2)- (C) which distance of maize to nearest soybean row is 50cm, and 2 maize rows (row distance 35cm) and 4 soybean rows (row distance 30cm) (2:4)- (D) which distance of maize to nearest soybean row is 40cm. Maize plant distance is 20cm and soybean plant distance is 30cm in all cropping patterns. Experiments repeated 3 times which using two-factor completely block design. Every unit area was 4.0cm ×5.0cm. Full irrigation and fertilizer were applied for every cropping system.

   HOBO U12-012 data loggers were put beside maize plants and soybean plants to measure temperature ranged from -20℃ to 70℃, RH ranged from 5% to 95%, light intensity ranged from 1 to 3000 footcandles (lumens/ft^2). The light intensity beyond 3000 footcandles was recorded as the max value. But the light distributing difference between mono-cropping and intercropping focus mostly in the upper, middle and botten part, and the light intensity is usually in the measurement range. The Li-6400 portable photosynthesis was used to measure the photosynthetic rate of maize leaves.

   The temperature, RH and light intensity on the position between maize and soybean, above ground 30cm (below the head of soybean canopy) and 70cm (above the head of soybean canopy) of field were measured at maize heading stage in our experiment. The HOBO loggers recorded the data one time every 30 minutes and the average of that during a period was obtained as the daily changes of temperature, RH and light intensity.

2   Results and Discussion

2.1   The Difference of Temperature and Relative Humidity Beside Maize Plant at Heading Stage in the Monocropping and Intercropping

Fig.1-(a) and (b) indicated that the similar trend of the temperature beside maize plants above ground 30cm and 70cm between monocropping and intercropping was observed,

viz., temperature in the daytime was higher than that in the nighttime. Wherever above ground 30cm or 70cm, in the daytime the temperature in intercropping was a little higher than that in monocropping, and the temperature in intercropping D was the most highest, followed by that in intercropping C and that in monocropping A. But in the night time, the contrary results were observed. The temperature in the nighttime in intercropping was a little lower than that in monocropping, and the temperature in intercropping D was the most lowest, followed by that in intercropping C and that in monocropping A. Fig.1-(c) and (d) also illustrated that similar daily change regulation of RH beside maize plants above ground 30cm and 70cm between monocropping and intercropping was also observed as the temperature, viz., the RH in daytime was lower than that in nighttime. Both above ground 30cm and 70cm, the RH in the daytime in monocropping A was the most highest, followed by that in intercropping D and that in intercropping C. But in the nighttime, the contrary results were observed. The RH in the nighttime above ground 30cm in intercropping D was the highest, followed by that in intercropping C and that in monocropping A. Above ground 70cm, the highest RH was found in intercropping D, and the lower one was found in monocropping A and in intercropping C. The results indicated that the microclimate in field such as temperature and RH in the intercropping were improved compared with that in monocropping.
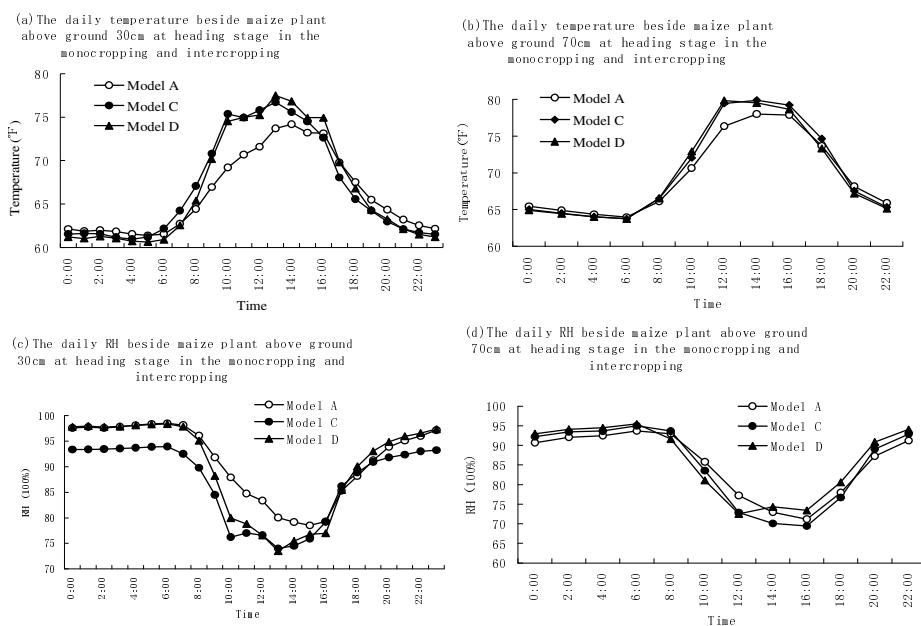
Fig. 1. Daily change of temperature and RH beside maize plants above ground 30cm and 70cm in monocropping and intercropping at heading stage

2.2   The Change of Light Intensity Beside Maize Plants above the Ground 30cm and 70cm in Monocropping and Intercropping at Heading Stage

Fig.2-(a) illustrated the light intensity during July 4th to July 19th, when the head of soybean canopy is below 30cm and the difference of light intensity was presented evident. Fig.2-(a) and (b) indicated that there were similar (parabolic curve) trends of daily light intensity above the ground 30 cm and 70cm beside maize plants in monocropping and intercropping. In other words, the maximum was found in the noon, while the minim was observed in the morning and evening in whatever cropping systems, which were consistent to the rules in meteorology. Moreover, light intensity in intercropping was significant higher than that in monocropping, especially with the increase of radiation angle.

Fig. 2. Daily light intensity change beside maize plants above ground 30cm and 70cm and the relationship between light intensity and photosynthetic rate in monocropping and intercropping at heading stage

Fig. 2-(a) indicated that above ground 30cm in monocropping only from 11:30 am to 13:30 pm, the light intensity was above the compensation points. But in intercropping (C and D), the light intensity from 8:30 am to 15:30 pm was always above the light compensation points. It was implied that the time length of light intensity above the compensation points was significantly longer (nearly 5 hours) than that in monocropping. Further more, the light intensity was over 400 footcandles in intercropping from 10:00 am to 4:00 pm. Therefore, no matter in the time length and the intensity, effective radiation among the maize canopy above the ground 30cm in intercropping was markedly

higher than that in monocropping during July 4th to July 19th. We conducted a Pearson correlation between daily average light intensity and the photosynthetic rate of 7th leaves (nearly on the position above ground 30cm) during July 20th to July 28th. The results showed that strong correlation (R^2=0.92, P<0.001) can be observed in Fig.2-(c).

Fig.2-(b) indicated that above ground 70cm (beyond the head of soybean canopy) in monocropping, the light intensity was significant lower than the light compensation points, viz., the light intensity was below 100 footcandles. However, in intercropping C and D, the light intensity from 9 am to 5 pm (nearly 8 hours) was above the light compensation points, and from 9 am to 4 pm (7 hours) the light intensity was more than 200 footcandles. Maybe it becaused the canopy above the ground 70 cm in intercropping was not shadowed and can fully accept radiation. In Fig.2-(d) the linear correlation (R^2=0.95, P<0.001) between the light intensity and the photosynthetic rate in ear leaves (nearly on the position above ground 70cm) was also observed in monocropping and intercropping.

2.3   Comparing Biological Characters of Maize in Monocropping and Intercropping

It was easy from above results to see that compared with monocropping, the improvement of microclimate of fields in intercropping cause the raise of the photo-synthetic rate of maize's leaves, which could resulte the improvement of the biological characters of maize.

Table 1 indicated that maize's yield characters in intercropping, including thousand kernel weight, yield per plant and area of ear leaves are all greater than that in monocropping, in which the thousand kernel weight showed D>C>A, yield per plant showed C>D>A, the area of ear leaves showed C>D>A. Accordingly, from the trial results it was seem that the biological characters of maize in intercropping were better than that in monocropping.

Table 1. Biological characters of maize in monocropping and intercropping

Biological characters / Pattern	Thousand kernel weight (g)	Yield per plant (g)	Leaf area(cm^2)
Monocropping	323.8	120.8	626.0
Intercropping 2:2	339.2	167.2	877.3
Intercropping 2:4	344.9	160.7	666.6

3   Conclusions

Intercopping leads to variation of microclimate, especially for light intensity, RH and temperature. Maize and soybean are staple crops in the world, intercropping of the

crops have long history in Asian countries, mainly for increasing the yield and nourishing the soil. However, mechanisms and efficiency of intercropping is unclear. It will be the limitation for improving intercropping system. In this study, microclimate of maize field including temperature, RH and light intensity in monocropping and intercropping were continously investigated by using Multi-channel Data Logger (Hobo H8) analyzed and compared. The results show that the microclimate was improved siginifically.

Firstly, wherever above the ground 30cm or 70cm, in the daytime the temperature in intercropping was a little higher than that in monocropping, but in the night time, the contrary results were observed. Both above the ground 30cm and 70cm, the RH in the daytime in monocropping was higher then that in intercropping, but in the nighttime, the contrary results were observed.

Secondly, from the daily change, no matter in the time length and the intensity, effective radiation among maize plants above the ground 30cm and 70cm in intercropping was markedly higher than that in monocropping. Moreover, a remarkably linear correlation between daily average light intensity and the photosynthetic rate of leaves   was observed.

Finally, the yield components of maize in intercropping, including thousand kernel weight, yield per plant and leaf area were increased than that in monocropping.

In summary, our results showed microclimate variation, including increasing the radiation duration and light intensity at different height position in intercropping field significantly, and remarkable correlation between light intensity and photosynthetic rate implied the variation of light intensity play important role for improving the yield component of maize. However, intercropping is a very complex system, biological variation of maize plants play roles for increasing the yield as well. To determine the better combination and more efficiency of resource utilization in intercropping, it is need to understand the biological, physical, chemical and microclimate effectors and their interaction.

References

1. Zhu, Y., Chen, H., Fan, J., et al.: Genetic diversity and disease control in rice. J. Nature 406, 718–722 (2000)
2. Li, C., He, X., Zhu, S., et al.: Crop Diversity for Yield Increase. PLoS ONE 4(11), e8049–e8057 (2009)
3. Zhu, Y., Li, C., et al.: Genetic Diversity for Crops Diseases Sustainable Management. Science Publisher, Beijing (2007)
4. Gómez-Rodrígueza, O., Zavaleta-Mejía, E., et al.: Allelopathy and microclimatic modification of intercropping with marigold on tomato early blight disease development. Field Crops Research 83(1), 27–34 (2003)
5. Monteith, J.L.: Climate and the efficiency of crop production in Britain. Phil. Trans. R. Soc., London 281, 277–294 (1997)

6. Sivakumar, M.V.K., Virmani, S.M.: Crop productivity in relation to interception of photosynthetically active radiation. Agric. For. Meteorol. 31, 131–141 (1984)
7. Jahansooz, M.R., Yunusa, I.A.M., Coventry, D.R., Palmer, A.R., Eamus, D.: Radiation- and water-use associated with growth and yields of wheat and chickpea in sole and mixed crops. Eur. J. Agron. 26, 275–282 (2007)
8. Tsubo, M., Walker, S., Mukhala, E.: Comparisons of radiation use efficiency of mono-/intercropping with different row orientations. Field Crops Res., 17–29 (2001)
9. Zhang, L., van der Werf, W., Bastiaans, L., Zhang, S., Li, B., Spiertz, J.H.J.: Light interception and utilization in relay intercrops of wheat and cotton. Field Crops Research 107, 29–42 (2008)
10. Tsubo, M., Walker, S.: A model of radiation interception and use by a maize–bean intercrop canopy. Agric. Forest Meteorol. 200, 203–215 (2002)

Estimating Foliar Pigment Concentration of Rice Crop Using Integrated Hyperspectral Index

Zhanyu Liu[1,2,*], Wenjiang Huang [3], Guofu Mao[4], Cunjun Li[3], Xingang Xu [3], Xiaodong Ding[1], Jingjing Shi [5], and Bin Zhou[1]

[1] Institute of Remote Sensing and Earth Sciences, Hangzhou Normal University, Hangzhou 311121, China
`liuzhanyu@zju.edu.cn`
[2] Key Laboratory of Urban Wetland and Region Change in Zhejiang Province, Hangzhou 311121, China
[3] Beijing Research Center for Information Technology in Agriculture, Beijing 100097, China
[4] Institute of Agricultural Sciences in Xiaoshan District, Hangzhou 311202, China
[5] Institute of Agricultural Remote Sensing & Information Technology, Zhejiang University, Hangzhou 310029, China

Abstract. Photosynthetic pigment concentration has strong relationship with nitrogen (N) concentration which is an essential element of plant growth and plays an important role in estimating the net primary productivity (NPP) in terrestrial ecosystem research and precision agriculture (PA). In this study, hyperspectral reflectance and pigment concentration of the upper three leaves of rice crop (*Oryza sativa* L.) with five N fertilization rates were measured in the laboratory. The results showed that there was no significant difference between the leaf hyperspectral reflectance and pigment concentrations in the visible and near-infrared (NIR) spectral regions at each leaf position. But in the shortwave infrared (SWIR) region, the difference of each two of the three leaf positions was obviously significant at level 0.05. The integrated hyperspectral index MCARI/OSAVI$_{[670,800]}$ had been proved to be better linear related with leaf pigment concentrations at different leave position. The result demonstrated that MCARI/OSAVI$_{[670,800]}$ was a reliable and stable hyperspectral index for estimating pigment concentration at leaf level.

Keywords: Hyperspectral reflectance, Nitrogen fertilization rate, Pigment concentration, Rice.

1   Introduction

Net primary productivity (NPP) is a function of an efficiency coefficient defining the carbon dioxide (CO_2) fixed in short-lived (foliage and fine roots) and long-lived (wood) tissue per unit area and time[1-2], which plays a vital role in essential materials and suitable environments for human society[3-4]. The precise estimation of NPP over a continental and global scale would facilitate an improved interpretation of dynamic physiological processes within terrestrial ecosystems research and precise

---

[*] Corresponding author.

agriculture (PA) management, and is also crucial for the research on the relationship between global climate change and carbon cycling processes[5-6].

NPP has been shown to be related to the fraction of absorbed photosynthetically active radiation (f_{apar}) which is a function of the sum of the concentrations, and more precisely the molar extinctions, of chlorophyll a (Chl *a* hereafter), chlorophyll b (Chl *b*), and the carotenoids (Cars). In addition, leaf characteristics, especially nitrogen (N) content, and specific leaf area affect NPP directly by constraining photosynthesis and transpiration rates[7-9]. Moreover, N is an essential element for plant growth and is frequently the major limit nutrient in most agricultural soils. Excessive N fertilizer may move into surface water and groundwater and accelerate eutrophication of lakes and streams[10]. Farmers must balance the competing goals of supplying enough N to their crops in PA management while minimizing the loss of N to the environment, which represents both a threat to water quality and an economic loss. Since N is the component element of chlorophyll molecules, there is a close link between leaf chlorophyll content and leaf N content[11]. Thus remote sensing techniques have the potential to evaluate the chlorophyll variability over large fields quickly[12].

Developments in hyperspectral remote sensing made it possible to estimate N and NPP variability through quantifying individual photosynthetic pigments within vegetation. Spectral reflectance measurements of corn (*Zea mays* L.) and wheat (*Triticum aestivum* L.) canopies have been used to detect characteristic of different N status of crop and provide reliable information for developing variable-rate fertilizer N application technique[13-15]. The variety of N treatments resulted in the differences in leaf chlorophyll content, leaf area index (LAI), biomass, and foliage cover which contributed to the differences in spectral reflectance at the leaf and canopy scales. The relatively subtle differences in leaf and canopy reflectance associated with changes in leaf chlorophyll are often confounded with major changes in plant growth and development due to N treatments[12].

Vegetation indices (VIs) have been developed as an attempt to reduce spectral effects caused by external factors such as the atmosphere and the soil background[16]. Peñuelas et al.[17] have advocated a shift towards narrow-band VIs for estimating the absolute and relative concentrations of Chl *a*, Chl *b*, and Cars in plant leaves. Zarco-Tejada et al.[18] performed a study of VIs for chlorophyll estimation in open-canopy forest from leaf level to the canopy through *SAIL and Kuusk* canopy reflectance model and demonstrated that nominal canopy reflectance model parameters appear to be sufficient to allow accurate application of the optical index/bioindicator algorithm to airborne data. Research by Wu et al.[19] suggested that the integrated VIs, namely, TCARI/OSAVI $_{[705,750]}$ and MCARI/OSAVI$_{[705,750]}$, were most appropriate for chlorophyll estimation with high correlation coefficients R^2 of 0.881 and 0.941, respectively, because more disturbances such as shadow, soil reflectance and nonphotosynthetic materials were taken into account.

Therefore, this study aims at examining the performance of integrated hyperspectra indices MCARI/OSAVI$_{[670,800]}$ developed by Rondeaux et al.[20] and Daughtry et al.[12] in photosynthetic pigment concentration estimation within the rice leaf. Meanwhile, the feedback response of leaf spectral characteristics to leaf pigment concentration due to N treatments will be investigated.

2   Materials and Methods

2.1   Field Experiment

The field experiment was conducted on sandy loam soil from early June to late October in 2008 at the research farm of Xiaoshan Agricultural Science Research Institute (30°20′N latitude, 120°31′E longitude and altitude approximately 6m), Hangzhou, Zhejiang Province, China. The original soil had 13.1g/kg organic C, 5.6mg/kg bicarbonate extractable P, 35.2mg/kg exchangeable K, 1.26g/kg total N and Ph 7.5 (soil: water = 1:1 (w/v)). Three rice (*Oryza sativa* L.) cultivars, namely, Yongyou 8, Zhongzheyou 1, and Zhejiang 22, were employed in this study.

A random block design was employed with the three rice cultivars and five N fertilization rates (0, 75, 180, 285 and 390 kg/ha). Forty percent of the N fertilizer (i.e. urea) was fertilized during the pre-transplanting period, 30% at the tillering stage, and 30% at the initial heading stage. Each treatment was replicated three times. The rice plants were transplanted on July 5th, 2008 and harvested on October 31st, 2008.

2.2   Hyperspectral Reflectance Data Acquisition

One whole rice plant from each plot was collected, placed in a plastic bucket with water, and transported to the laboratory for spectral measurement on August 21st, September 27th, and October 10th, 2008. The first, second and third uppermost leaves were cut from the main stem of rice plant and referred to as L1, L2 and L3, respectively. The hyperspectral reflectance of L1, L2 and L3 was measured with a portable spectrophotometer (Analytical Spectral Devices, Inc., Colorado, CO, USA) in the wavelength range of 350-2500 nm. The spectral resolution of the instrument is 3 nm for the region of 350-1000 nm and 10 nm for the region of 1000-2500 nm.

Rice leaves were positioned on a dark background so that the fiber optic sensor with a 25° instantaneous field of view (IFOV) vertically pointed to centre of the leaf surface with about 3.5 cm height, equivalent to 1.9 cm^2 observed area. An incidence angle of 45°was maintained at a standard distance of 50 cm throughout the study in a closed chamber with a halogen lamp (50 Watt). A sampling spectrum was consisted of ten readings for each leaf. The average spectra derived from the sampling spectra represented each leaf using ViewSpec Pro (version 5.6.10).

2.3   Foliar Pigment Concentration Acquisition

After the hyperspectral measurements, a leaf disk with a weight of 0.1 g was cut from each leaf for pigment analysis. Each leaf disk was crushed, and dipped in 20 ml solution (acetone: ethanol: distilled water = 4.5:4.5:1) for 24 hours in the dark environment to extract pigment. The optical density (OD) of the extraction solution was measured at 440 nm, 645 nm, and 663 nm by spectrophotometer (Shimadzu UV 2550, Tokyo, Japan). Pigment consists of Chl *a*, Chl *b*, and Car in the study. The formula of pigment concentration (mg/g) was referenced the literature of Tang et al.[21].

2.4   Data Analysis

2.4.1   Data Preprocessing

The hyperspectral reflectance was smoothed with a five step moving average to suppress instrumental and environmental noise before these data were further analyzed[22]. Then, the raw hyperspectral reflectance less than 400 nm and more than 2400 nm were ineffective due to severe instrument and system noise[23].

2.4.2   Statistical Analysis

Analysis of variance (ANOVA) is useful for assessing what proportions of the variations in a dependent variable can be accounted for by one or more independent variables[20, 24]. The hyperspectral reflectance (Table 1) and pigment concentration (Table 2) were expressed as mean and standard deviation (SD) for the remote sensing and biophysical data and were analyzed by ANOVA, followed by Duncan's multiple-range test when appropriate. Differences between groups were considered significant when $p<0.05$. Pearson's linear regression coupled to ANOVA was used to verify the effects of pigment concentration on hyperspectral reflectance and integrated hyperspectral VIs. Values of the determination coefficients were obtained by the linear regression analysis (Fig.3). ANOVA was then performed with SPSS software (Statistical Package for the Social Science, version 16.0.0).

2.4.3   Integrated Hyperspectral Index

The chlorophyll absorption ration index (CARI) developed by Kim[25] could minimize the effects of nonphotosynthetically materials on spectral estimates of absorbed photosynthetically active radiation (PAR). Daughtry et al. [12] simplified the calculation equation of CARI to obtain the modified CARI (MCARI). MCARI is the depth of chlorophyll absorption at 670 nm relative to the reflectance at 550 nm and 700 nm and is defined as the following equation:

$$\text{MCARI}_{[670,700]}=[(R_{700}-R_{670})-0.2(R_{700}-R_{550})]\frac{R_{700}}{R_{670}} \tag{1}$$

Rondeaux et al.[20] concluded that the optimized soil-adjusted vegetation index (OSAVI) could reduce the sensitivity of object reflectance to the underlying soil. OSAVI is defined as follows:

$$\text{OSAVI}_{[670,800]} = \frac{(1+0.16)(R_{800}-R_{670})}{(R_{800}+R_{670}+0.16)} \tag{2}$$

Therefore, the integrated form of MCAI and OSAVI could be defined as

$$\frac{\text{MCARI}}{\text{OSAVI}}[670,800] = \frac{[(R_{700}-R_{670})-0.2(R_{700}-R_{550})](R_{700}/R_{670})}{(1+0.16)(R_{800}-R_{670})/(R_{800}+R_{670}+0.16)} \tag{3}$$

Wu et al.[19] proved that MCARI/OSAVI is one of the most appropriated integrated hyperspectral indices for chlorophyll estimation because more disturbances such as shadow, soil reflectance and nonphotosynthetic materials are taken into account.

3   Results

3.1   Leaf Hyperspectral Reflectance

The mean (n = 45) hyperspectra reflectance of L1, L2, and L3 of rice crops collected on three sampling dates for three cultivators at five N fertilization rates were shown in Fig.1. They showed very little detail information in the full wavelength range (400-2400 nm), especially in the visible region.

To get more specific information on the difference between different leaf positions in leaf hyperspectral reflectance due to changes of leaf pigment concentration, the hyperspectral reflectance as waveband means was calculated to simulate the Enhanced Thematic Mapper Plus (ETM+) of Landsat-7 (Table 1). ANOVA followed by Duncan's test with 0.05 of alpha was used to verify the effects of pigment concentration at different leaf positions on hyperspectral reflectance (Table 1 and 2).
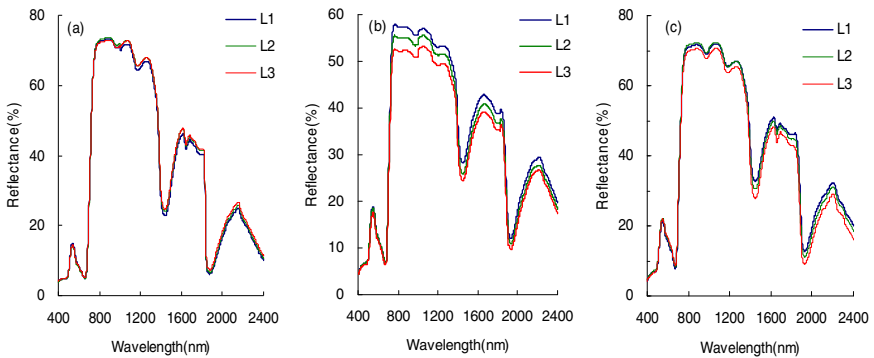
Fig. 1. Mean hyperspectral reflectance of rice crop at the first, second and third uppermost leaf (L1, L2 and L3) collected on August 21st (a),   September 27th (b) and October 10th (c), 2008

Table 1. Average hyperspectral reflectance of rice crop at the first, second and third uppermost leaves (L1, L2 and L3) at six wavelength intervals denoting the Enhanced Thematic Mapper Plus (ETM+) of Landsat-7 (Unit :%)*

Specta range	August 21st, 2008			September 27th, 2008			October 10th, 2008		
(nm)	L1	L2	L3	L1	L2	L3	L1	L2	L3
Blue(450-515)	6.0a	5.8a	6.0a	7.5a	7.5a	7.1a	8.0a	8.2a	7.7a
Green(525-605)	11.8a	11.2a	11.5a	16.0a	15.8a	14.9a	18.7a	19.3a	19.3a
Red(630-690)	6.4a	6.1a	6.5a	8.1a	8.1a	7.7a	10.2a	10.6a	11.1a
NIR(775-900)	72.7a	73.1a	72.4a	57.4a	55.1b	52.2c	71.2a	71.8a	70.0a
$SWIR_1$(1550-1750)	43.6b	44.7a	45.0a	41.4a	39.3a,b	37.7b	48.6a	47.7a,b	45.9b
$SWIR_2$(2090-2350)	19.9b	20.5a	21.4a	26.9a	25.5a,b	24.4b	28.4a	27.1a	24.8b

* Means followed by the same letter in each row are not significantly different from each other by Duncan test at
  alpha = 0.05.

Table 2 Statistical description of pigment concentration for the first, second and third uppermost leaves (L1, L2 and L3)

Date	Pigment Position	Chlorophyll a			Chlorophyll b			Carotenoids		
		L1	L2	L3	L1	L2	L3	L1	L2	L3
08/21	Mean	1.97a	2.12a	2.07a	0.66b	0.71a,b	0.75a	0.74b	0.81a	0.80a
	SD	0.32	0.33	0.45	0.12	0.14	0.16	0.12	0.12	0.13
09/27	Mean	2.40a	2.19a	2.11a	0.82a	0.80a	0.81a	0.91a	0.81a	0.81a
	SD	0.82	0.89	0.80	0.31	0.36	0.35	0.27	0.26	0.26
10/10	Mean	1.63a	1.39a	1.02b	0.53a	0.48a	0.37b	0.75a	0.65a	0.50b
	SD	0.76	0.74	0.55	0.28	0.27	0.22	0.27	0.28	0.21

* Means followed by the same letter in each row are not significantly different from each other by Duncan test at alpha = 0.05.

Photosynthetic pigments control the hyperspectral reflectance and transmittance in the visible region. Because there was almost no significant difference among L1, L2 and L3 in concentration of Chl *a*, Chl *b* and Cars (Table 2), the hyperspectral reflectance had no corresponding significant difference in the visible blue-green (450-515nm), green (525-605nm), red (630-690nm), near-infrared (NIR, 775-900nm) spectral regions except the samples collected on September 27th, 2008 as shown in Table 1.

In the shortwave infrared spectral region (SWIR$_1$:1550-1750nm; SWIR$_2$:2090-2350nm), the differences between L1 and L2, and between L2 and L3 were significant only for the latter two sampling dates. Comparing with L1 collected on September 27th, 2008, the hyperspectral reflectance of L2 decreased 5.1% and 5.2% in SWIR$_1$ and SWIR$_2$ regions, respectively; and comparing with L2, the hyperspectral reflectance of L3 decreased 4.1% and 5.3% in SWIR$_1$ and SWIR$_2$ regions, respectively(Fig.1 *b* and Table 1). Same phenomena occurred for the third sampling date (Fig.1 *c* and Table 1). But the very reverse changes in hyperspectral reflectance appeared for the first sampling date (Fig.1 *a* and Table 1).

3.2   Relationship between Hyperspectral Reflectance and Leaf Pigment Concentration

The correlation between the hyperspectral reflectance and the leaf pigment concentrations was investigated to interpret the changes of spectral characteristics through the full wavelength range (400-2400 nm) (Fig.2).

As shown in Fig.2(*a*), the hyperspectral reflectance in the green light region had greatest negative correlation with leaf pigment concentrations. The correlation coefficients were 0.735, 0.660 and 0.708 for Chl a, Chl b and Cars, respectively; and the corresponding wavebands of the former pigment located at 560nm, the latter two pigments at 545nm. In the red light region, the highest and lowest coefficients of hyperspectral reflectance with leaf pigment concentrations appeared near 667nm and

694nm wavebands, respectively. The inflexion points of correlation coefficient in the NIR region came forth near 760nm. The highest and lowest coefficients in SWIR region rose near 1458nm and 1650nm, but the former waveband would in the vapor absorption region if the hyperspectral reflectance was measured at the canopy level from the airborne or spaceborne platforms. Although the growth stages were different (Fig.2 *a-d*), the sensitive wavebands of hyperspectral reflectance to leaf pigment concentrations always occurred near 550nm, 670nm, 700nm and 1450-1460nm in the green, red, NIR and SWIR regions, respectively.
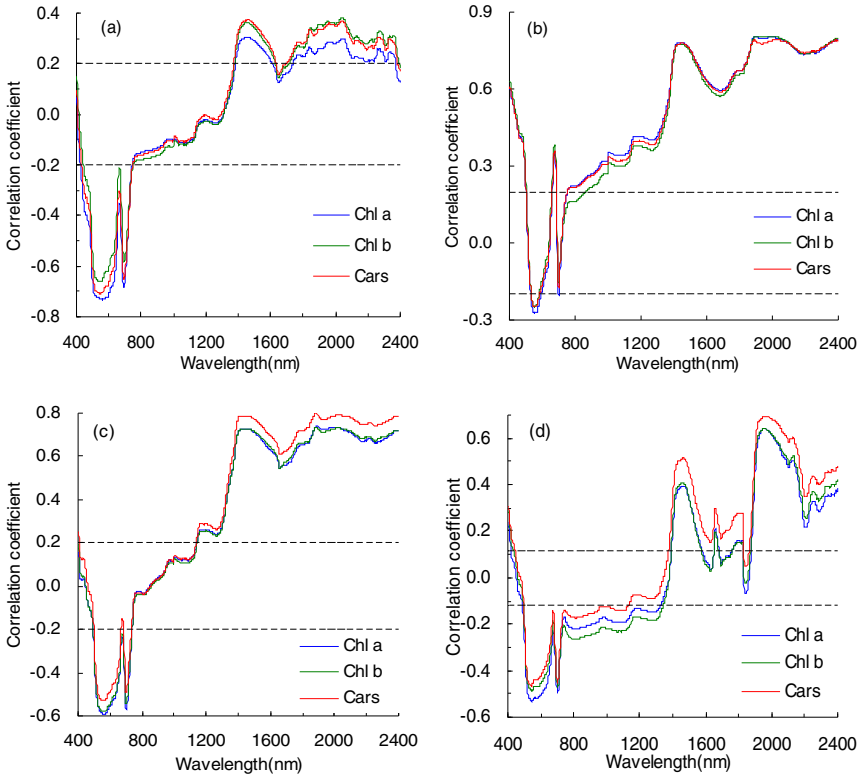
Fig. 2. Correlograms of hyperspectral reflectance to leaf pigment concentration collected on August 21st (a), September 27th (b) and October 10th (c), 2008, and all the three sampling dates (d). The dashed line denoted obviously significant level at p<0.001.

3.3   Regression between Pigment Concentration and Integrated Hyperspectral Index

As analyzed above, the sensitive wavebands were consistent with the applied wavebands of integrated hyperspectral index (MCARI/OSAVI$_{[670,800]}$). MCARI/ OSAVI$_{[670,800]}$ had better linear relationship with leaf pigment concentrations (Fig.3).

As shown in Fig.3 (*a*) to (*l*), the correlation between MCARI/OSAVI$_{[670,800]}$ and leaf pigment concentrations were obviously significant (*p*<0.05) in any leaf positions and total samples. Strong correlation existed for all the data collected in the three sampling dates (Fig.3 *m-p*). MCARI/OSAVI$_{[670,800]}$ were most appropriate for Chl *a* estimation with high determination coefficient (R^2) of 0.749 for L1 collected on August 21st, 2008 (Fig.3 *a*). It was worst appropriate for all the data collected in the three sampling dates without regard to leaf positions (Fig.3 *p*). The determination coefficients of total samples for any sampling dates (Fig.3 *d*, *h*, *l* and *p*) were always lower than those of each leaf position (Fig.3 *a-c*, *e-g*, *i-k* and *m-o*).
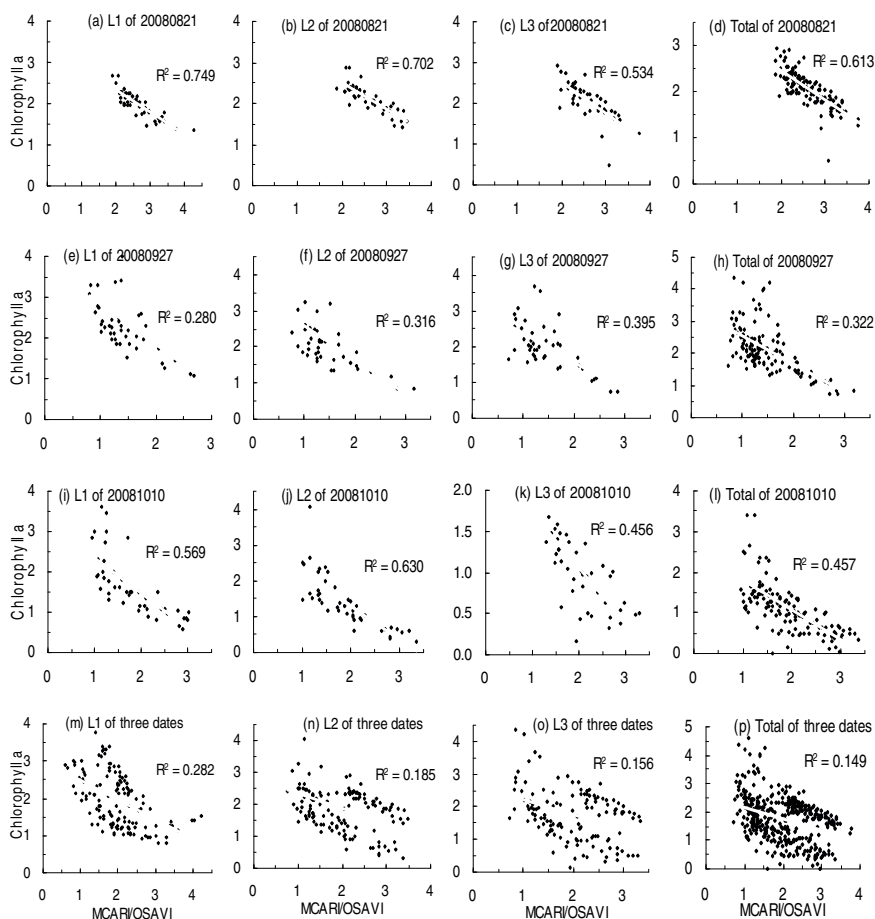
Fig. 3. Linear relationship between foliar chlorophyll a concentration of rice crop and the integrated index (MCARI/OSAVI$_{[670,800]}$)

4   Discussion and Conclusion

In the present study, the hyperspectral reflectance were found that there was no significant difference (n=45 in each group) in the visible blue (450-515nm), green (525-605nm), red (630-690nm), near-infrared (NIR, 775-900nm) spectral regions for difference leaf positions (Fig.1, Talbe1) because the difference in concentrations of photosynthetic pigments Chl *a*, Chl *b* and Cars, which absorb the visible spectrum by electronic transitions among L1, L2 and L3 was almost not significant (Table 2). In SWIR regions, however, the difference was obviously significant (Duncan test at alpha = 0.05) by twos of all the three leaf positions (Table 1). The leaf hyperspectral reflectance in SWIR regions decresed gradatially from the top to below in the two out of three sampling dates (Fig.1, Table 1). Such results were similar to others reports by Zhao et al.[3] who concluded that the hyperspectral reflectance of maize would change in different spectral regions due to N treatments.

The correlation analysis of hyperspectral reflectance indicated that the most sensitive wavebands mainly located at near 550nm, 670nm, 700nm, 1460nm, and 1650nm. These wavebands always presented the greatest correlation coefficient with significant difference ($p<0.05$) even very obviously significant difference ($p<0.001$). The result in this study was incompletely with the current research which suggested that the most sensitive wavebands distributed in the green (525-605nm) and yellow (605-650nm) regions, and the worst in SWIR region[26]. The waveband at 1460nm was efficient to estimate the absolute and relative concentration of photosynthetic pigments and other biophysical or biochemical parameters at leaf level in the laboratory condition, but it's inefficient at the canopy level from the near ground, airborne and spaceborne platforms because of the presence of water vapor[12,27].

The integrated hyperspectral index MCARI/OSAVI$_{[670,800]}$ had the better linearity with leaf pigment concentrations(Fig.3). Perhaps because the component wavebands were derived from the aforementioned sensitive wavebands (Equ. (3)). As shown in Fig.3 (*a*) to (*l*), The regression analysis denoted that MCARI/OSAVI $_{[670,800]}$ had a great relationship with leaf pigment concentrations in any leaf positions and total samples. To take into account the leaf positions, MCARI/OSAVI$_{[670,800]}$ was a reliable hyperspectral index for estimating the pigment concentration.

However, there are more external factors including canopy structure, LAI, plant cover density, soil types, soil moisture, sunlight illumination, cloudy and plant shadow, and so forth, which coexist in the field[23], and bring lots of noise in remote sensing application. Therefore, it will be necessary to make multiangular measurement to acquire more detailed information on plant structure than vertical measurement, and then the hyperspectral VIs will obtain high accuracy for estimating the concentration of photosynthetic pigments and other agricultural parameters.

References

1. Ahl, D.E., Gower, S.T., Mackay, D.S., Burrows, S.N., Norman, J.M., Diak, G.R.: Heterogeneity of light use efficiency in a northern Wisconsin forest: implications for modeling net primary production with remote sensing. Remot. Sens. Environ. 93(1-2), 168–178 (2004)
2. Zeng, H.Q., Liu, Q.J., Feng, Z.W., Wang, X.K., Ma, Z.Q.: Modeling the interannual variation and response to climate change scenarios in gross and net primary productivity of Pinus elliottii forest in subtropical China. Act. Ecol. Sini. 28(11), 5314–5321 (2008)
3. Zhao, M.S., Heinsch, F.A., Nemani, R.R., Running, S.W.: Improvements of the MODIS terrestrial gross and net primary production global data set. Remot. Sens. Environ. 95(2), 164–176 (2005)
4. Jia, X.X., Shao, M.A., Wei, X.R., Horton, R., Li, X.Z.: Estimating total net primary productivity of managed grasslands by a state-space modeling approach in a small catchment on the Loess Plateau, China. Geoderma 160(3-4), 281–291 (2011)
5. Blackburn, G.A.: Spectral indices for estimating photosynthetic pigment concentrations: a test using senescent tree leaves. Int. J. Remot Sens. 19(4), 657–675 (1998a)
6. Steele, B.M., Reddy, S.K., Nemani, R.R.: A regression strategy for analyzing environmental data generated by spatio-temporal processes. Ecol. Model. 181(2-3), 93–108 (2005)
7. Fownes, J.H., Aber, J.D.: Forest canopy chemistry from Blackhawk Island, Wisconsin. In: Proceedings of the Airborne Imaging Spectrometer Data Analysis Workshop, pp. 100–105. NASA-Jet Propulsion Laboratory Publication, Pasadena (1985)
8. Gholz, H.L.: Environmental limits on aboveground net primary production, leaf area, and biomass in vegetation zones of the Pacific Northwest. Ecol. 63(2), 469–481 (1982)
9. Ingestad, T.: Nitrogen stress in birch seedlings. 2. N, K, P, Ca and Mg nutrition. Plant Physiol. 45(1), 149–157 (1979)
10. Zhou, Q.F., Liu, Z.Y., Huang, J.F.: Detection of nitrogen-overfertilized rice plants with leaf positional difference in hyperspectral vegetation index. J. Zhejiang Univ.-Sc. B. 11(6), 465–470 (2010)
11. Yoder, B.J., Pettigrew-Crosby, R.E.: Predicting nitrogen and chlorophyll content and concentrations from reflectance spectra (400-2500 nm) at leaf and canopy scales. Remot. Sens. Environ. 53(3), 199–211 (1995)
12. Daughtry, C.S.T., Walthall, C.L., Kim, M.S., de Colstoun, E.B., McMurtrey III, J.E.: Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. Remot. Sens. Environ. 74(2), 229–239 (2000)
13. Blackmer, T., Schepers, J.S., Varve, l.G.E., Walter-Shea, E.A.: Nitrogen deficiency detection using reflected shortwave radiation from irrigated corn canopies. Agro. J. 88, 1–5 (1996)
14. Walburg, G., Bauer, M.E., Daughtry, C.S.T., Housley, T.L.: Effects of nitrogen nutrition on the growth, yield, and reflectance characteristics of corn canopies. Agro. J. 74, 677–683 (1982)
15. Hinzman, L.D., Bauer, M.E., Daughtry, C.S.T.: Effects of nitrogen fertilization on growth and reflectance characteristics of winter wheat. Remot. Sens. Environ. 19(1), 47–61 (1986)
16. Demarez, V., Gastellu-Etchegorry, J.P.: A modeling approach for studying forest chlorophyll content. Remot. Sens. Environ. 71(2), 226–238 (2000)
17. Peñuelas, J., Gamon, J.A., Fredeen, A.L., Merino, J., Field, C.B.: Reflectance indices associated with physiological changes in nitrogen- and water-limited sun flower leaves. Remot. Sens. Environ. 48(2), 135–146 (1994)

18. Zarco-Tejada, P.J., Miller, J.R., Mohammed, G.H., Noland, T.L., Sampson, P.H.: Optical indices as bioindicators of forest condition from hyperspectral CASI data. In: Proceedings of the 19th Symposium of the European Association of Remote Sensing Laboratories (EARSeL), Valladolid, Spain (1999)
19. Wu, C.Y., Niu, Z., Tang, Q., Huang, W.J.: Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. Agr. Forest Meteorol. 148(8-9), 1230–1241 (2008)
20. Rondeaux, G., Steven, M., Baret, F.: Optimization of soil-adjusted vegetation indices. Remot. Sens. Environ. 55(2), 95–107 (1996)
21. Tang, Y.L., Wang, R.C., Huang, J.F.: Relations between red edge characteristics and agronomic parameters of crops. Pedosphere 14(4), 467–474 (2004)
22. Kobayashi, T., Kanda, E., Kitada, K., Ishiguro, K., Torigoe, Y.: Detection of rice panicle blast with multispectral radiometer and the potential of using airborne multispectral scanners. Phytopathology 91(3), 316–323 (2001)
23. Liu, Z.Y., Wu, H.F., Huang, J.F.: Application of neural networks to discriminate fungal infection levels of rice panicle using reflectance measurements and principal components analysis. Comput. Electron. Agr. 72(2), 99–106 (2010)
24. Daughtry, C.S.T., Bauer, M.E., Crecelius, D.W., Hixson, M.M.: Effects of management practices on reflectance of spring wheat canopies. Agro. J. 72(6), 1055–1060 (1980)
25. Kim, M.S., Daughtry, C.S.T., Chappelle, E.W.: The use of high spectral resolution bands for estimating absorbed photosynthetically active radiation (Apar). In: Proceedings of the Sixth Symposium on Physical Measurements and Signatures in Remote Sensing, Val D'Isere, France, pp. 299–306 (1994)
26. Zhao, D., Reddy, K.R., Kakani, V.G., Read, J.J., Carter, G.A.: Corn (Zea mays L.) growth, leaf pigment concentration, photosynthesis and leaf hyperspectral reflectance properties as affected by nitrogen supply. Plant Soil 257(1), 205–217 (2003)
27. Blackburn, G.A.: Quantifying chlorophylls and caroteniods at leaf and canopy scales: an evaluation of some hyperspectral approaches. Remot. Sens. Environ. 66, 273–285 (1998b)

Identifying Leaf-Scale Wheat Aphids Using the Near-Ground Hyperspectral Pushbroom Imaging Spectrometer

Jinling Zhao[1,2], Dongyan Zhang[1,3], Juhua Luo[1], Dacheng Wang[1], and Wenjiang Huang[1,*]

[1] Beijing Research Center for Information Technology in Agriculture, Beijing Academy of Agriculture and Forestry Sciences, Beijing, 100097, P.R. China
`yellowstar0618@163.com`
[2] Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, 100101, P.R. China
[3] Institute of Agricultural Remote Sensing & Information Technology Application, Zhejiang University, 310029, Hangzhou, P.R. China

Abstract. This study is to identify leaf-scale wheat aphids using the near-ground hyperspectral Pushbroom Imaging Spectrometer (PIS). Firstly, the spectral characteristics between normal and aphid-infested wheat leaves were compared in spectral reflectance. Concerning the serious aphid damage level, it is obvious that its spectral curve is badly flattened such as green peak (centered around 550 nm), red valley (centered around 680 nm), due to the influence of aphid. Specifically, in the visible spectrum (500-701 nm), the maximum delta (the maximum value minus the minimum value) is 3.3 and it is 7.5 in the near-infrared spectrum (701-900 nm). Then, the spectral difference and change rate were further analyzed. It seems that both curves show the mirror symmetry and their maximum values are 55.8% and 17.4%, respectively. For the difference curve, the value is negative in the visible spectrum (400-700 nm), which shows that the reflectance of normal wheat leaf is less than that of the serious level. Conversely, it is greater in the near-infrared spectrum (700-900 nm). Finally, based on the high spatial resolution PIS image, ENvironment for Visualizing Images (ENVI-EX) was utilized to extract aphids and the overall accuracy reaches 97%. The result indicates that the PIS is sufficient to identify the wheat aphids and this study can lay a foundation for further applications in precision agriculture using such a hyperspectral imaging system.

Keywords: Hyperspectral remote sensing, leaf-scale, pushbroom imaging spectrometer (PIS), spectral characteristics, wheat aphids.

1   Introduction

Wheat (Triticum aestivum L.) is one of the most important cereal crops and it has been used as major consumable commodity by human beings in most areas of the

---

world. However, global climate change has exerted a negative impact on wheat production [1]. As a result, the incidence of various kinds of stress is becoming more often, among which diseases and insects have strongly affected the wheat crop. As one of the most important insects in winter wheat, aphids (Rhopalosiphum padi L.) have caused significant losses in its quantity and quality [2]. In addition, insects and diseases in wheat can spread at a fast pace and serious hazards can be induced if they are not timely detected and prevented. Therefore, some effective measures must be taken to reduce production losses and ensure food safety. Nevertheless, on-farm pest management and crop protection strongly depend on diagnosis of disease or insect stresses in fields, while traditional field survey methods are labor-intensive and time-consuming, and they cannot yield full coverage [3]. Accordingly, it is of great importance to accurately monitor and evaluate the wheat aphids in a noninvasive and nondestructive way. On the other hand, it is also essential for ensuring that pesticide spraying can be limited and more time- and site-specifically [4].

With the fast development of advanced sensors and image processing techniques, quantitative remote sensing has facilitated extraordinary advances in monitoring and identifying wheat insects [5]. When it is infested, the reflectance spectra are obviously different between healthy and infested plant. To detect such a stress using the remote sensing, it is based on the assumption that insect stress interferes with photosynthesis or physical structure of plant. Furthermore, such destruction affects the absorption of light energy, and thus alters the reflectance spectra at different damage levels [6]. In previous studies, multi-spectral remotely sensed data were usually utilized to identify and monitor the crop stresses in several bands at a large scale in the visible and near-infrared regions [7-9]. However, the crude spectral resolution of the reflected and emitted energy from the earth is the primary limiting factor in differentiating subtle differences of plant stresses [10]. Consequently, hyperspectral remote sensing is currently being investigated by scientists with regard to the detection and identification of plant health/stress status [11]. Although airborne and spaceborne hyperspectral sensors can also acquire simultaneously imaging and spectrum in a single scanning process, lack of near-ground prior knowledge of stressed plants has greatly limited such data in the wide applications of precision agriculture. Conversely, near-ground hyperspectral imaging devices can derive the refined spectral characteristics of stressed plants in fields, which can lay the methodological foundation for those hyperspectral data at large spatial scales and further provide the ground truth data for accuracy validation.

By comparing with commonly used hyperspectral spectrometers such as the ASD (Analytical Spectral Devices, Boulder, CO, USA), near-ground hyperspectral imaging device can acquire simultaneously imaging and spectroscopy in a single system, which can be an effective tool in describing the variability between healthy and infected plants. In the last several years, more attention has only been paid to such an imaging device. Several studies have shown that the near-ground hyperspectral imaging systems can detect insect- and disease-infested plants in a non-invasive and non-destructive way [12-14]. However, corresponding applications and processing techniques have not yet been enough explored in detecting wheat insects or diseases [15-16]. Therefore, based on the Pushbroom Imaging Spectrometer (PIS), the primary

objective of this study lies in characterizing the spectral characteristics of leaf-scale wheat aphids and identifying the aphids using the high spatial resolution image.

2   Materials and Method

2.1   Near-Ground Hyperspectral Imaging Device

The hyperspectral imaging device used in our study is the PIS, which was jointly developed by Beijing Research Center for Information Technology in Agriculture and University of Science and Technology of China. This system acquires images by linear array push-broom imaging and Fig. 1 is the demonstration of collecting hyperspectral imaging data. Table 1 shows some key specification parameters of the PIS. It is obvious that PIS is superior to the commonly used ASD field spectrometer in spectral resolution and sampling interval. Additionally, the PIS can also acquire spectro-images with a very high spatial and spectral resolution by which land features of interest can be identified.

Table 1. Some key specification parameters of the PIS imaging system

Sensor parameters	PIS
Spectrum range	400~1000 nm
Spectral resolution	2 nm
Sampling interval	0.7 nm
FOV	16°
Spatial resolution	5-10 mm
Pixel dimension	7.4 μm×7.4 μm
Image resolution	1400 (Spatial dimension) × 1024 (Spectral dimension)

2.2   Data Acquisition and Preprocessing
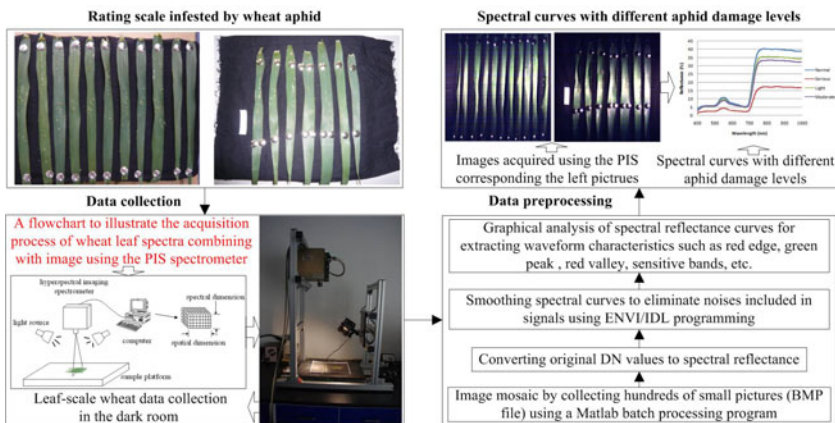
Fig. 1. A demonstration of acquiring the spectra combing with image of different aphid damage levels using the PIS

An experiment was carried out at the experimental farm of Beijing Academy of Agriculture and Forestry Sciences (39.93° N, 116.27° E) on 26 May 2010. At that time, it was just the grain-filling stage of winter wheat, which is the key yield-forming period and is also extremely sensitive to various kinds of biological disasters. In this study, wheat aphids were used as the study object. When aphid-infested symptoms of wheat leaves were evaluated, relative insect damage levels were assessed according to the aphid populations under the help of experienced pathologist.

Before using the device to measure wheat leaves, strict laboratory calibration must be firstly performed for ensuring a high radiometric accuracy. When collecting the reflectance spectra, the PIS lens was set at 80 cm over the wheat leaves and halogen lamp irradiation was fixed at a 45° angle. Then, the inverse 2nd leaves were picked up and immediately put onto the scanning platform, and the PIS moved evenly along the track and the image and spectra of wheat leaves on the black cloth were acquired (Fig. 1). Before and after collecting the leaf spectra, white reference panel was used to optimize the instrument. Finally, further processing were required to covert the original DN (Digital Number) values to reflectance by the empirical linear method (Eq. 1). Afterwards, a moving-average smoothing algorithm was utilized to exclude abnormal values and smooth the spectral curves in order to accurately derive the reflectance characteristics between healthy and aphid-infested wheat leaves.

$$\rho = a * DN + b \ . \tag{1}$$

Where ρ is the real reflectance, a and b are the coefficients. When putting the measured spectral value and corresponding DN into Eq. 1, a and b can be obtained by Least-Square Method (LSM), and then ρ can be obtained.

3   Results and Discussion

3.1   Analysis of Spectral Characteristics of Leaf-Scale Wheat Aphids

On the high-resolution PIS image (Fig. 2(a)), two pixels including a healthy leaf point and an aphid-covered leaf point were selected. Fig. 2(b) is the comparative result of their spectral reflectance curves in the 500-900 nm spectral range. It is obvious that the spectral reflectance values are markedly different between normal and aphid leaves. Furthermore, between 500 nm and 701 nm, the reflectance of aphid leaf is larger than that of normal leaf, while it is smaller in the 701-900 nm spectral regions. In the visible spectrum (500-701 nm), the maximum delta (the maximum value minus the minimum value) is 3.3 and it is 7.5 in the near-infrared spectrum (701-900 nm). Concerning the locations of the minimum value, they are very similar between normal and aphid pixels, which are 664 nm and 661 nm, respectively. However, they are very different for the locations of maximum values which are 785 nm and 852 nm, respectively.

Comparing the spectral curves between normal and aphid-infested leaves, it can be found that both of them show the typical spectral characteristics of green vegetation: green peak centered around 550 nm, red valley centered 680 nm and high near-infrared reflectance centered around 780 nm, but they are different at specific spectral values for different damage levels. Furthermore, comparing the reflectance between

the visible and near-infrared spectra, the spectral differences are more obvious in the near-infrared spectrum. Nevertheless, the spectral curve of aphid-affected leaf is badly flattened especially for the characteristic bands such as green peak, red valley owning to the influence of aphid.
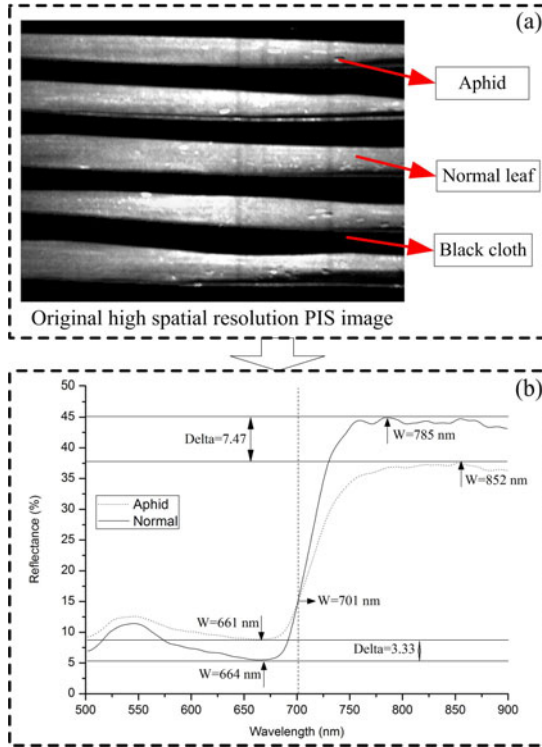
Fig. 2. The comparative of spectral curves (b) between normal and aphid-infested leaves on the PIS image (a)

Additionally, the analysis of spectral difference and change rate were also performed between normal and seriously aphid-affected leaves (Fig. 3). Due to the constraint of spectral range, the PIS can only obtain the spectral characteristics in the visible and partial near-infrared wavelength. In the 400-900 nm spectral regions, it can be found that they are very different in the spectral difference and change rate curves. It seems that they show the mirror symmetry and the maximum change rate and difference are 55.8% and 17.4%, respectively. As can be seen in the difference curve, the value is negative in the visible spectrum (400-700 nm), which shows that the reflectance of the normal wheat leaf is less than that of the serious level. Conversely, it is greater in the near-infrared spectrum. This phenomenon shows that the reflectance is determined by different factors in the visible and near-infrared spectral ranges. For the change rate curve, it fluctuates more severely in the visible spectrum than in the near-infrared spectrum, which shows that aphids exerted more evident impact in the visible spectrum.
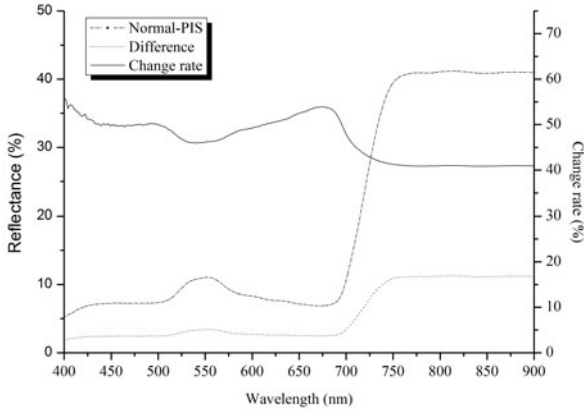
Difference=f$_{Normal}$-f$_{Serious}$; Chang rate=Difference/f$_{Normal}$; f is the spectral reflectance.

Fig. 3. Analysis of spectral difference and change rate between healthy and seriously aphid-affected wheat leaves

The phenomenon for the change trend of spectral curves is that aphids cover the top of wheat leaf and they destroy the cell structure and change the chlorophyll content. In the visible spectral range, the reflectance of wheat leaf is mainly determined by various pigments, among which, chlorophyll is a leading one. When aphids break the wheat leaf, chlorophyll content will accordingly decrease which result in an increase of reflectance owning to the decreasing ability of chlorophyll absorption. Additionally, the secretions of aphids adhere to the leaf top can absorb some dust and other materials and increase the reflectance to a certain degree. Conversely, the reflectance depends on the cell structure in the near-infrared spectrum. Aphids destroy the cell structure of wheat leaf, so the reflectance of aphid-infected leaf decreases compared with the normal leaf.

3.2   Identifying Aphids Using the PIS Image

Based on the high spatial resolution PIS image, aphids were identified. Due to the only leaf-scale analysis in this study, it is easy to separate aphids from leaf background. Here, the ENVI-EX (ENvironment for Visualizing Images, Research Systems, Inc.) feature extraction module was used owning to the obvious aphid texture and single background. As shown in Fig. 4(a), there are five necessary steps to complete the object-oriented classification including choosing the scale parameter, merging the object primitives, refining the objects using a threshold value for just one band of the image and it is an optional step, extraction of the attributes, and object-oriented classification based on rules or examples. As can be seen on the aphid identified image (Fig. 4(b)), it is obvious that most aphids were extracted from normal

wheat leaves. To evaluate the classification accuracy, 30 aphid pixels were randomly selected in the ENVI environment, and then they were found on the classification map using the link displays tool o ENVI. The comparative result showed that 29 points were correctly identified and the overall accuracy reached 97%.
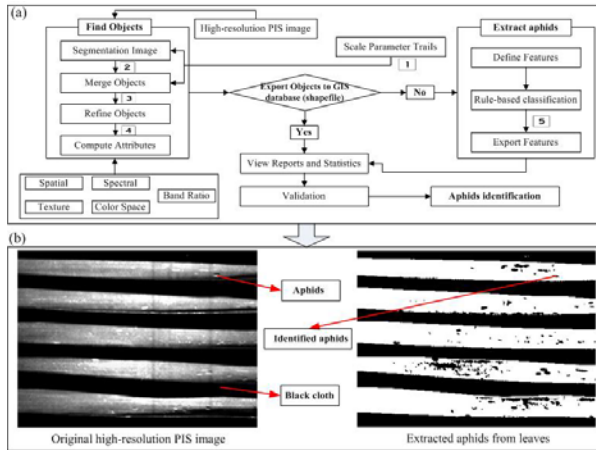
Fig. 4. The technical flowchart for identifying leaf-scale aphids using the ENVI-EX object-oriented classification module (a) and (b) is the high-resolution image acquired by the PIS spectrometer and the identification result

4   Conclusion

In our study, near-ground hyperspectral imaging data (PIS) were used to identify leaf-scale wheat aphids. By analyzing the spectral characteristics between normal and aphid-affected wheat leaf and classifying aphids using the PIS image, it has been proved to be an effective tool for identifying aphids using such a hyperspectral imaging system. According to the above analysis, we drew several conclusions: (1) the PIS can combine imaging with spectrum in a single system compared with commonly used non-imaging ASD spectrometer; (2) the PIS can collect the pixel-to-pixel spectrum, so it can obtain the pure spectral information on the basis of identification of targets of interest; (3) when the PIS is used in the canopy-scale wheat aphids, it will be more accurate to identify the targets of interest by using simultaneously spectral and spatial information of the PIS; (4) owning to very high spatial resolution of the PIS image, it will be very suitable for identifying the targets of interest using the object-oriented classification method.

References

1. Ortiz, R., Sayre, K.D., Govaerts, B., Gupta, R., Subbarao, G.V., Ban, T., Hodson, D., Dixon, J.M., Ortiz-Monasterio, J.I., Reynolds, M.: Climate Change: Can Wheat Beat the Heat? Agr. Ecosyst. Environ. 126, 45–58 (2008)
2. Burd, J.D., Porter, D.R., Puterka, G.J., Haley, S.D., Peairs, F.B.: Biotypic Variation among North American Russian Wheat Aphid (Homoptera: Aphidae) Populations. J. Econ. Entomol. 99, 1862–1866 (2006)
3. Lucas, J.A.: Advances in Plant Disease and Pest Management. J. Agr. Sci. 149, 91–114 (2011)
4. Delalieux, S., Auwerkerken, A., Verstraeten, W.W., Somers, B., Valcke, R., Lhermitte, S., Keulemans, J., Coppin, P.: Hyperspectral Reflectance and Fluorescence Imaging to Detect Scab Induced Stress in Apple Leaves. Remote Sens. 1, 858–874 (2009)
5. Yang, Z., Rao, M.N., Elliott, N.C., Kindler, S.D., Popham, T.W.: Differentiating Stress Induced by Greenbugs and Russian Wheat Aphids in Wheat using Remote Sensing. Comput. Electron. Agr. 67, 64–70 (2009)
6. Moran, M.S., Inoue, Y., Barnes, E.: Opportunities and Limitations for Image-Based Remote Sensing in Precision Crop Management. Remote Sens. Environ. 61, 319–346 (1997)
7. Huang, W.J., Lamb, D.W., Niu, Z., Zhang, Y.J., Liu, L.Y., Wang, J.H.: Identification of Yellow Rust in Wheat using in-situ Spectral Reflectance Measurements and Airborne Hyperspectral Imaging. Precis. Agric. 8, 187–197 (2007)
8. Qin, Z.H., Zhang, M.H.: Detection of Rice Sheath Blight for In-Season Disease Management using Multispectral Remote Sensing. Int. J. Appl. Earth Obs. 7, 115–128 (2005)
9. Lan, Y., Huang, Y., Martin, D.E., Hoffmann, W.C.: Development of an Airborne Remote Sensing System for Crop Pest Management: System Integration and Verification. Appl. Eng. Agric. 25, 607–615 (2009)
10. Govender, M., Chetty, K., Bulcock, H.: A Review of Hyperspectral Remote Sensing and its Application in Vegetation and Water Resource Studies. Water SA 33, 145–152 (2006)
11. Yang, C.M.: Assessment of the Severity of Bacterial Leaf Blight in Rice using Canopy Hyperspectral Reflectance. Precis. Agric. 11, 61–81 (2010)
12. Nillson, H.E.: Remote Sensing and Image Analysis in Plant Pathology. Annu. Rev. Phytopathol. 15, 489–527 (1995)
13. Mirik, M., Michels Jr., G.J., Kassymzhanova-Mirik, S., Elliott, N.C., Catana, V., Jones, D.B., Bowling, R.: Using Digital Image Analysis and Spectral Reflectance Data to Quantify Damage by Greenbug (Hemitera: Aphididae) in Winter Wheat. Comput. Electron. Agr. 51, 86–98 (2006)
14. Singh, C.B., Jayas, D.S., Paliwal, J., White, N.D.G.: Identification of Insect-Damaged Wheat Kernels using Short-Wave Near-Infrared Hyperspectral and Digital Colour Imaging. Comput. Electron. Agr. 73, 118–125 (2010)
15. Ye, X.J., Sakai, K., Okamoto, H., Garciano, L.O.: A Ground-Based Hyperspectral Imaging System for Characterizing Vegetation Spectral Features. Comput. Electron. Agr. 63, 13–21 (2008)
16. Inoue, Y., Penuelas, J.: An AOTF-Based Hyperspectral Imaging System for Field Use in Ecophysiological and Agricultural Applications. Int. J. Remote Sens. 22, 3883–3888 (2001)

Research on Application of Web Log Analysis Method in Agriculture Website Improvement

Jian Wang

Agricultural information institute of CAAS, Beijing 100081, China
`wangjian@caas.net.cn`

Abstract. With the advance of agricultural modernization, agriculture website was increasingly becoming a major tool for farmers getting information about life and production. How to make the analysis of the needs of farmers effectively to help them to find the information from the ocean of information and resources of the Internet they were interested in had become an urgent and important issue. In this paper, we used the website of Agridata as an example and focused on the solution for the problem. A way was proposed for analyzing and mining the web log of Agridata, which integrated statistical analysis and cluster analysis. By this method, could information behaviors of users be grasped when browsing the website. It was important significance in improving the structure and content of agriculture website, which could provide better services for farmers and improve the level of modernization of agricultural production.

Keywords: Web log analysis, information behavior, agricultural users, agricultural website.

1 Introduction

As the emphasis on the rural information continues to deepen in our country, the development environment for rural information was constantly optimized and farmers' awareness of information technology was growing. More and more famers knew the technology for getting information of production and living from Internet, which could fundamentally change the situation of information isolated from the outside world to remote villages. Therefore, agriculture web site provided a great convenience to farmers in eliminating poverty, agricultural sales, employment, medical treatment and education. Thus, agriculture web site played a positive role for improving living conditions in rural areas and safeguarding national unity and stability. However, during the process of building agriculture web site, as the special nature of agriculture, farmers' behavior of access information had some special needs. How to grasp the information needs of agricultural users was important for improving agriculture web site. This was an urgent problem.

Web log was a kind of file which record some information of user access to web pages which included records of Browse, search, information downloads, and other message of site visits. We could obtain varies information about visits and utilization of

website resource and properties of users by calculating and analyzing web server log. This information could help us optimize the information resources and improve quality and efficiency of web services. At present, many commercial Web sites had analyzed user behavior through the log analysis techniques to improve web services and this area had become a research focus. For example, Zhang xuehong(2005) discussed the actual value of user behaviors analysis, home page designing and customer service with the method of log analysis and introduced one process of homepage log analysis. Feng-chunhui(2010) studied the web log applied in distance education using data mining technology and obtained a good result. Furthermore, Xuping (2010) and Weng-changping(2010) also studied web log playing the role of mining users information behaviors to improve website with their background.

However, the study background of user information behavior analysis based on web log which had referenced from existing literatures were for academic professionals such as library users and distance education students. This method was very little used for analyzing the information behaviors of web users in rural areas. To resolve this problem, in this passage, the Web access log of the agricultural scientific data center was used as an object and we discussed the method for studying the information behavior of users of agricultural methods with statistical analysis and cluster analysis. Then we conducted experiments with the Web access log of the national agricultural scientific data center and analyzed the pattern of information behavior of its user. With these results, several advices on this website of the agricultural scientific data center were made for providing a better service to agricultural users.

2   Materials and Methods

2.1   Web Log of the Agricultural Scientific Data Center Collection

The agricultural scientific data center (Agridata) is one of the experimental data centers supported by National Facilities and Information (NFII) of Ministry of Science and Technology. Based on agricultural scientific data sharing Standard, Agridata Integrated 12 types of agricultural scientific data which included crop science, animal science & veterinary medicine, agricultural resource and environment, grassland science, food science and standards, etc. these data resources could support the agricultural technology innovation and management decision greatly. Up to 2009, Agridata Formed a stable user group including more than 150 group users and 8,000 individual registered users. The total visit count of Agridata was 1.8 million. It showed that Agridata was an important information source of network users in rural areas.

The web log of Agridata was some log files about website visits existing in the web server which included IP address of website visitors, URL of explored pages, date and time of visits, path of exploring, etc. Usually, as different setting of web server, the format of log files had three types: National Center for Supercomputing Applications Combined (NCSA), Microsoft IIS Format and W3C Extended Log File Format. The log file format of Agridata was W3C Extended Log File Format. The format and meaning of the fields in the log file was shown in table 1.

Table 1. The format and meaning of the fields

ID	Index	W3C Extended Log File Format	Meaning of fields
1	Host IP	c-ip	IP address or domain name of the client computer which send HTTP request to sever.
2	User name	cs-username cs-host	user ID and host of users who send HTTP request to sever
3	Date and time	Date、Time	Date and time of HTTP request to sever send by user
4	HTTP Request	cs-method、 cs-uri-stem cs-uri-query	HTTP request to sever send by user
5	HTTP service status	sc-status sc-substatus sc-win32-status	Service status codes of HTTP request which indicated success or failure
6	The number of bytes	sc-bytes、 cs-bytes Time-taken	The number of bytes and time taken of processing HTTP request and web server receiving and sending
7	Backlinks	cs(cookie) cs(Referer)	URL and cookie visited by user before the last HTTP request send
8	User Agent	cs(User-Agent)	browser and operating system software information user used to visit web server
9	Protocol Version	cs-version	format version of log file
10	Server circumstances	s-sitename s-computername s-ip s-port	information of website sever included name of computer, name of website, IP address and port number

2.2 Web Log Data Processing

Web log file gotten from website sever directly contained very rich information. Only a part of this information was needed to analyze information behavior of website users. How to extract this useful information from the log file correctly and completely was the key for analysis of user information behavior. This process was called web log data processing.

As usual, there were four steps in web log data processing: data cleaning, user identification, session identification and access fragment identification. The flow chart of web log data processing was shown as figure 1.
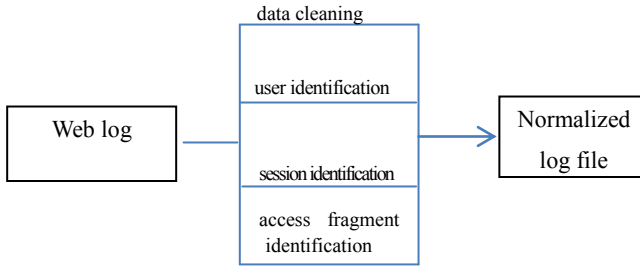
Fig. 1. The flow chart of web log data processing

2.2.1 Data Cleaning

The purpose of data cleaning was deleting the information unrelated to analysis the user information behaviors from web log files, combining similar records, summarizing and dealing with records of error accessing. As usual, HTML files in web log were associated with the user session. On the other hand, during the process of visiting page, images, video and audio files in the page were also downloaded and visit recorders of these files were written in the log. These information were useless for analysis the user information behaviors and should be deleted with the method of URL suffix filter. But it was worth noting that the method of URL suffix filter was used with understanding website content to avoid the loss of some important user sessions. For example, for a page which main content was image, we should not simply delete records of images access from log file when using the method of URL suffix filter for data cleaning. Meanwhile, we should define a set of rules to retain this information as a basis for analysis. On the other hand, the operating records of network administrators, web spider and web crawler could be filter out during the process of data cleaning.

2.2.2 User Identification

The purpose of user identification was analyzing information of users. As the local cache, proxy servers and firewalls were used in web, such as different users using the same proxy servers, the computer with the same IP address at the same time having different systems and browsers, etc. it was difficult for identifying users with IP address. Usually, during the process of research, we could set some rules to simplify the classification process: when the same IP address used different systems and browsers at the same time, users with the IP address could be considered as different users; when the same IP address had visited a group of pages which were not topological link, users with the IP address could be considered as different users and so on. On the other hand, as people in research institutions were relatively concentrated geographically while farmers were more scattered, we could set a rule to infer these two kinds of people: the grope with high concentration IP address were researchers, otherwise the group with low concentration IP address were farmers.

2.2.3   Session Identification

A session was a valid user to access the service. Details on session identification were processes of identifying users' requests for consecutive pages to get users' message of information behaviors and interesting objection. The purpose of session identification was creating page clustering for each user's information behaviors, which could provide "material" for analysis information behaviors of user. Usually, for a log file of long series access time, the same user may repeatedly visit the same page. Thus, we should set a rule: when time intervals of request for any two adjacent pages were over a set threshold, we could regard this process as restarting a session.

2.2.4   Access Fragment Identification

The nature of access fragment identification was supplementing the above process of session identification. It could mean that this process determined whether any important request in file log could be identified. In this process, the method we applied was similar to the way of user identification which set a rule: when pages between current request and last request were not hypertext links, this page the user current request was called from the local cache. In this case, we should check the reference log to determine the source of the current user request. When several pages in visit history contained links to current requested page, we used the page that visit time was closest to the current one as a current source. If some web log was missing, we could add the missing page to the user's session file by the method of analysis of site topology.

2.3   Log Data Mining and User Information Behavior Analysis

After data of web log processing, a large number of redundant information was filtered out and the useful information could be mined to get the information behavior of network users. Usually, there were two steps in this process: user behavior information mining and user information behavior analysis.

2.3.1 User Behavior Information Mining

User information behavior is a concept of diversity, on one hand, only when information was useful to users, had the information existence value; on the other hand, users could not do anything without information. It was meaning that information and users relied on each other. Thus, we can determine that during the process of users' information behaviors, users could absorb certain information and create information related to it. Information was given new life during the process. In general, the user's information behavior was determined by a variety of characteristics of the user, such as the user's knowledge structure, behavior, etc. In this paper, we focused on the need for Improve the Agridata website efficiency with the method of analysis agricultural network user's information behavior. So characteristics we chosen from the web log data mining in this study were as follows: click-through of a page, users' traffic, the view number of pages, traveling time of users, access duration, network traffic, the visit number of a single page, etc.

2.3.2 User Information Behavior Analysis

According to the characteristics of users' information behavior, we could analyze the information behavior of a user to grasp the user's psychology, which was a great significant for improving services effect of website. As usual, this analysis included two methods: statistical analysis and classification clustering.

(1) Statistical analysis

This was a most commonly method for analyzing user access patterns, which studied how to measure, observe, express and summary the number characteristics of objection. In this passage, we could compute parameters such as mean, maximum, minimum, etc, with a variety of information of users who visited Agridata website to estimate the behavior information pattern of users. To some characteristics that were not easy to see from data, we used graphics-assisted method for analysis, such as Scatter plots, histograms, line charts, box diagrams and so on, to show trend of changes intuitively. In the actual analysis process of this passage, we computed the distribution of site visits and page views changed with the length of visit. These results of statistical analysis could help to determine rationality of Agridata topology map and convenience of users' access to resources, which could useful for improving structure of the website.

(2)Cluster analysis

This was an important method for data partitioning or packet processing. Currently, the cluster analysis algorithm could be divided into the method, the level of methods, density-based methods, grid-based methods and model-based methods and so on. As there were large amounts of data resources in Agridata and the purpose of users was not clear (the meaning of users could not be accurately expressed with mathematical expressions) when they searched data in Agridata, in this passage, we applied fuzzy clustering analysis method for analyzing users' information behaviors. This method combined with fuzzy set theory and the traditional clustering method. The principle of this approach was as follows: first, the given objection was divided into a number of equivalence classes with the way of fuzzy equivalence relation, which was meaning that establishing fuzzy similar matrix to express similarity between each sample. Then, we clustered with these equivalence classes directly. During the process of specific implementations, we clustered pages of the same database users visited and regarded these pages as a whole for analysis. To these pages, we could compute some parameters such as click-through, access duration and navigation path, for analysis and displaying the users' information behaviors totally. Furthermore, this method for analysis could one-sidedness caused by reference to a single characteristic value

Based on the above analysis, in the cluster analysis, we let $X = \{X_1, X_2, ..., X_n\}$ be a set of characteristic parameters which included values of click-through, access duration and navigation path, etc. The set of pages of one database could be expressed as $W = \{W_1, W_2, ..., W_k\}$. The set of users could be expressed as $U = \{U_1, U_2, ..., U_m\}$. To any user in the set of U, that was $\forall U_i \in U (i = 1,2,..., m)$, the process of the user visited pages could be

expressed as $U_{iWj} = \{y_{i1}, y_{i2}, ..., \quad y_{in}\}$, where $y_{ib}(b = 1,2,..., \quad n)$ represented the value of characteristic in the set of X. After normalizing the elements of U_{iWj}, we could get a multiple fuzzy set of view features for any user.

$$
\begin{aligned}
K = \{\{ & y_{i11}, y_{i12}, ..., \quad y_{i1n}\} \\
, \{ & y_{i21}, y_{i22}, ..., \quad y_{i2n}\} \\
, ..., & \\
\{ & y_{ik1}, y_{ik2}, ..., \quad y_{ikn}\}\}
\end{aligned}
$$

As a result, the interest value of a user U_i to a set of pages could be expressed as (1).

$$
\theta_{ij} = \frac{\sum\limits_{t=1}^{k} y_{itj}}{\sum\limits_{p=1}^{k} \sqrt{\sum\limits_{t=1}^{n} y^2_{ipt}}}
\tag{1}
$$

Where $j = 1,2,....,\quad n$, θ_{ij} represented the interest value of a user U_i to a set of pages W for the characteristic X_j

3   Experiments and Analysis

In the experiment of this passage, we chose the website of Agridata as the representative of agricultural sites and analyzed information behaviors of its users. As thinking about the life and production situation of agriculture network users, in this passage, we chose web log files in July 2010 as the objection of experiments. Such sample we chosen could not only reflect needs of agriculture users during the busy farming season, but also avoid deviations of information behaviors in the general sense caused by various holidays.

In this passage, we used the analysis method combining with statistical analysis and cluster analysis. The flow chart of analysis was shown in figure 2.

During the process of analysis, after processing 40873 records in the log file chosen as the objection of the experiment, we computed some parameters such as time distribution of access site, access duration, the amount of download data and the concentration of IP address, etc, to provide the basis for analysis information behaviors of agriculture users. These parameters also helped us to determine the future direction of

development and related factors of agriculture users' information behaviors. In this passage, we analyzed the chosen log files with WebLog Expert, a common log analysis tools. Some results gotten from the experiment were shown in figure 3.
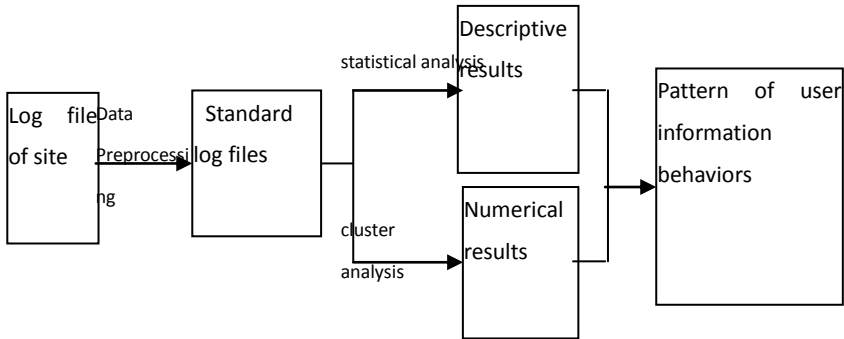
Fig. 2. The flow chart of analysis
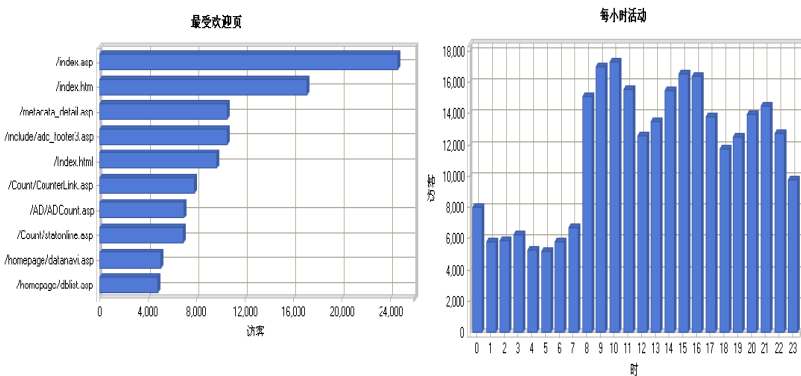
Fig. 3. Results from analysis with WebLog Expert

In this passage, to simplify the analysis, we chose 7 databases from 62 databases of Agridata as objections of cluster analysis during the process of cluster analysis. These 7 databases represented text-based database, numerical databases and image databases. Details of these databases were shown in table 2.

Records of these 7 databases which users visited from the website in log files should be extracted and clustered based on databases for the analysis. To simplify computing, we chose page residence time of users as a characteristic for analyzing. After clustered this feature, we computed the interest value for the database we chosen as an objection for the experiment with these parameters. In the actual calculation, the interest value could be computed with formula (2) which derived from formula (1).

Table 2. Details of chosen databases

Number	Name of the database	Number of recorders	Data type
1	agricultural insect in china database	788	Image
2	natural enemies spider database	338	image
3	Parameter of organic fertilizer database	109	data
4	Fertilizer parameters database	145	data
5	quality standard of rice database	53	text
6	quality standard of tea database	55	text
7	quality standard of green food database	72	text

$$\theta_j = \frac{F_j}{T},$$
(2)

Where $j = 1,2,3..., 7$, F_j represented the time a user stayed in the page set of database j; T represented the total time a users used when he visited the website of Agridata.

The overall interest value of these seven databases could be gotten from the mean of all users' interest values for these seven databases, which were shown in table 3.

Table 3. the overall interest value of databases

number	Name of database	the overall interest value
1	agricultural insect in china database	0.51
2	natural enemies spider database	0.49
3	Parameter of organic fertilizer database	0.29
4	Fertilizer parameters database	0.35
5	quality standard of rice database	0.48
6	quality standard of tea database	0.35
7	quality standard of green food database	0.33

From analysis results above, we could infer information behavior characteristics of users who visited the website of Agridata.

(1) From data of table 3, we could infer that the interest value of images was the highest, which were 28% higher than the interest value for text database and 56% higher than the interest value for data database. It fully confirmed that acceptance of

users for image was higher than that of text and data for data. After further studying data in table 3 and results of the statistical analysis, we could find that IP addresses of users who access image databases were many and scattered, the one for text databases was second and IP addresses for data databases were the most concentrated. It could mean that agricultural workers and farmers preferred image data while numerical data were often used by agricultural researchers.

(2) From results of the statistical analysis, we could determine that the access time period of users was very scattered, which could mean that users could visit the website at any time of a day. In general, in the day time (about 8:00 to 18:00), click-through of the website was maximum, around 74% of all click-through and IP addresses of visited were more concentrated. So we could infer that agricultural researchers in institutions were more willing to use Agridata than that of farmers.

(3) From results of statistical analysis for main page and search pages residence time of users, we determined that IP addresses of users whose pages residence time was shorter had displayed more concentrated while longer pages residence time of users whose IP addresses were more scattered. Thus, we could infer that farmers were much weaker for adapting to main page and search pages than agricultural researchers did.

From results of the analysis in this passage, the website of Agridata should be improved in three aspects. Firstly, we should strengthen publicity efforts to farmers for Agridata; secondly, the main page and search pages of Agridata should be simplified to make them easier to use; finally, the number of image databases should be increased to meet the need for farmers. If more farmers visited the website of Agridata for life and production, we could sure that Agridata should provide better services for farmers and promote agricultural production and research.

4   Conclusions

In this passage, the Web access log was used as an object and we discussed the method for studying the information behavior of users of agricultural methods with statistical analysis and cluster analysis. Then we conducted experiments with the Web access log of the national agricultural scientific data center and analyzed the pattern of information behavior of its user. With these results, several advices on this website were made for providing a better service to agricultural users and promoting agricultural production and research.

References

1. Ding, J.-D.: Analysis and Research on the Web-journal of University Library. Sci-Tech Information Development & Economy 16(6), 5–6 (2006)
2. Feng, C.-H.: Research on Application of Web Log Mining in Network Teaching. Computer Technology and Development 20(6), 183–187 (2010)

3. Li, J., Sun, J.: Empirical Study on Website Quality,User Perception and Technology Adoption Behavior. Journal of the China Society for Scientific and Technical Information 28(3), 227–236 (2011)
4. Liang, X.-X., Wang, F.: A survey and prospect of log analysis based on clustering. Journal of Yunnan University(Natural Sciences Edition) 31(S1), 52–55 (2009)
5. Peng, Q.: Research of Library Web- based Resources Log Analysis. The Journal of the Library Science in Jiangxi 31(1), 30–31 (2004)
6. Ping, T.: The application of Server Log Analysis Method in Website Improvement. Journal of Jiujiang University(Natural Science Edition) (4), 31–33 (2010)
7. Weng, C.-P.: A Web Log Based University Library Users Information Behavior _ Taking the AnHui University Library as the Example. AnHui University, Anhui (2010)
8. Zhang, X.: Log File Analysis of Peking University Library Homepage. New Technology of Library and Information Service (5), 81–83 (2005)
9. Tang, R., Solomon, P.: Toward an Understanding of the Dynamics of Relevance Judgment: An Analysis of One Person'S Search Behavior. Information Processing & Management 34(2/3), 237–256 (1998)

Spatial Pattern of Plant Specimen and Its Implications in Conservation Biology in Hengduan Mountains of Southwest China[*]

Hanming He[1] and Jianmeng Feng[2,**]

[1] Key Lab. of the Ministry of Education for Agro-Biodiversity and Pest Control, Yunnan Agricultural University, Yunnan, Kunming 650201, China
[2] Department of Life Science and Chemistry, Dali University, Dali 671000, China
`fjm@pku.org.cn`

Abstract. In the past, the spatial patterns of specimen have not been paid enough attention. In this study, we probed the spatial pattern of plant specimen and its implications in conservation biology in Hengduan Mountains, one of the hotspots of global biodiversity, based on the dataset extracted from Chinese Virtual Herbarium. The results showed that there were big differences of specimen density among the units of the studied area. High specimen density was mainly found in the boundary regions of Northwest Yunnan, Southwest Sichuan and Southeast Tibet, the central part and the southern part of the studied area. The interest of the collectors or botanists was mainly focused on the units with low population density (low disturbance of human activity) and high complexity of topography. With the increase of specimen density, an increasing trend of species density was observed. Most of the units studied were not paid enough attention in the history of specimen collection. If we collected specimen at higher density in blank or marginalized area, we may have much more chances to find more taxonomies. The spatial pattern of specimen density may shape our understanding of the spatial pattern of specie diversity. We should not only inherit specimen from our predecessors, but also, we should probe into specimen repositories to understand biodiversity status and its spatial pattern.

Keywords: Altitudinal difference against area, Hengduan mountains, Population density, Spatial pattern, Specimen density, Species density, Vascular plants.

1   Introduction

Taxonomic records and their conservation are the basis of biodiversity research (Bickel, 1999; Ponder et al. 2001; Rudolf & Torsren, 2004). With the accumulation of specimen in the long history of specimen collection, its implications in conservation biology are being paid much more attention in these days (Ponder et al. 2001). The inquiry into large specimen discussions can help us understand the characteristics and

---

the spatial patterns of biodiversity in local or global scale (Chew & Oheim, 2007; Greenstreet & Piet, 2008; Rudolf & Torsren, 2004). Also, it lends us a hand to detect the hotspots, in which lots of energy and time have been inputted in the past. Moreover, the probe into huge specimen discussions can help us to find or predict the blank or marginalized area neglected in the history of specimen collection more or less, and in which we may find more Taxonomies un-discovered in the past (Radim, 2007; Rudolf & Torsren, 2004). Therefore, the investigation on the spatial pattern of specimen and its implication in conservation biology can not be under-estimated.

Myers et al. (2000) published their research on biodiversity hotspots at global scale in Nature Magazine. The conclusions disclosed that Hengduan Mountains in Southwest China held high richness of biodiversity, and was one of 25 biodiversity hotspots in the world. Therefore, the study on the biodiversity in this region has drawn extensive and intensive attention (Sun, 2002). So, there was a long history of plant specimen collection, and a large amount of plant specimen from the studied area were accumulated (Wu, 1988), which made the inquiry into the spatial pattern of plant specimen and its implications feasible. However, until now, no relevant study was reported.

Supported by China National Science & Technology Infrastructure Center (SCNSTIC), Chinese Virtual Herbarium (CVH, www.cvh.org.cn) is aimed to provide essential and comprehensive information of plant specimen in China (Institute of Botany, Chinese Academy of Sciences, 2010). The herbarium covers most of plant specimen resources in China, with more than 2.5 million plant specimen digitalized and updated online. Until now, the herbarium is believed as the most complete database of plant specimen in China, serving us with taxonomic and distribution information of plant specimen across China with high reliability. So, our aim is to explore the spatial patterns of plant specimen and its implication in conservation biology in Hengduan Mountains, based on the dataset extracted from CVH. We hope this study can help us understand and protect plant diversity in the studied area. It must be noted that in this paper, we only focused the vascular plants in Hengduan Mountains.

2   Data Collection and Methods

All information of plant specimen in the studied area, including taxonomic and distribution information, was extracted from Chinese Virtual Herbarium. Then, we built a database of plant specimen of Hengduan Mountains, with more than 120,000 specimens and the information of plant specimen in each unit (county). Species richness in each unit was estimated from the specimen information. The topographical information and population density were from literatures (Zhao, 2005). To exclude the effects of area on the spatial pattern of specimen, we adopted specimen density to represent the amount of plant specimen in each unit. In the similar way, we adopted species density to stand for species richness (Logarithm transformation of area was conducted). In this paper, we adopted altitudinal difference against area (logarithm transformed) to represent topographical complexity, and the altitudinal difference was the difference between the highest altitude and lowest altitude in each unit studied.

Based on the dataset, dispersion coefficient was estimated to probe the differences of specimen density among 87 units studied. With the help of ArcGIS Desktop 8.3, we

examined the spatial pattern of specimen density. To explain the pattern, Pearson's correlation between specimen density and population density, altitudinal difference against area was conducted. To check our hypothesis that the spatial pattern of specimen may shape our understanding of the patterns of species diversity, we conducted Pearson's correlation between specimen density and species density. Above analysis and drawings were completed with the help of ArcGIS Desktop 8.3, MS Excel and SPSS.

3   Results

3.1   Frequency Distribution of Specimen Density and the Discreteness Analysis

There were big differences of specimen density among 87 units studied (Fig.1). The lowest density was zero, while the highest density was 3.4 specimens per km2. There were 46 units (accounting for 52.9% of all units) whose specimen density was lower than 0.1 specimen per km2. Among 87 units studied, there were 78 units whose density was lower than 1 specimen per km2. It implied that most of the units displayed low specimen density, while high density was only observed in a small fraction of units studied. The discreteness analysis showed that the discreteness coefficient was 1.7, markedly higher than 1, which meant that the distribution of specimen density among 87 units studied was extremely uneven.
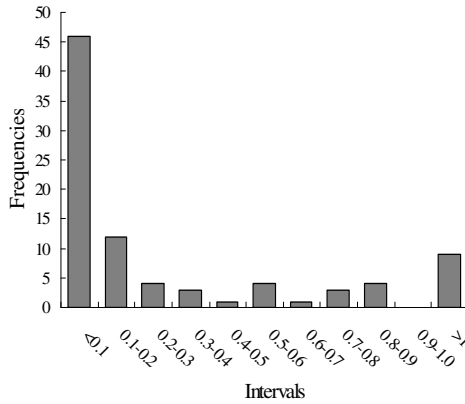
Fig. 1. Frequency distribution of specimen density in the units of Hengduan Mountains

3.2   Spatial Pattern of Specimen Density and Its Interpretation

High specimen density was observed in the boundary regions of the Northwest Yunnan, Southwest Sichuan and Southeast Tibet, and we also found the high density in the southern units and the central units of Hengduan Mountains (Fig.2). Except these units, low specimen density or blank units were detected in most of the units studied. Pearson's correlation indicated that with the increase of population density, there was a

slight decreasing trend of specimen density among 85 units studied, except irregular data points of two units (Fig.3). There was strong correlation between specimen density and altitudinal difference against area (topographical complexity), and with the increase of the altitudinal difference against area, an increasing trend of specimen density can be predicted (Fig.4).
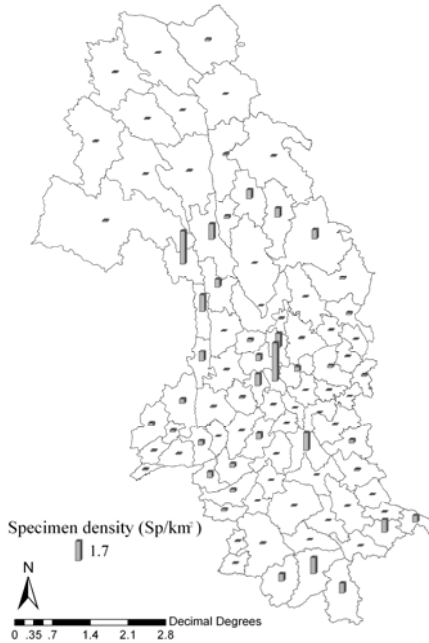
Fig. 2. Spatial patterns of specimen density of vascular plants in Hengduan Mountains
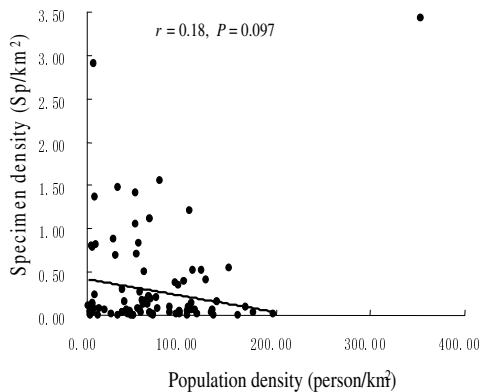
Fig. 3. Correlation between population density and specimen density
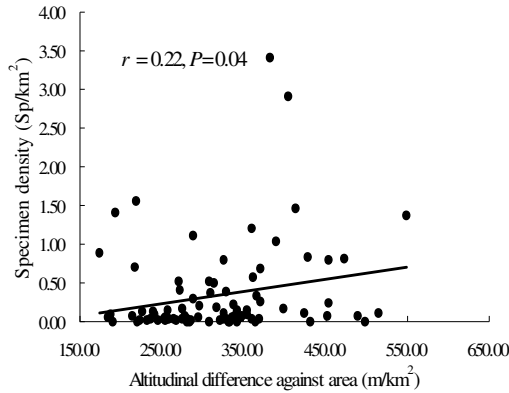
Fig. 4. Correlation between Altitudinal difference against area and specimen density

3.3   Correlation between Specimen Density and Species Density

We observed that there was strong correlation between specimen density and species density, and with the increase of specimen density, an increasing trend of species density can be perceived (Fig.5). It implied that in the units with lower specimen density, lower species density can be probably predicted, and vice versa.
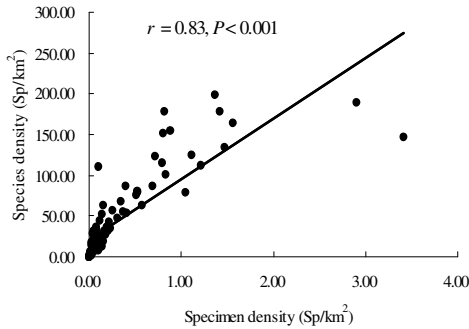
Fig. 5. Correlation between species density and specimen density

3.4   The Implications in Conservation Biology

The discreteness analysis indicated that the extreme unevenness of specimen density can be observed among the units studied, and the spatial pattern of specimen density also showed that the higher density was mainly found in the boundary regions of the three provinces, the central units and the south units of the studied area, which implied that these units were hotspots in the history of specimen collection and biodiversity research. However, in most of the units studied, the density was very low, and even some of them

were blank area. So, we can infer that most of the units studied were paid little attention, or were overlooked in the history of specimen collection, more or less. The Pearson's correlation indicated that with the increase of altitudinal difference against area, an increasing trend of specimen density can be observed. Also, the results showed that there was slight negative correlation between population density and specimen density. It may imply that the interest of the collectors or botanists was mainly focused on the units with lower population density (low human disturbance) and higher complexity of topography, probably because raw or undisturbed vegetation and high richness of plant diversity in units with complex topography attracted their attention.

Previous study showed that if we conducted intensive specimen collection in the blank or marginalized regions, there were much more chances to find more taxonomies un-discovered in the regions in the past, compared with the attempt in biodiversity hotspots (Rudolf & Torsren, 2004). In this paper, both discreteness analysis and the spatial pattern indicated that lots of blank or marginalized units were overlooked in the history of specimen collection. Therefore, we tentatively deduced that if we collected specimen at higher density in these blank or marginalized area, we may discover more species or taxonomies, which may enrich and refine our understanding of biodiversity in Hengduan Mountains and its spatial pattern.

In the results, we observed that there was strong correlation between species density and specimen density. It implied that higher intensity of specimen collection may enrich our findings in plant taxonomy in the studied area. It also indicated that specimen density would possibly affect our understanding of biodiversity in the studied area. If the intensity of specimen collection was low or not enough, we may underestimate the biodiversity in the studied area. However, if we conducted specimen collection at higher or enough intensity, we may find more species or taxonomies, and the results may more accurately reflect the biodiversity status in the studied area. Recently, Van Liefferinge (2010) observed that the investigation at higher sampling density may more reasonably reflect the biodiversity of fish in small brooks in Belgium. The studies in fungus also obtained similar conclusion (Whitcomb & Stutz, 2007). Moreover, in the field of genetics diversity research, the studies testified that sampling density is an important factor affecting our understanding of genetics diversity (Alberton et al. 2006; Miyamoto et al. 2008; Muirhead et al. 2008). Other relevant studies also supported our conclusions (Chew & Oheim, 2007; Greenstreet & Piet, 2008). Therefore, the implications of the intensity of specimen collection can not be under-estimated.

The strong correlation between specimen density and species diversity also hinted that the spatial pattern of specimen density may shape our understanding of the spatial pattern of species richness. Recently, most of biodiversity research in the studied area focused on interpreting species richness patterns with climatic and topographical factors (Fu et al. 2007; Sherman et al. 2008; Yao et al. 2010; Zhang et al. 2009). But, our results implied that the spatial pattern of specimen density may be a strong potential factor, and may shape or even distort our understanding of spatial patterns of biodiversity. Considering the unevenness of the spatial pattern of specimen density and lots of blank or marginalized units, we tentatively assumed that we had not fully understood the spatial pattern of plant species diversity in Hengduan Mountains. Rodrigo & Michael (2007) investigated the fern diversity in Bolivia, and found that

sampling density strongly shaped their understanding of spatial patterns of fern diversity, which in some extents supported our conclusions.

The results showed that with the increasing of specimen density, the increasing trend of species density can be evidently observed. It may imply that the overall intensity of specimen collection was not enough. We should increase our intensity of specimen collection to enrich our understanding of plant diversity in the studied area. And it may also imply that the plant specimen in our dataset only represent a small fraction of plant species in the studied area. In 2004, Rudolf & Torsren explored the species diversity of Euscelidia, and revealed that the specimen in museums only represented 50% of species in reality (Rudolf & Torsren, 2004), which in some extents supported our results.

The specimen was important and fundamental resources in conservation biology, and we should input much energy and time on the collection, arrangement of specimen and the construction of specimen repository. However, in the past, the specimen's implications in conservation biology have not been paid enough attention in the studied area. In this paper, we suggested that we should not only inherit specimen from our predecessors. But also, we should probe into the specimen discussions to understand, assess the biodiversity status and its spatial pattern, and evaluate its hotspots and marginalized area. Furthermore, compared with other methods, the research based on specimen repository is one of the most efficient methods at the lowest costs (Rudolf & Torsren, 2004), therefore, we should fully exploit the resources of specimen repository (Ponder et al., 2001; Rudolf & Torsren, 2004).

4   Conclusions

In this study, we probed the spatial pattern of plant specimen and its implications in conservation biology in Hengduan Mountain. The results showed that there were big differences of specimen density among the units of the studied area. High specimen density was mainly found in the boundary regions of Northwest Yunnan, Southwest Sichuan and Southeast Tibet, the central part and the southern part of the studied area. The interest of the collectors or botanists was mainly focused on the units with low population density (low disturbance of human activity) and high complexity of topography. With the increase of specimen density, an increasing trend of species density was observed. Most of the units studied were not paid enough attention in the history of specimen collection.

References

1. Alberton, O., Kaschuk, G., Hungria, M.: Sampling effects on the assessment of genetic diversity of rhizobia associated with soybean and common bean. J. Soil Biol. Biochem. 38, 1298–1307 (2006)
2. Bickel, D.J.: What museum collections can reveal about species accumulation, richness, and rarity: an example from the Diptera. In: Ponder, W., Lunney, D. (eds.) The other 99%: the Conservation and Biodiversity of Invertebrates. J. Royal Zoological Society of New South Wales, Mosman, Australia, Mosman, Australia, pp. 174–181 (1999)

3. Chew, A., Oheim, K.: Teasing apart the effects of taphonomic and sampling bias on species diversity estimates using gis. J. Vertebr Paleontol. 27, 58A (2007)
4. Chinese Virtual Herbarium: Institute of Botany, Chinese Academy of Sciences, Beijing, `http://www.cvh.org.cn` (cited March 1, 2010)
5. Fu, C.Z., Wang, J.X., Pu, Z.C., Zhang, S.L., Chen, H.L., Zhao, B., Chen, J.K., Wu, J.H.: Elevational gradients of diversity for lizards and snakes in the Hengduan Mountains. J. China. Biodivers & Conserv. 16(3), 707–726 (2007)
6. Greenstreet, S.P.R., Piet, G.J.: Assessing the sampling effort required to estimate alpha species diversity in the groundfish assemblages of the North Sea. J. Mar. Ecol.-Prog. Ser. 364, 181–197 (2008)
7. Miyamoto, N., Fernandez-Manjarres, J.F., Morand-Prieur, M.E., Bertolino, P., Frascaria-Lacoste, N.: What sampling is needed for reliable estimations of genetic diversity in Fraxinus excelsior L. J. Ann For Sci. 65, 403–410 (2008)
8. Muirhead, J.R., Gray, D.K., David, W.K., Sandra, M.E., Damiel, D.H., Hugh, J.M.: Identifying the source of species invasions: sampling intensity vs. J. Genetic Diversity. Mol. Ecol. 17, 1020–1035 (2008)
9. Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Kent, J.: Biodiversity hotspots for conservation priorities. J. Nature 43, 853–858 (2000)
10. Ponder, W.F., Carter, G.A., Flemons, P., Chapman, R.R.: Evaluation of museum collection data for use in biodiversity assessment. J. Conserv. Biol. 15, 648–657 (2001)
11. Radim, H.: Is Sampling subjectivity a distorting factor in surveys for vegetation diversity? J. Folia Geobot. 42, 191–198 (2007)
12. Rudolf, M., Torsten, D.: Significance of Specimen Databases from Taxonomic Revisions for Estimating and Mapping the Global Species Diversity of Invertebrates and Repatriating Reliable Specimen Data. J. Conserv. Biol. 18, 478–488 (2004)
13. Sherman, R., Mullen, R., Haomin, L., Fang, Z.D., Wang, Y.: Spatial patterns of plant diversity and communities in Alpine ecosystems of the Hengduan Mountains, Northwest Yunnan, China. J. Plant Ecol-UK 1(2), 117–136 (2008)
14. Soria-Auza, R.W., Kessler, M.: The influence of sampling intensity on the perception of the spatial distribution of tropical diversity and endemism: a case study of ferns from Bolivia. J. Divers Distrib. 14(1), 123–130 (2007)
15. Sun, H.: Evolution of Arctic-Tertiary Xora in Himalayan-Hengduan Mountains. J. Acta Bot. Yunnan 24, 671–688 (2002)
16. Van Liefferinge, C., Simoens, I., Vogt, C., Cox, T.J.S., Breine, J., Ercken, D., Goethals, P., Claude, P.M.: Impact of habitat diversity on the sampling effort required for the assessment of river fish communities and IBI. Hydrobiologia 644, 169–183 (2010)
17. Whitcomb, S., Stutz, J.C.: Assessing diversity of arbuscular mycorrhizal fungi in a local community: role of sampling effort and spatial heterogeneity. J. Mycorrhiza 17, 429–437 (2007)
18. Wu, Z.Y.: The Hengduan Mountains flora and her significance. J. Jpn. Bot. 63(9), 1–14 (1988)
19. Yao, Y.H., Zhang, B.P., Han, F., Pang, Y.: Diversity and Geographical Pattern of Altitudinal Belts in the Hengduan Mountains in China. J. Mt Sci. 7(2), 123–132 (2010)
20. Zhang, D.C., Zhang, Y.H., Boufford, D.E., Sun, H.: Elevational patterns of species richness and endemism for some important taxa in the Hengduan Mountains, southwestern China. Biodivers. Conserv. 18(3), 699–716 (2009)
21. Zhao, D.H.: Album of Yunnan, 1st edn. Sinomap Press of China, Beijing (2005)

Determination of Corn Nutrient Status under N&K Stressed Condition Using Hyperspectral Analysis

Haihua Wang, Minzan Li[*], and Yane Zhang

College of Information and Electrical Engineering, China Agricultural University,
Beijing 100083, P.R. China
{wanghaihua,gpac,zye}@cau.edu.cn

Abstract. Variable fertilization for crops, like corn, depends on monitoring nutrition condition. Thus, hyperspectral reflectance was used to predict chlorophyll and total nitrogen content under different N,K treatments during corn growth. The area of the experimental field was divided into 3 strips with different nitrogen treatment (N1:0 kg/ha–low, N2:314 kg/ha–normal, and N3:653 kg/ha–high). In each strip, there were 3 repetitions of 3 potassium treatments(K1:0 kg/ha–low, K2:214 kg/ha–normal, and K3:500 kg/ha–high). The results show that growth stages happened in advance under low nitrogen treatment and reflectance intensity, total N, and chlorophyll content are all largest under N2K1 in shooting stage. GNDVI(R2=0.88, RMSE=0.08) performs well for chlorophyll prediction under N3K2. In addition, MLR has the potential of determination of chlorophyll(R2=0.94, RMSE=0.02) and total N(R2=0.97, RMSE=0.09) in trumpet stage, as well as PLSR for chlorophyll(R2=0.99, SEC=0.01;SEP=0.09) and total N (R2=0.96, SEC=0.11;SEP=0.47) in trumpet stage.

Keywords: Hyperspectral Reflectance, Corn, Nutrient, MLR, PLS.

1 Introduction

The corn (*Zea mays*) is cultivated large scale as significant food and economic plant. More inorganic fertilizers are undertaken to gain higher yield in conventional way, however it has been proved that excessive nitrogen(N) fertilizer will reduce the yield and harm environment. So, it is important to give growers a reasonable advice based on growth nutrition condition.

Since chlorophyll content is related to response to nutrition condition, the traditional fertilization is executed based on observing leaf color. But, in fact, this method is not reliable because growth is under complex conditions. Compared to chemical analysis, optical method is nondestructive and fast. Near infrared spectral technology has been used widely in past decades and proved the potential of prediction of growth condition[1-7]. Tang shows that the contents of chlorophyll and carotenoid are corresponded to spectra[8]. Cheng reports LAI and Chlorophyll can be predicted by the first derivative spectra of corn canopies[9].

Different treatment of nitrogen(N) and potassium(K) can induce growth nutrition change. This paper described the changes of total nitrogen and chlorophyll content, tested by chemical method, in corn leaves under different match of N and K, then studied how to use hyperspectral analysis to predict total N and chlorophyll content. Wang reported that potassium(K) could affect the spectra of corn leaf[10]. Walburg considered that ratio of NIR to red was more advantage than red in N detection[3]. Cheng found some sensitive wavelength regions in different growth stages under different nutrient treatments (N,P, and K) [9].

2   Materials and Methods

2.1   Experiment Design

Nongda 86 cultivar was selected as experimental corn and was sowed on April 25, 2009 at Shangzhuang Experimental Station of China Agricultural University, Beijing. Fertility of soil was detected before fertilization control. The area of the experimental field was about 12 ha and was divided into 3 strips with different nitrogen treatment (N1:0 kg/ha–low, N2:314 kg/ha–normal, and N3:653 kg/ha–high). In each strip, there were 3 repetitions of 3 potassium treatments(K1:0 kg/ha–low, K2:214 kg/ha–normal, and K3:500 kg/ha–high). So the total number of samples was 54. In addition, water and other fertilizers were treated under normal condition.

2.2   Spectra

Spectra was obtained under no wind condition between 10:00 and 13:00 and the dates were related to corn growth stages(Table 1). Trumpet and anthesis-silking stage were illustrated in Figure 1a and 1b respectively.

Table 1. Growth stages of corn (2009)

Stage No.	1	2	3	4	5	6	7
Name	Seedling		Shooting		Trumpet	Anthesis-silking	Filling
Date	5-27	6-02	6-11	6-23	6-30	7-07	7-15
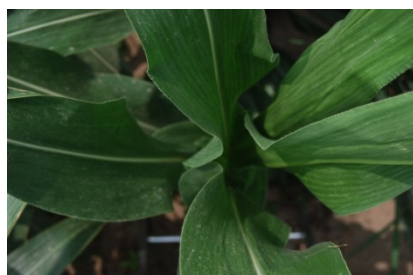
(a)   (b)

Fig. 1. Images of (a) Trumpet stage and (b)Anthesis-silking stage

FieldSpec spectrometer (ASD Co., USA) was used to obtain reflectance spectra which covered from 325 to1075nm with 1nm resolution. Spectra were collected in the middle part of the fourth(up to down) corn leaves. Three repetitions for each leaf were averaged for data processing. The standard white board was used as reference for each 3 leaves by equation (1):

$$R = (R_S - Dark) / (R_W - Dark) , \tag{1}$$

where R_S is reflectance of corn leaf; R_W is reflectance of the standard white board; and Dark means that no light enters the spectrometer. Considered that nitrogen may shift in filling stage, data was sampled from ear leaves.

2.3   Determination of Chlorophyll and Total Nitrogen Content

The same leaves were packaged with plastic bags for determination of chlorophyll and total nitrogen content after spectra collection. Chemical experiment and SP-2102 UV spectrophotometer (Cany Precision Instruments Co., Ltd, Shanghai China) were used together for chlorophyll content. After veins in leaf were removed, leaf was cut into small fragments. Then 0.4g fragments and 25ml solution with 1 for 95% ethanol and 2 for 99% acetone were mixed in a test tube. 24 hours later, the chlorophyll-containing solution was put into the spectrophotometer and A645, A652 and A663 readings were recorded for calculate the chlorophyll content by equation (2):

$$C_T = 20.29A645 + 8.05A663 \text{ or } C_T = 20.29A652 , \tag{2}$$

where C_T is a response value to total chlorophyll content (mg . L^{-1}). The equation (3):

$$Chl = C_T \cdot V \times 10^{-3} / W , \tag{3}$$

was used calculate chlorophyll content Chl(mg . g^{-1}), where V is volume of extract liquor, 25ml; W is fresh weight of leaf, 0.4g.

To total nitrogen detection, rest fragments were dried in oven and grinded into powder. 0.4g powder and 6.2g $CuSO_4 \cdot 5H_2O$ were mixed, then reacted with 60ml concentrated sulfuric acid in a tube. The tube was heat at 500℃ for 2 hours, then was cooled. Kjeldahl (FOSS NIRSystems, Inc., US) was used to determinate total nitrogen content(mg. g^{-1}) of the leaf in solution.

2.4   PLSR Analysis and Multiple Linear Regression

Log (1/R) was often used as absorbance for spectral preprocessing. First and second derivative were also taken into eliminate the baseline drift error. Compared the derivative spectra of high and low total N content, we selected a set of G, R, NIR wavelengths. Then single-band spectra and some VIs(Vegetable Indexes), like red edge slope, NDVI, GNDIV, DVI, RVI and SVI, were used to predict chlorophyll and total nitrogen content.

MLR(Multi Linear Regression) and PLSR (PLS Regression) were also introduced to determinate chlorophyll and total nitrogen content in this paper, which were all multivariate data processing methods and fit to hyper spectra. PLS (Partial Least Squares) was based on MLR and PCA (Principle Component Analysis), and PCA had the potential for dimension reduction under considerable variables. So, PLSR was more powerful to calibrate spectrum data than MLR. However, Over-fitting can happen to validation sometimes. In this paper, MATLAB (Math Works Inc., Natick, MA, USA) was used for data processing tool.

3   Result and Discussion

3.1   Chlorophyll and Total Nitrogen Content

Figure 2 and 3 show chlorophyll and total N content under different N treatments in different growth stages respectively. Chlorophyll content increased early, then decreased. Total N content had the same change as Chlorophyll, but increased again in filling stage(July 15) for the reason that corn steps into reproductive growth.

To low nitrogen treatment, the chlorophyll content reach highest in early shooting stage(June, 11) , and the total content decreased at shooting stage(June 23). The same growth stages happen in advance for low nitrogen treatment. Under low nitrogen treatment, the high N content in filling stage also proved that grain is difficult to get more nitrogen nutrient. In addition, Figure 2 and 3 indicated that the total N content under high N treatment deceased later than chlorophyll content, and later than normal N treatment.
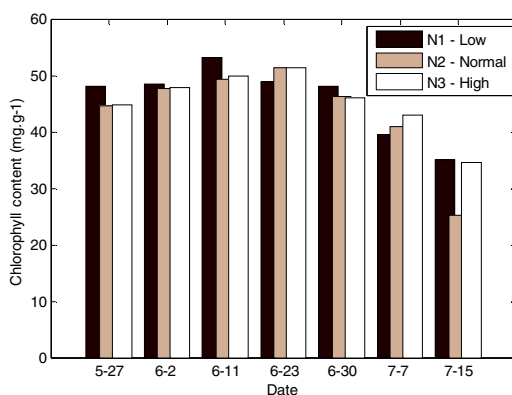
Fig. 2. Chlorophyll content in each stage under different nitrogen treatments
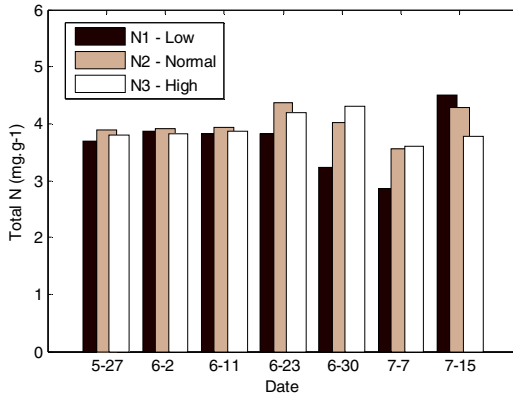
Fig. 3. Total N content in each stage under different nitrogen treatments

Different NK treatments were studied in-depth except low nitrogen treatment, which showed growth ahead and poor performance of nitrogen absorption in filling stage based on the above analysis. Figure 4 and 5 show that total N and chlorophyll content of leaves under N2K1 in the shooting stage are highest. Followed by N3K3 and other normal N, the lowest under N3K2 and N3K1. It can be concluded that normal nitrogen treatment is fit to low potassium and high nitrogen needs high potassium.
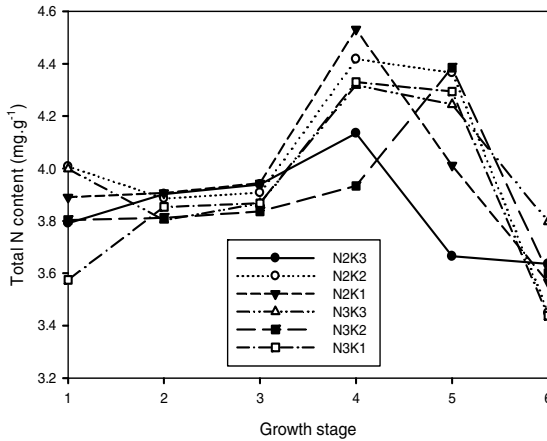
Fig. 4. Total N content in each stage under different N,K treatments
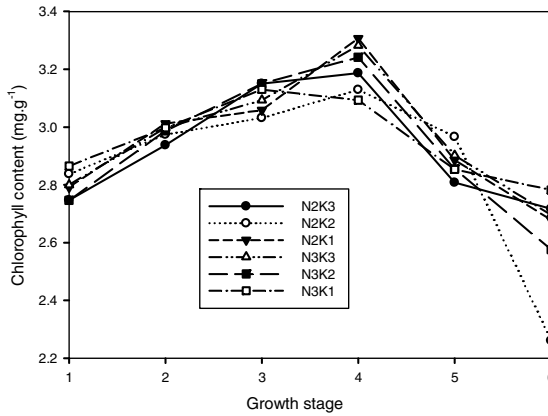
Fig. 5. Chlorophyll content in each stage under different N,K treatments

3.2   Correlation between Total N Content and Chlorophyll Content

Total nitrogen content and chlorophyll content of leaf showed low correlation at different growth stages. In Figure. 6 and 7 , the top two performances happened in shooting stage ($R^2 = 0.83$; RMSE = 0.10) with positive correlation and trumpet stage ($R^2 = 0.69$; RMSE = 0.27) with negative correlation.
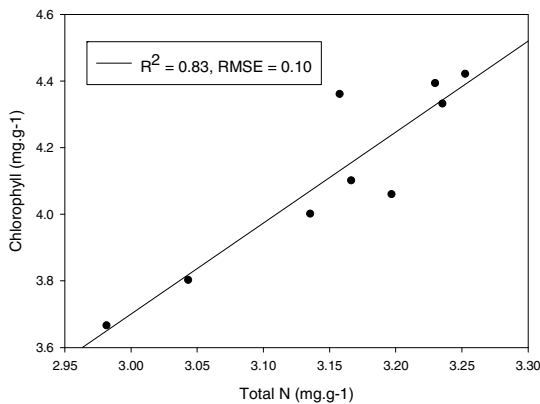
Fig. 6. Correlation between total N and chlorophyll at Shooting stage (June 23)
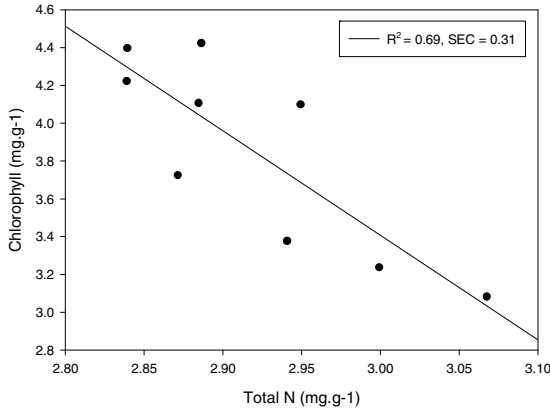
Fig. 7. Correlation between total N and chlorophyll at Trumpet stage (June 30)

The correlation coefficient between total N content and chlorophyll content was given in Table 2, excluding filling stage where they are significant different. Three treatments, where the correlation was higher than 0.8, are N1K3(R2 = 0.83; RMSE = 0.15), N2K1(R2 = 0.81; RMSE = 0.12), and N2K3(R2 = 0.81; RMSE = 0.07). Under low N,K treatment, although growth stage was earlier, the correlation of chlorophyll and total N was higher. In contrast, high nitrogen treatments had low correlation, which indicated that excess nitrogen would affect transform nitrogen to chlorophyll.

Table 2. Correlation between total N and Chlorophyll under different treatments

N,K treatment	R2	RMSE
N1K3	0.83	0.15
N1K2	0.78	0.19
N1K1	0.31	0.32
N2K3	0.81	0.07
N2K2	0.67	0.19
N2K1	0.81	0.12
N3K3	0.24	0.18
N3K2	0.06	0.23
N3K1	0.22	0.29

3.3   Leaf Spectrum

Typical spectra of corn leaves (Anthesis-silking stage; The sample S5, total N = 4.1234%; S26, total N = 2.0707%) are illustrated in Figure. 8. Wavelength ranges below 405 and above 978 were cut with high noise. In the visible region, most of the light was absorbed and there was a small peak in 553nm related to green band. The spectra curve jump to strong reflectance in near-infrared region through red edge (685~750nm), which can be related to the internal organizational structure of corn leaf.
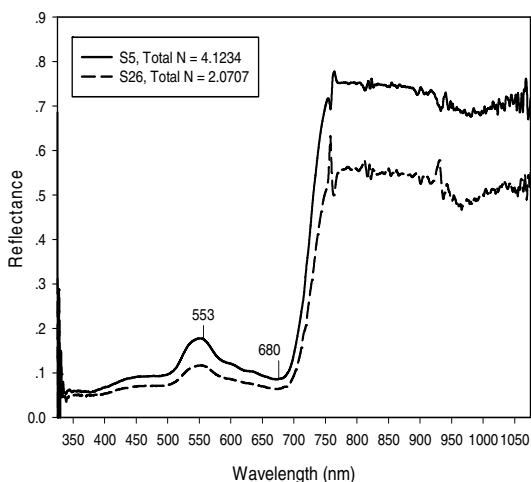
Fig. 8. Reflectance spectra of two corn leaves with high and low total N content

Figure 9 and 10 show spectral reflectance at 553 and 818nm under different N, K treatments in different stages respectively. Overall, changes of reflectance intensity are like wave, and there are two peaks in the shooting stage(Stage 4) and anthesis-silking(Stage 6) stage. The reflectance intensity at 553nm is lowest in trumpet stage(Stage 5).

N2K1, N3K3 were more strong and N3K2, N2K3, N1K2 were lower in shooting stage at both wavelength. It was similar to the result of leaf total N content and chlorophyll content. Compared to other N,K treatments, the intensity of reflectance under N2K1, N3K2 change obviously in trumpet and anthesis-silking stage(Stage 6). High nitrogen treatments increased gradually, and low nitrogen treatment declined in general.
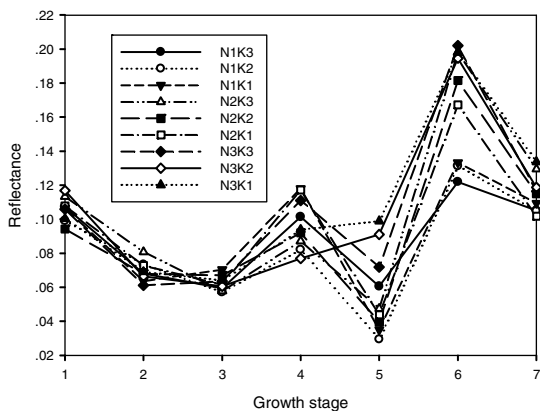
Fig. 9. Reflectance in each stage at 553nm
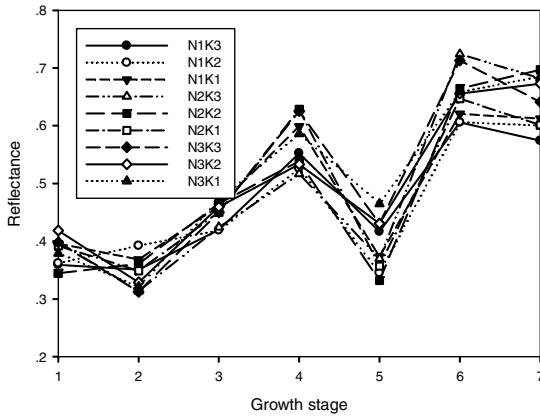
Fig. 10. Reflectance in each stage at 818nm

3.4   Wavelength Selection and VIs Regression

Figure 11 shows Second derivative spectra of log(1/R) of the same corn leaves shown in Figure 8. 553,680,760 and 818nm were selected as green, red, red edge, and near-infrared wavelengths.

Single-band spectra and some VIs, like red edge slope, NDVI, GNDIV, DVI, RVI and SVI, were used to predict chlorophyll and total nitrogen content. The best performance was the correlation between GNDVI(R^2 = 0.88, RMSE = 0.08) and chlorophyll content under N3K2, followed by G (R^2 = 0.79, RMSE = 0.13) under N2K2 and NDVI (R^2 = 0.78, RMSE = 0.11) under N3K2.
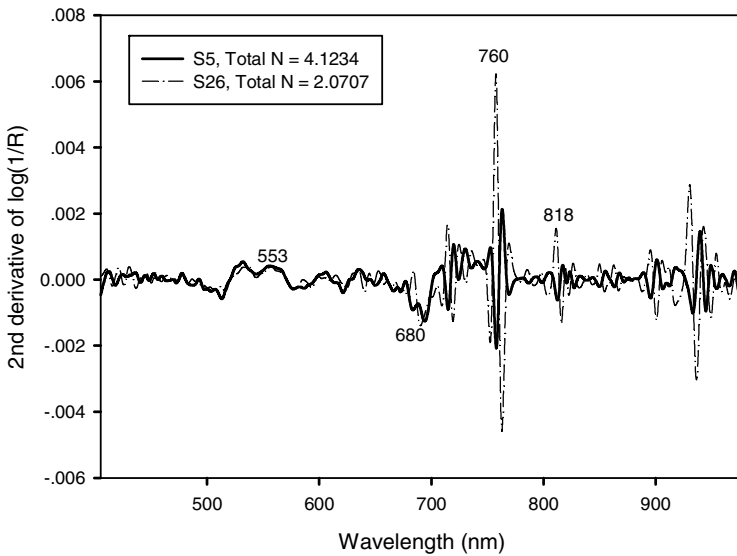
Fig. 11. Second derivative spectra of log(1/R) of the same corn leaves shown in Figure 8

To total N, all top 3 performances happened in anthesis-silking stage(Stage 6), including G ($R^2 = 0.88$, RMSE = 0.12), R ($R^2 = 0.83$, RMSE = 0.14) and GDNVI ($R^2 = 0.81$, RMSE = 0.15) respectively.

3.5   PLSR for Total N and Chlorophyll Content

Some efficient results is shown in Table 3 by PLSR for total N and chlorophyll content in different growth stage. Better performances for Chlorophyll content were shooting stage ($R^2 = 0.95$, SEC = 0.03; SEP = 0.12) and trumpet stage ($R^2 = 0.99$, SEC = 0.01; SEP = 0.09) respectively. Better results for total N content were filling stage ($R^2 = 1.00$, SEC = 0.02; SEP = 0.58) and trumpet stage ($R^2 = 0.96$, SEC = 0.11; SEP = 0.47) respectively. Compare the four results, prediction for total N content is not perfect with high SEP in validation. In addition, The PLSR results for total N and chlorophyll content under different N,K treatment are good for calibration, but over-fitting for validation.

In general, total N and chlorophyll content can be predicted in trumpet and shooting stage using reflectance spectra for assessment of corn nutrition condition.

Table 3. PLSR for total N and chlorophyll content in different growth stage (5 factors)

Stage	Chlorophyll content				Stage	Total N content			
	Calibration		Validation			Calibration		Validation	
	R2	SEC	R2	SEP		R2	SEC	R2	SEP
Shooting	0.95	0.03	0.71	0.12	Trumpet	0.96	0.11	0.72	0.47
Trumpet	0.99	0.01	0.70	0.09	Filling	1.00	0.02	0.53	0.58

3.6   MLR Analysis for Total N and Chlorophyll Content

The intensity of log(1/R) at R, G, NIR, and the slope of red edge were selected to implement MLR to predict total N and chlorophyll content. The result is described in Table 4. The best correlation ($R^{2)}$ for chlorophyll content prediction was 0.94 (RMSE = 0.02), and for total N content was 0.97 (RMSE = 0.09) in trumpet stage(June 30).

Table 4. MLR for N and chlorophyll content in different growth stage

Date	R^2 for Chlorophyll	SEC	R^2 for total N	SEC
5-27	0.72	0.07	**__0.26__**	0.13
6-02	0.54	0.02	0.66	0.04
6-11	0.79	0.05	0.89	0.02
6-23	0.75	0.05	0.78	0.13
6-30	**0.94**	0.02	**0.97**	0.09
7-07	0.51	0.14	0.92	0.10
7-15	**__0.48__**	0.23	0.60	0.21

Table 5 shows that the best correlation (R^2) for chlorophyll content prediction is 0.96 (RMSE = 0.04) under N3K3 treatment, and for total N content is 0.99 (RMSE = 0.04) under N3K1 treatment.

Table 5. MLR for N and chlorophyll content under different N,K treatment

N/K treatment	R^2 for Chlorophyll	SEC	R^2 for total N	SEC
N1K3	0.85	0.09	0.91	0.11
N1K2	0.78	0.17	0.81	0.17
N1K1	**<u>0.59</u>**	0.12	0.68	0.22
N2K3	0.61	0.12	**<u>0.50</u>**	0.12
N2K2	0.88	0.10	0.92	0.09
N2K1	0.80	0.09	0.76	0.14
N3K3	**0.96**	0.04	0.93	0.05
N3K2	0.94	0.06	0.85	0.09
N3K1	0.82	0.05	**0.99**	0.04

4   Conclusion

Chlorophyll and total N content increased early, then decreased until filling stage when reproductive growth started. The correlation between Chlorophyll and total N content was analyzed because growth stages happened in advance under low nitrogen treatment. Chlorophyll and total N content under the treatments of normal nitrogen and low potassium performed best in shooting stage, which means high nitrogen is not necessary.

Reflectance under N2K1 and N3K3, similar to total N content and chlorophyll content, were more strong than N3K2, N2K3,N1K2 in shooting stage at both 553 and 818nm wavelength. GNDVI(R^2 = 0.88, RMSE = 0.08) was better than other VIs for prediction of chlorophyll content under N3K2, as well as G (R^2 = 0.88, RMSE = 0.12) was better for total N content prediction in anthesis-silking stage. In trumpet stage, the correlation of Multiple Linear Regression(553, 680, 760 and 818) for prediction of Chlorophyll and total N content are 0.94(RMSE = 0.02) and 0.97 (RMSE = 0.09) respectively, and PLSR also had good performance for total N content (R^2 = 0.96, SEC = 0.11; SEP = 0.47) and chlorophyll content   (R^2 = 0.95, SEC = 0.03; SEP = 0.12). In a word, total N and chlorophyll content can be predicted in trumpet and shooting stage or some N,K treatments using reflectance spectra for assessment of corn nutrition condition.

References

1. Norris, K.H.: Design and development of a new moisture meter. Agric. Eng. 45, 370 (1964)
2. Gausman, H.W., Allen, W.A.: Optical Parameters of Leaves of 30 Plant Species. Plant Physiology 52, 57–62 (1973)
3. Walburg, G., Bauer, M.E., Daughtry, C.S.T.: Effect of Nitrogen Nutrition on Growth, Yield and Reflectance Characteristics of Corn Canopies. Agronomy Journal 74, 677–683 (1982)
4. Shibayama, M., Akiyama, T.A.: A Spectroradiometer for Field Use: V I. Radiometric Estimation for Chlorophyll Index of Rice Canopy. Japanese Journal of Crop Science 55(4), 433–438 (1986)
5. Yan, Y.-L.: Near-Infrared Spectrum Analysis and Application, p. 518. China Light Industry Press, Beijing (2005)
6. Li, M.-Z.: Spectral Analysis Technology and Application, p. 45. Science Press, Beijing (2006)
7. Wang, J.-H., Zhao, C.-J., Huang, W.-J.: Remote Sensing Quantitative Theory and Application in Agriculture. Science Press, Beijing (2008)
8. Tang, Y.-L., Huang, J.-F., Wang, X.-Z.: Study on hyper Spectral characteristics of corn leaves and their correlation to chlorophyll and carotenoid. Journal of Maize Sciences 16(2), 71 (2008)
9. Cheng, Y., Hu, C., Wang, C., Yu, G.: Physiological response and spectral characteristics of summer corn under nutrient stress condition. Resources Science 23(6), 54–58 (2003)
10. Wang, L., Bai, Y.: Nutrients Change and Spectral Response of Spring Corn Leaf for Varying Amounts of Potassium Fertilization. Journal of Remote Sensing 11(5), 641–647 (2007)

Research in Crop Land Suitability Analysis Based on GIS[*]

Guobing Pan[1,2] and Jianping Pan[2]

[1] Structure Engineering Lab, Chongqing Jiaotong University,
No.66 Xuefu Road Nan'an District Chongqing China, China, 400074
[2] School of Civil Engineering & Architecture, Chongqing Jiaotong University,
No.66 Xuefu Road Nan'an District Chongqing China, China, 400074

Abstract. This paper based on GIS land suitable assessment principle and method, use supermap to analysis suitability of crop land of liao city. The suitability of comprehensive analysis was conducted in Study area for three crops .First, choosing representative and having the same impact the suitability of natural factors and social economic factors of the three kinds of crops; Then, interpolation and empowerment of overlap and analysis in supermap, and the appropriate level of land were classified according to land on crops. Thus can adjust measures to local conditions to crops layout and farming system reform and adjustment, the biggest production potential play land. Experimental results show that in the complex crop suitability assessment process, GIS is an effective, possible means.

Keywords: crop land, Suitable assessment, The geographic information system.

1   Introduction

The land suitable assessment is to evaluate the suitability for a certain purpose land suitable degree of process. Through the comprehensive analysis of natural factors and the social economic factors which have influence on the land, divide the land into some level according to the suitability under the specified way, in order to show that the appropriateness of appropriate degree for a variety of use .At present, the wide application of GIS in land suitable assessment, make land suitable assessment more flexible, science. A single use land suitable assessment has been much attention, but evaluation basically all around economic crop, lack major crops on land suitability analysis research [1~3].

Liao city located in the western of shandong province, north latitude 35 ° 47 '-37 °, 03 ° east longitude'115°16′-116°30′, the total area is 8590 square kilometers, the city is in the Yellow River alluvial plain, flat and tilt from southwest to northeast, average rail slope is about 1/7500, altitude is 27.5 -49.0 meters, belong to temperate and monsoonal climate area, has the remarkable season change and monsoon climate

---

characteristics, belong to semi-arid continental climate. Large land degrees is for 62.8 -64.8, dry degree is in 1.7 -1.9. Overall, agricultural climate resource is more abundant and suitable for growing DuoZhong crops. The climate resources is also rich and from the Yellow River water for irrigation, the climate for DuoZhong crops has better adaptability, therefore, the planting industry has a long history, become the main crops north China cotton, wheat, corn on the region [2~5]. Due to the different influence on the land in liaocity of each factor differences, And with the quickening of the process of urbanization, unreasonable planting and reclamation, so as to make the land resources will not be used reasonably, efficiently. This paper discuss how to apply GIS t0 crop land the suitability analysis in liaocity.

2   Research Methods

The research area (HuangFanOu) of this paper is located in north China plain, the main crops are wheat, corn and cotton. The region in the terrain, sunshine and rainfall, some natural factors change caused by small scale, the requirements of the conditions of land crop of great similarity, that is, in this regions have three crops in the suitability of basic level is consistent. Based on this law, this paper this can be three crops suitability degree analysis. First, choose is representative of the three kinds of crops, to have the same impact the suitability of natural factors and social economic factors; Then, in supermap interpolation, and empowerment of overlap and analysis, and according to land on crops, the appropriate level of land were classified. Thus can adjust measures to reform and adjustment for local conditions to crops layout and farming system, play the biggest production potential of land and create conditions for the rational utilization of land include making land planning.

3   Evaluation Factors Selected

Land suitability level of each assessment unit is a comprehensive effect of Comprehensive role, reasonable choice of evaluation factors is the premise to ensure the suitability of the land quality, therefore, in selecting a factor should be followed in several principles

Leading-the main factor restricting the use of land.

Difference-choosing the factors which have obvious difference and can appear near the value in the study area.

Independence-each factors can't appear causality.

The possibility-factors must have the corresponding information.

According to the above principles, this paper selects 5 major factor as a land of liao city suitable assessment factors, namely: soil types (the class), soil texture, buried deep underground water level, soil fertility and soil pollution, among them soil fertility was expressed through the organic matter, total nitrogen, alkali solution nitrogen, rapidly-available phosphorus and potassium, soil pollution was expressed through the cadmium, mercury, lead, arsenic, chromium and copper six kinds of

heavy metal. Through these five factors, we will get soil type distribution, soil texture map, buried deep underground water level distribution, soil fertility distribution, soil pollution distribution, then, this five distribution empowerment overlap add to get comprehensive evaluation distribution, and then get comprehensive evaluation distribution based on the land suitability of crops degree.

4   The Experiment

4.1   Data Preparation

1、   First, find a paper map, the map evenly distributed sampling points, then conduct field sampling according to the sample point distribution and the needed factor data.

2、   Through the scanner, scanning liao city land map and forming grid data, using this grid data as a screen for reproduction, working in a vectorial. After scanning, the grid data was stored for JPG format. Then use PhotoShop to pretreatment grid data, improve the contrast of raster images and other quality, in order to improve the quality and efficiency of the vector.

3、   Map in digital and attribute editor in supermap, getting sample point distribution (figure 1).
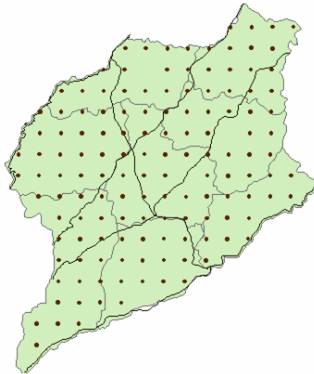
Fig. 1. Sample point distribution

4.2   Suitability Analysis Process

1、 Soil Type Analysis

(1) First, score different soil types (table 1), then build " Soil types _v" column in the sample point data attributes list, the value is listed in table 1 for the "score". Set up the analysis environment.

Table 1. Soil type allocation

Soil types	score
Sand soil meadow	1
Meadows solonchak	1
Alluvial soil	1
solonchak	1
Chao soil salinization	2
Alkalize chao soil	2
Damp soil	3
Take off the tide soil	3
Chao soil	4

(2) Choosing inverse distance square weights interpolation method to do interpolation analysis of soil types, making special charts to get soil type distribution (figure 2).
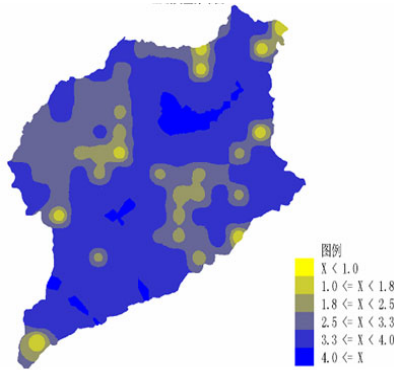
Fig. 2. Soil type distribution

2、 The Soil Texture Analysis

As for soil types analysis, first of all expert scoring (such as table 2), and then build " Soil types _v" column in the sample point data attributes list, the value is listed in table 2 for the "score". Finally interpolation soil texture map obtained (figure 3).

Table 2. Score soil texture

Soil texture	score
Soil mass sand	1
Clay soil mass	1
Stick loam accounting	2
Sandy loam stick	2
Sticky loam	2
Sandy loam	3
loam	4
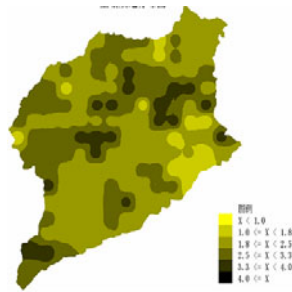
Fig. 3. Soil texture map

3、Groundwater Depth Analysis

According to the underground water level in the region, score suitability of major crops (table 3), then interpolation analysis in the supermap and get in groundwater depth distribution, according to table 3 classify the buried depth of underground water level distribution, get buried depth underground water level classification figure (figure 4).

Table 3. Groundwater depth allocation

Buried deep underground water level(m)	score
2 – 10	4
10 – 30	3
30 – 50	2
50 – 60	1
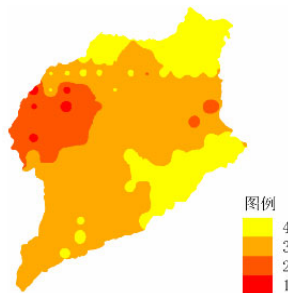
Fig. 4. Buried deep underground water level classification figure

4、 Soil Fertility Analysis

(1) The soil fertility index system

Content of soil organic matter, the soil of nitrogen, phosphorus and potassium, etc fertility factor is the important influence factors of soil fertility and assessment of soil fertility changes, not limited to individual single fertility factors. Therefore, we

combined with the local cropping system and the requirements of the fertility of the soil, the choose organic matter, total nitrogen, alkali solution nitrogen, phosphorus and phosphorus rapidly-available potassium as soil fertility evaluation index, and obtains the single soil fertility index level (table 4).

Table 4. Soil fertility index level

Organic matter (g/kg)	Total nitrogen (g/kg)	Alkali solution nitrogen (ppm)	Rapidly-available phosphorus (ppm)	Rapidly-available potassium (ppm)	level
0.329 – 0.6	0.018 – 0.05	20.65 – 30	Null	6.45 – 30	6
0.6 – 1	0.05 – 0.075	30 – 60	3.8 – 5	30 – 50	5
1 – 2	0.075 – 0.1	60 – 90	5 – 10	50 – 100	4
2 – 2.1072	0.1 – 0.1372	90 – 120	10 – 20	100 – 150	3
Null	Null	120 – 123	20 – 40	150 – 200	2
Null	Null	Null	40 – 79	200 – 331	1

(2) The index weight determination

The index weights mean every soil fertility factor refer to comprehensive fertility soil of contribution. It is the key issue of comprehensive evaluation of fertility how to determine the weight of single fertility index. This paper represents the weight of each index by the analysis of correlation coefficient between each index. Calculation method is: first it calculates the correlation coefficient of single fertility index with SPSS software (table 5), and then it calculates average value $(\bar{r})$ of the correlation

Table 5. Correlation coefficient

	Organic matter	Total nitrogen	Alkali solution nitrogen	Rapidly-available phosphorus	Rapidly-available potassium
Organic matter	1				
Total nitrogen	0.959	1			
Alkali solution nitrogen	0.035	0.076	1		
Rapidly-available phosphorus	-0.0	-0.38	0.563	1	
Rapidly-available potassium	0.036	0.92	0.222	0.137	1

Table 6. Each fertility index correlation coefficient and weight

Fertility index system	Correlationcoefficient average	Weight coefficient
Organic matter	0.238	0.192
Total nitrogen	0.394	0.316
Alkali solution nitrogen	0.224	0.180
Rapidly-available phosphorus	0.06	0.048
Rapidly-available potassium	0.329	0.264

coefficients of a fertility index and other fertility index. And it uses the average of all fertility index correlation coefficient average absolute value and the sum total $(\sum \bar{r})$ ratio $(\bar{r}/\sum \bar{r})$ as the single fertility index in soil fertility of the comprehensive contribution value or weight (table 6).

(3) Soil fertility changes evaluation

According to the results of the above and the application of the soil quality index and method, it quantitative reveals different areas of comprehensive fertility changes computation formula is:

$$IFI = \sum (W_i \times I_i)$$

Among them, for soil fertility index IFI comprehensive, Wi said each single evaluation indexes weights, Ii for each single evaluation index level, so, the size of the values IFI said the comprehensive soil fertility levels. And according to the change of the value range, IFI will soil fertility is divided into four levels (table 7), and then based on the above analysis and calculation get soil fertility grading figure (figure 5).

Table 7. Soil fertility level

IFI	≥5	4 – 5	3 – 4	≤3
The soil fertility level range	1	2	3	4
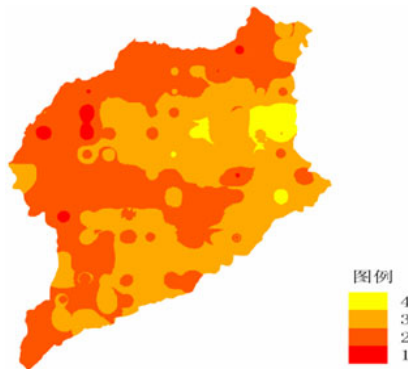
Fig. 5. Soil fertility grading figure

5   Soil Pollution Analysis

At present heavy metal pollution of the cultivated land area accounts for about 1/5 of the total area of arable land in China, every year because of pollution land it cuts food about 10 million Ton , and another 12 million Ton food pollutants exceed standard, both of them the direct economic loss are more than 20 billion RMB. The soil pollution are provided with latent, hysteresis, cumulative, not reversible and difficult management. through the analysis of the data acquisition, data and expert consultation, we think it's scientific and reasonable that the soil pollution as important factor. We use six heavy metals such as mercury, cadmium, lead, arsenic, chromium and copper to show pollution of soil condition in this analysis. According to each of the mean and the maximum metal content and the maximum value of elements, it calculates the heavy metal pollution index (nemero index), formula is:

$$P_g = \sqrt{\frac{(\frac{C_i}{L_i})^2_{最大} + (\frac{C_i}{L_i})^2_{平均}}{2}}$$

Ci is the measured elements of i; Li is the highest tolerances elements of i. (the secondary standard of GB 15618-1995); Ci /Li is elements pollution value of i; （Ci /Li ）最大 is the largest maximum of Ci/Li; （Ci /Li ）平均 is the sum of the average of Ci/Li . According to above analysis calculate comprehensive index, with comprehensive pollution index method pollution level division standard, it divides heavy metal composite pollution values of the division standard grades of the study area in the study area and the heavy metal pollution values (table 8). According to

Table 8. The comprehensive pollution index of soil and classification

Comprehensive soil pollution index	Pollution degree	Pollution levels	Level (score)
0.7 – 1.0	warning limit	Is clean	3
1.0 – 1.4	Light pollution	Pollution than starting pollution value, the crop began to pollution	1
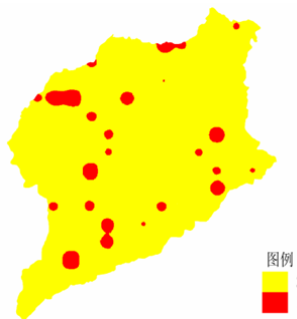
Fig. 6. Soil pollution grading figure

above analysis results, first it constructs " Pg " column in attributes list, second enters each Pg value, then interpolation analysis and heavy classification, last it gains the soil pollution grading graph (figure 6).

6   Land Comprehensive Evaluation Analysis

Thought the above 5 results in this paper, the analytical process gives a certain weight stack according to the local conditions to grow crops on soil types factor, the soil texture factor, the underground water level buried depth factor, soil fertility factor and the soil pollution factor. In the paper it uses three scale two steps analytic hierarchy processes to identify weight

　　The basic principle: the first step is to compare between the two factors to a comparison matrix by three scale; The second step is to use comparison matrix and mathematical formula to obtain structure matrix; The third step is to use consistency check with the matrix, use the important factor of the maximum eigenvalue to represent the corresponding level. The specific procedure is as follows:

　　1、It compares between the two factors for quantitative with the next type, then composes comparative matrix by numerical value (table 9).

$$K_{ij} = \begin{cases} 0(\text{i factor is less important than j factor}) \\ 1(\text{i factor is as important as j factor}) \\ 2(\text{i factor is more important than j factor}) \end{cases}$$

Kij is the comparison quantitative value of factor "i" and factor "j", also it's the relevant elements of the matrix. If i = j, K11, K22,----, K66, it means the factor compare themselves. Diagonal of comparative matrix the number is 1. If the comparison matrix appear 0, 1, 2 in the triangle then other value is 2, 1 and 0 in the corresponding position under triangle in table 9.

Table 9. Comparative matrix

Each factor	K1	K2	K3	K4	K5	$\sum\limits_{i=1}^{5} Ki$
Soil types K1	1	0	2	2	0	5
Soil texture K2	2	1	2	2	0	7
Buried deep underground water level K3	0	0	1	2	0	3
Soil pollution K4	0	0	0	1	0	1
soil fertility K5	2	2	2	2	1	9

2、 The second step that by comparison matrix stack value ($\sum_{i=1}^{5} Ki$), press type can obtain hierarchical analysis, calculation of judgments matrix structure (table 10).

$$R_{ij} = \begin{cases} (K_i - K_j)*(B_m - 1)/(K_{max} - K_{min})+1(K_i > K_j) \\ 1/[(K_i - K_j) * (1 - B_m)/(K_{max} - K_{min})+1](K_i < K_j) \end{cases}$$

Rij is the table structure in the judgment matrix element, Ki, Kj respectively represent each table corresponding factor accumulate value ($\sum_{i=1}^{5} Ki$), Kmax, Kmin are the maximum and the minimum of $\sum_{i=1}^{5} Ki$; Bm generally instead by Kmax, Kmin, means the sum of the most important factor and the least important factor related coefficient, then in the paper Bm=9+1=10; So we can get tectonic judgment matrix (table 10). Then it uses matlab software to solve the maximum eigenvalue λmax =5.2855, and（5.2855-5）/（5-1）=0.071375<0.1, it can pass the consistency check, so the maximum eigenvalue vector matrix: W=[0.2034, 0.4286, 0.0961, 0.04886, 0.8737]

Then, it gets the weight value factor by unification (table 11).

Table 10. Structure judgment matrix

factor	R1	R2	R3	R4	R5
Soil types R1	1.0000	0.3077	3.2500	5.5000	0.1818
Soil texture R2	3.2500	1.0000	5.5000	7.7500	0.3077
Buried deep underground water level R3	0.3077	0.1818	1.0000	3.2500	0.1290
Soil pollution R4	0.1818	0.1290	0.3077	1.0000	0.1000
soil fertility R5	5.5000	3.2500	7.7500	10.0000	1.0000

Table 11. Each factor weight

factor	K1	K2	K3	K4	K5
weights	0.1232	0.2597	0.0582	0.0296	0.5293

Through the above analysis, the soil type distribution, soil texture map, groundwater depth grading figure, soil pollution grading figure and soil fertility hierarchical graph empowerment overlap add (Table 11), we can get the land comprehensive evaluation distribution (Fig. 7).

图例
X < 1.0
1.0 <= X < 4.0
4.0 <= X
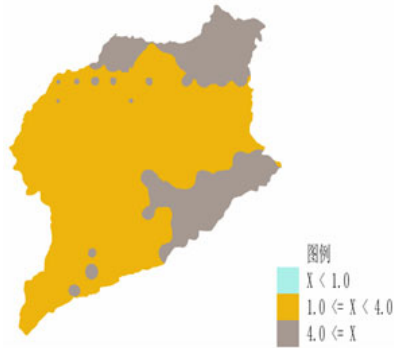
Fig. 7. Land comprehensive evaluation distribution4 conclusion

7   Conclusion

Based on GIS land suitability assessment principle and method, it used supermap to analyzing suitably crops of Liao city land, than we get the following:

1、 With united analyzing respectively of rivers and soil type distribution map, groundwater depth buried depth distribution, it can be found soil salinity is most serious, groundwater depth is the biggest in GuanXian. But five big rivers through Liao city are not through GuanXian (WeiHe and Maxia river are all in the edge) and some of the main canal perennial dry, therefore, in the region we should increase the water resources and avoid the soil salinity become heavier. Also we should adjust the agricultural structure and change the agricultural development model.

2、 Liao city agricultural land quality gradually reduce from the east to the west, therefore, it should pay attention to adjusting agricultural layout. In the east we must give full play to the rich soil water resources and fertile soil advantages to improve crop quality and output. In the west land we can adjust measures to local conditions reform land conditions to develop forestry, orchards and other production.

3、 The soil heavy metal pollution reach light pollution levels in some areas, the crops has caused the damage in a certain degree in the region. Therefore, some polluting factories should be banned that prevent soil pollution become worse.

It is concluded that the present situation and the existence problems in Liao city crop land suitability assessment process of agricultural land. It shows that GIS is an effective and operational technology in complex evaluation process to the crops suitability.

References

1. Pan, J., Fang, S., Sun, W., et al.: Use soil information system for rice nitrogen decisions in Dafeng city in Jiangsu province as an example. Journal of Nanjing Agricultural University (3), 53–56 (2000)
2. Meng, Z., Zhao, C., Liu, H., et al.: The variable fertilization operating systems design and implementation based on prescription figure. Journal of Jiangsu university (natural science edition) 30(4), 338–342 (2009)
3. Sun, Z., Yu, F., Zheng, C., et al.: The GPS variables fertilization system to CPLD control modules design. Farm Machinery Research (3), 83–86 (2008)
4. Li, A., Wang, X., Wang, Z.Q., et al.: Study on electro-hydraulic drive type variable fertilization closed-loop control system. China Agriculture Bulletin 25(7), 272–275 (2009)
5. Li, S., Qin, L.: Variable precision agriculture fertilizer technology and its research progress. World Agriculture 335(3) (2007)

Research on WSN Channel Fading Model
and Experimental Analysis in Orchard Environment

Wei Zhang[1,2], Yong He[1], Fei Liu[1], Congcong Miao[2],
Shitao Sun[2], Chengfeng Liu[2], and Jianfang Jin[3]

[1] College of Biosystems Engineering and Food Science,
Zhejiang University, Hangzhou, China, 200082
[2] Mechanical & Electrical Engineering College,
Jiaxing University, Jiaxing, China, 314001
[3] Service Center of Agricultural Technology,
Fengqiao Town, Jiaxing, China, 314007
zdwnh@126.com

Abstract. The wireless sensor networks used in large-scale orchard must be able to run in various environments including different heights and densities of fruit trees. Moreover, as the radio channels have different propagation characteristics in different environments, it is important to study the propagation characteristics of wireless channel in orchards environments. This paper studied the propagation characteristics of wireless channel of 2.4GHz and 433MHz in orchard with different peach tree heights, respectively, and the channel fading and PLR (Packet Loss Rate) were tested. After that, the transmission range and PL (Path Loss) of wireless channel with different antenna heights were analyzed, and the communication channel fading model was also established. It can be observed that antenna height influence on the frequency of 433MHz is slightly larger, but the initial path loss of 2.4GHz wireless channel is greater. And it also can be found that the antenna height of 3.5 meters is better.

Keywords: Wireless sensor networks (WSN), channel fading model, precision agriculture, orchard.

1   Introduction

Favorable orchard environment is the basis for fruit production. A continuous monitoring on the orchard environment and the growth information of fruit trees was an important technical method to improve agricultural land productivity and resource utilization efficiency, to increase crop production management level and to develop modern agriculture [1-4]. Using WSN technology to solve related issues has become a hot research focus currently [5, 6]. The research on the propagation characteristics of wireless channel provides important guidance for its application in orchard environment [7].

2   Principles and Methods

2.1   Path Loss Model

By subtracting received power (in dBm) from transmitted power (in dBm), the path loss in dB is resulted.

Radio signal meet the formula in propagation:

$$PR=PT / dn . \tag{1}$$

Where d is the distance between the sending and receiving node; n is propagation factor, whose value depends on the propagation environment of wireless signal.

Taking the logarithm on both sides of the formula:

$$10nlgd=10lg(PT / PR). \tag{2}$$

Using least-squares fitting the measured power attenuation data in Matlab based on the approximate logarithm distance path loss model, we found the data met Model (3):

$$PL = k+10nlgd. \tag{3}$$

Where n is the power decay index; K is the model parameters.

Considering that the antenna height h had a greater link with loss index n, Formula (3) could be written as:

$$PL= k+10n（h）*lgd. \tag{4}$$

Parameter k and n determined the relationship between received signal strength and signal transmission distance.

2.2   Experimental Device

The experiments selected CC2430 nodes and 1100SE RF module(CC2430 sends signal of 2.4GHz while RF1100SE sends 433MHz), and DS18B20 Digital Temperature Sensor covered by stainless steel was connected to send real-time temperature information to the receiving node. The transmitting node and receiving node were required to be consistent to ensure that no additional loss occurred to the system. Moreover, the node path loss less demanding, it could be powered by a 3V battery. The following is a brief description of the node.

CC2430 is TI's new ZigBee wireless RF module. It can work in the 2.4 GHz band, requiring low voltage (2.0-3.6 V) power supply with low power consumption (27mA when receiving data and 25mA when sending). Its sensitivity reaches up to -90dBm, and its maximum output is +0.6dBm at the maximum transfer rate of 250 kbps.

The micro-power wireless transmission module RF1100SE adopts high performance radio chip CC1100. Its maximum transmission rate can be up to 500kbps, and the baud rate can be modified using software. In addition, it can be wake up with wireless. With sensitivity up to-110dBm and high reliability, it is widely used in various short-range wireless communications.

2.3   Experimental Methods

Antenna of transmitting and receiving nodes must be consistent, so that the antenna gain would no be generated. The nodes send real-time temperature information by the temperature sensor DS18B20, then sent out with wireless signal through the antenna, monitor the changes of the wireless signal with wirelessmon[6] software on a real-time basis. This software can monitor the received radio signal strength (RSSI), show receive rate(can easily figure out packet loss rate), draw graphics and record data.

When deploying sensor nodes in the orchard, T – R distance d and the antenna height h (distance between receiving and sending node and the antenna height) are two important factors leading to large-scale channel fading [8]. Through experiments in the orchard, we analyzed the attenuation of energy under different antenna heights, different distances between receiving and sending nodes and different wireless channels, respectively. Given the fact that the peach trees were measured to be about 2.5m high on average, we selected antenna heights (above ground) of 1.5, 2.5, 3.5 and 4.5m, respectively; chose 2.4GHz and 433MHz wireless channels; unified transmit power as 0.6dbm, which was easier for the comparison between two types of nodes. And the receive range was chosen to be 10, 20, 30m…, until the received signal was very weak, which was the case with packet loss rate of 1. The PLR (packet lost rate) was defined as the ratio of the loss of data frame bytes to the number of complete data frame bytes. If the data frame received by receiving node was not complete, then there was packet loss.

3   Results and Discussions

3.1   Path Loss Model

Table 1 and Table 2 show the path loss parameters in the band of 433MHz and 2.4GHz, respectively. From the table it can be found that path loss index increased as the antenna height decreased, while the model parameter k appeared in two situations: as the antenna height increased, 2.4GHz wireless channel of the path loss index became smaller, but that for 433MHz gradually became larger. This raised the first doubt in the experiment. If one considered the causes such the differences between the nodes in terms of sensitivity and frequency, and discarding the zero, so the initial path loss is not the same.

Table 1. Parameters in path loss model in 2.4GHz

The antenna height(h)	1.5	2.5	3.5	4.5
n	2.6667	2.2234	2.1038	2.0857
k	46.6667	43.0711	41.7534	39.6255

Table 2. Parameters in path loss model in 433MHz

The antenna height(h)	1.5	2.5	3.5	4.5
n	3.6504	3.1299	2.1139	2.0206
k	21.6215	23.5031	25.0325	25.3814

3.2   The Effects of Different Antenna Heights on the 2.4GHz Band

Fig.1 and Fig.2 show the curve of the 2.4GHz wireless channel with the four different antenna heights, the relation between the transceiver distance and the path loss and the packet loss rate, respectively. It is more clearly seen from Fig.1 that the height of the antenna had a greater impact on path loss, and the different antenna heights in the decreasing order of the amount of path loss were: h = 1.5m> h = 2.5m> h = 3.5m> h = 4.5m. The packet loss rate curve in Fig.2 also confirms that when the packet loss rates
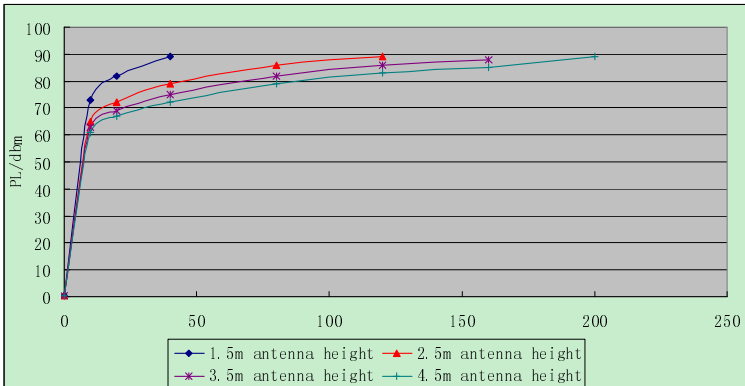
Fig. 1. Transceiver distance and path loss in different antenna heights curves
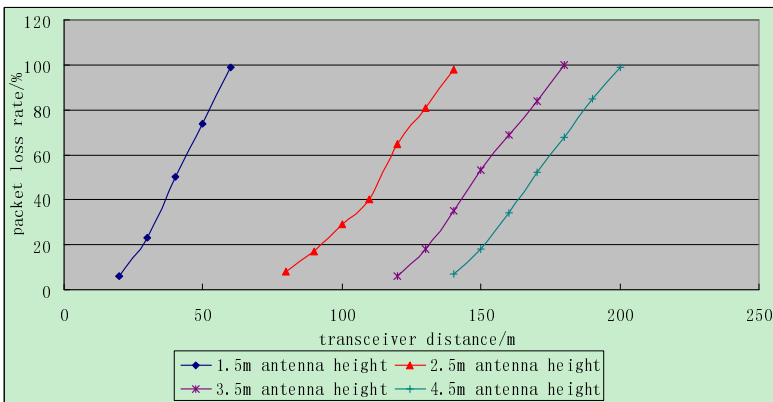
Fig. 2. Transceiver distance and packet loss rate in different antenna heights curves

at different antenna heights were 100%, the different antenna heights in an increasing order of transceiver distance were: h = 1.5m (60m) <h = 2.5m (140m) <h = 3.5 m (180m) <h = 4.5m (200m), the bracket were the transceiver distance. In addition, one can find that under the condition that the antenna height was 3.5m and 4.5m, the channel fading and loss rates were relatively similar. It could be concluded from the experiment that the antenna height should be set as 3.5m in 2.4GHz band.

3.3   The Effects of Different Antenna Heights on 433MHz Channel

Fig.3 and Fig.4 show the curves of 433MHz wireless channel with the four different antenna heights, respectively, the relation between the transceiver distance and the path loss and the packet loss rate. It is clear from Fig.3 that the height of the antenna had a greater impact on path loss, and the different antenna heights in the decreasing order of amount of path loss are: h = 1.5m> h = 2.5m> h = 3.5m> h = 4.5m. The packet loss rate curves in Fig.4 also confirmed that when the packet loss rated for four different antenna heights were 100%, the four different antenna heights in the increasing order of transceiver distance were: h=1.5m（190m）<h=2.5m（280m）< h=3.5m（360m）< h=4.5m（370m）, the bracket were the transceiver distance. Nodes can be found in the positioning of distance, because the frequency of 433MHz was smaller. At the same time it also had the advantages of less loss of communication channel, and greater penetration. It could be also found that with the antenna height of 3.5m and 4.5m, the situation of the packet loss rate and the path loss were more similar. Thus it could be concluded from the experiment above that the antenna height should be set as 3.5m for 433MHz channel.
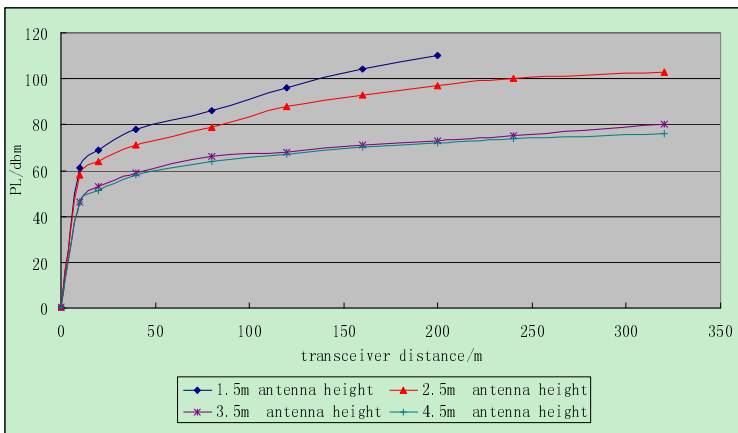
Fig. 3. Transceiver distance and path loss in different antenna heights curves
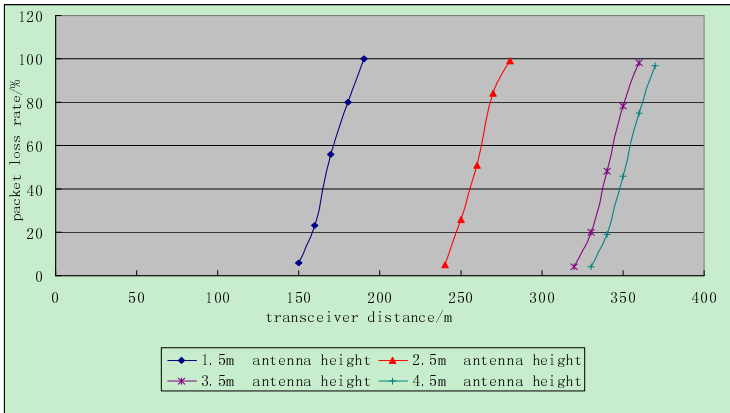
Fig. 4. Transceiver distance and packet loss rate in different antenna heights curves

3.4   Comparison between Different Channels

Fig. 5 and Fig.6 are based on the previous two experiments. The curves were plotted at the most suitable antenna height of 3.5m, the path loss and packet loss rate between different frequencies (2.4GHz, 433MHz).

This section described the differences between 2.4GHz and 433MHz namely: the different node receiver sensitivities, -90dbm for 2.4GHz while -110dbm for 433MHz. In Fig.5, the 2.4GHz's energy quickly reached to 63dbm, then gradually slow down the rate of decay, and ultimately rounded up at 90dbm; for 433MHz, reach to 50dbm quickly, began to be more slow rate of decay, and finally tend to 110dbm.

Fig. 6 shows that the 3.5m antenna height could be positioned under the two frequency ranges. For 2.4GHz, the packet loss occurring from the 110m and in the 180m at 100% packet loss rate. For 433MHz, appear to 360m from 320m at the end.
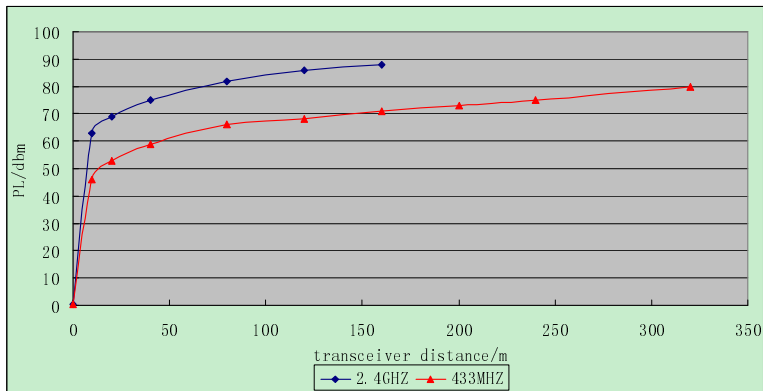
Fig. 5. Curves of two different frequencies and path loss at antenna height of 3.5m

And packet loss rate curve was steeper, after pass the receiver sensitivity of the node, the packet loss rate rapidly transition from 0 to 1. It could be known from the above experiment summary, it can give the result that lower path loss has been 433MHz, positioning a wider range, and a more suitable for the complex environment of peach orchards.
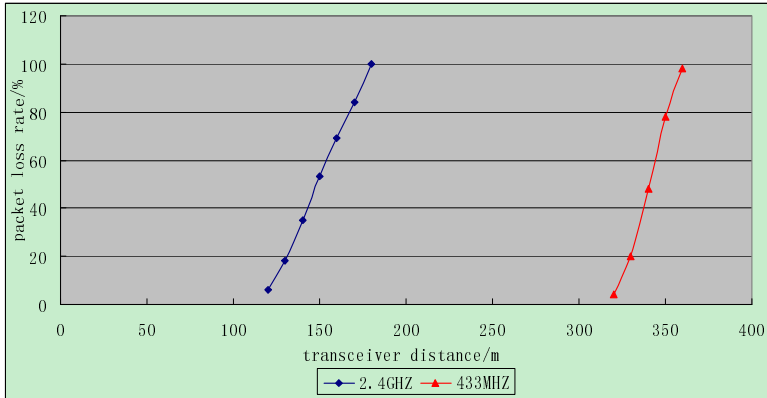
Fig. 6. Curves of two different frequencies and their packet loss rate at antenna height of 3.5m

4   Conclusions

4.1   The Impact of the Antenna Height

Based on the above experimental results and analysis of Fig.1 and Fig.3, it could be found that influence of antenna height on 433MHz wireless channel was slightly larger, but the initial path loss of 2.4GHz wireless channel was greater.

4.2   The Suitable Antenna Height

Through experimental results, it could be found that the antenna height of 3.5 meters was better. While the path loss and the packet loss rate at the height antenna of 4.5m was smaller in the same distance, there was little difference at the antenna height of 3.5m and 4.5m. The higher node height itself would result in the increased costs and higher layout requirements. Therefore, the optimal antenna height was set as 3.5m.

4.3   Wireless Channel Frequency Selection

From the comparison between different channels, it could be found that the optimal radio channel frequency was 433 MHz. At the same distance and the same antenna height, the path loss and the packet loss rate of the channel were relatively lower than those of 2.4GHz wireless channel.

4.4   Suggestions on Wireless Node Localization

When the antenna height was 3.5m, wireless channel frequency was chosen as 433MHz, and the optimal path loss model of the system was PL=21.139lgd+25.0325. This finding can provide guidance for the determination of the position of wireless nodes in large-scale WSN system.

References

[1] Zhang, W., He, Y., Qiu, Z., Cao, J., Qian, S., Lu, J., et al.: The design of precision irrigation system based on wireless sensor network and fuzzy control. Transactions of the CSAE 25(supp. 2), 7–12 (2009)

[2] Gao, F., Yu, L., Zhang, W., Xu, Q., Jiang, Q.: Preliminary study on precision irrigation system based on wireless sensor networks of acoustic emission technique for crop water stress. Transactions of the CASE 24(1), 60–63 (2008)

[3] Fang, X., Zhou, Y., Cheng, W., et al.: The design of wireless intelligent irrigation system based on ZigBee technology. Journal of Agricultural Mechanization Research (1), 114–118 (2009)

[4] Liu, H., Wang, M., Wang, Y., et al.: Development of farmland soil moisture and temperature monitoring system based on wireless sensor network. Journal of Jilin University 38(3), 604–608 (2008)

[5] Li, S.Y., Gao, H.J., Jiang, J.Z.: The effect of antenna height of wheat field on 2.4GHz radio channel propagation characteristics. Journal of Agricultural Engineering 25 (suppl. 2), 184–189 (2009)

[6] Pierce, F.J., Elliott, T.V.: Regional and on-farm wireless sensor networks for agricultural systems in Eastern Washington. Computers and Electronics in Agriculture 61(1), 32–43 (2008)

[7] Gu, J.-J., Chen, S.-C., Zhuang, Y.: Wireless Sensor Network-Based Topology Structures for the Internet of Things Localization. Chinese Journal of Computers (09) (2010)

[8] Kong, X.S., Zhao, D.G., Wang, D.H., Zhang, Z.J.: Analysis of low-channel's impact on wireless sensor networks. Journal of Sensor Technology 24(1), 106–110 (2011)

Analysis of Trace Elements in Leaves Using Laser-Induced Breakdown Spectroscopy

Xu Zhang, Mingyin Yao[*], Muhua Liu, and Zejian Lei

Optics-Electronics Application of Biomaterials Lab, College of Engineering,
Jiangxi Agricultural University, Nanchang, Jiangxi 330045, China
{zx8475396,mingyin800}@126.com

Abstract. Laser-Induced Breakdown Spectroscopy (LIBS) is a new way to analyze the plant ecology. The experimental used a Q-switched Nd:YAG laser to be the laser source and equipped with an eight-channel model spectrometer which's wavelength range between 200 and 1100 nm. Studying the spectrum of the air-drying leaves and the nature leaves and detected the elements which contain Fe, Ca, Na, Mg, K, Cu, Al and Mn. Displaying the list which shows the all spectrum and elements. Refer to Fe as the benchmark, obtain the relative content of trace elements. At the same time, this technology can be employed for food safety and environment pollution evaluation. It will be the based for studying the portable LIBS instrument of detecting the pollution of heavy metal.

Keywords: LIBS, Trace element, Leaves, Air-drying, Relative content.

1 Introduction

Laser-Induced Breakdown Spectroscopy (LIBS) is a new technology which can analyse the constitution and concentration of the matter. It is a laser-based omnipotent molecular and elemental analysis tool. This technology doesn't need to deal with the samples in complicated. And it can anslyse several elements at the same time. LIBS is non-destructive, rapid detection, high sensitivity, on-site and online analysis. It is widely applied in the detection of trace elements in solid, liquid and gas, such as soil[1] alloy steel[2] solution[3] and even in biomedicine[4]. There are some reports about analysing the leaf by laser-induced breakdown spectroscopy. Lidiane[5] and other partners used LIBS to analyse the leaf by neuro-genetic approach. Miloslav[6] and his partners analysed the crop leaves and detected six certified reference materials of leaf tissues by LIBS.

Recently, in order to develop the level of LIBS, there are some experiments include nanosecond, femtosecond, monopulse, dipulse[7-8]. Although LIBS has been made a very great achievement in the analysis of material. But there are still many problems to be solved, such as how to decrease the matrix effect[9] in detecting several elements and how to increase the signal to noise radio.

---

[*] Corresponding author.

In our work, we analyzed trace elements in leaf by LIBS. This works about the analysis of   leaf can help to study the information of tea leaf and tobacco in the future. In this paper, first of all, we introduce the information of sample and the experimental setup. We displayed the results of trace elements analysis and all spectral lines to the corresponding elements table in leaves by LIBS.

2   Experimental Setup

In our experiments, we used a Q-switched Nd:YAG laser (BeamTech, Nimma-200, China) to be the laser source and the experimental setup includes an eight-channel model spectrometer, mirror, DG535, lens, optical fiber(1.5 m length, 400 um core diameter), fiber-optical probe, rotating stage, computer as shown in figure 1.

The fundamental wavelength of   the laser is 1064 nm. And other parameters about laser are 8 ns pulse width, 10 Hz repetition rate. Laser beam was reflected to 45° by the mirror, and through the mirror with a hole focused on the sample which were put on the rotating stage by the lens which's focal length is 200 nm. The plasma was launched and focused on the fiber-optical probe through the lens which's focal length is 100 nm. The probe position is adjustable. We used an eight-channel model AVS-Rackmount-USB2 spectrometer(Avantes, France)   which's wavelength range between 200 and 1100 nm to collect and analyse the plasma emission. The spectrometer has eight wavebands which is 200-317nm, 315-417nm, 415-499nm, 497-565nm, 563-673nm, 671-750nm, 748-931nm, 929-1100nm. The integration time   and the delay time respectively were 2ms and 1.28 us by taking the signal-background   ratio and signal-noise   ratio into account.
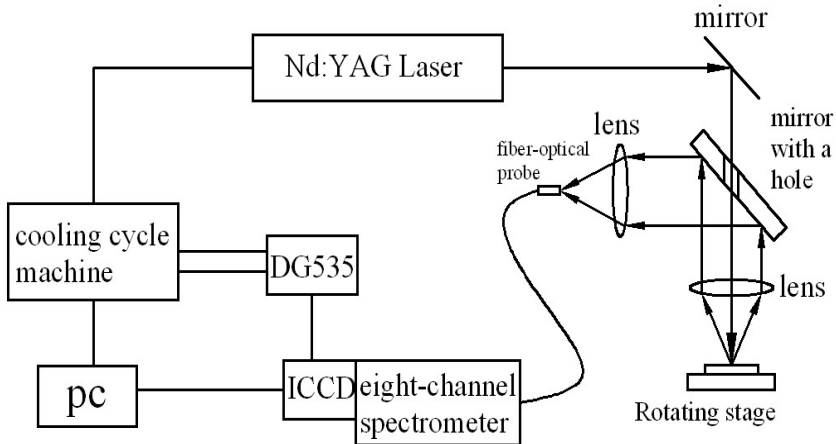
Fig. 1. LIBS experimental setup

3   Experimental Results and Analysis

The leaves came from the ecological orchard of Jiangxi Agricultural University. We picked the leaves in different orange trees. And then we washed the leaves by the

deionized water. The leaves were divided into two parts, a part of all was wiped dry using filter paper, the others was dry by air drying in the air. Because the laser would produce the points which the laser hit on samples into high temperature, we used a rotating stage in order to make the effects of the laser in different points.

The figure 2 shows that a part of the spectra of capsicum in leaves. Refer to the NIST atomic database[10] and the spectral data of the elements from the papers[11-13], we obtain the specific wavelength of trace elements and analysed all peak of waves showing table 1. From the table 1, we can find several elements such as Fe, Ca, Na, Mg, K, Cu, Al, Mn. The spectrum of Fe is most, reaching 61. At the same time, the elements C, H, O, N, Cl were detected by LIBS. But these elements which were detected maybe were in the air. In addition, we detected Mo which was the indispensable element in plants. Because Fe was the most element and spectrum. We choose Fe for the datum to compared with other elements.

As shown in table 2, the intensity of Na is strongest of all. According to the intensity of spectrum being proportional to the concertration, we can obtain the conclusion that content of Na is the highest in leaves. The intensity of the same elements in air-drying leaves is stronger than in nature leaves. It explaines that water in leaves will affect the detection of elements. But the spectral intensity of Al and Na almost is the same between air-drying and nature. The water has a little effect to Al and Na in leaves.
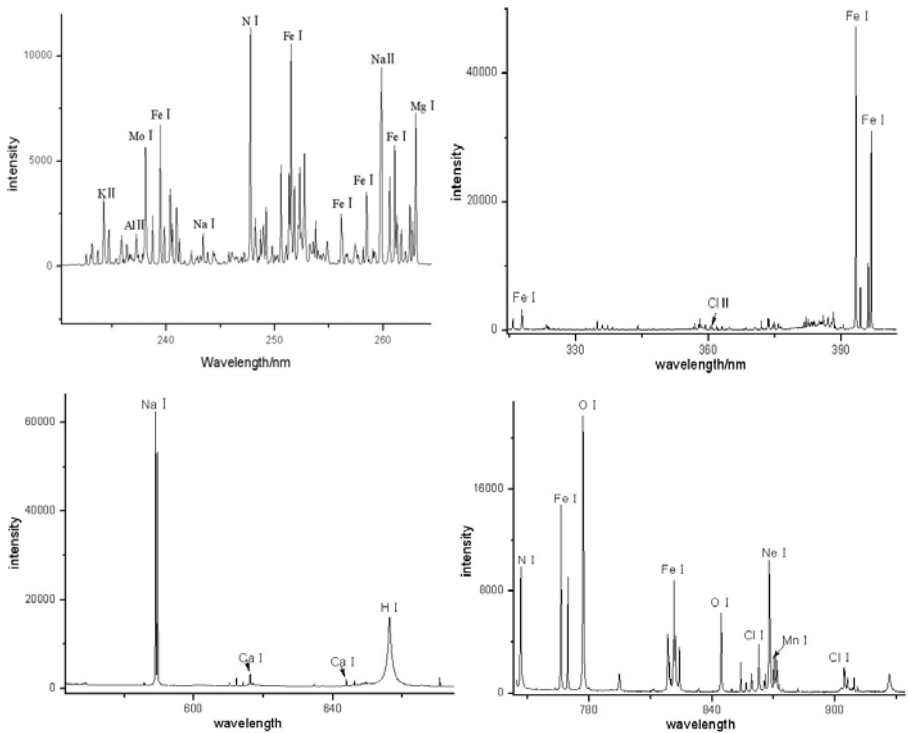
Fig. 2. A part of spectrum in leaves

Table 1. The spectrum of LIBS in leaves

Wavelength/nm	Elements	Wavelength /nm	Elements	Wavelength /nm	Elements	Wavelength /nm	Elements
232.606	O II	253.231	Mo II	381.977	He I	742.361	N I
233.120	Fe II	253.787	Mn II	388.193	Fe I	744.219	N I
233.698	Na II	254.834	C II	393.263	Fe I	746.814	N I
234.212	K II	256.125	Fe I	396.775	Fe I	766.340	Fe I
234.725	C II	257.412	Ne II	422.742	Fe I	769.678	Fe I
235.365	Fe II	258.452	Fe I	438.406	Na II	777.072	O I
235.877	K II	259.854	Na II	445.518	Na II	794.791	O I
236.325	Fe II	260.584	Fe II	498.216	Fe I	818.404	N I
237.218	Al I	261.070	Fe I	517.303	Fe I	821.545	Fe I
238.047	Mo I	262.463	Fe I	518.410	Fe I	824.163	N I
238.747	Mg II	262.705	Fe I	520.880	Fe I	844.481	O I
239.446	Fe I	263.008	Mg I	527.061	Fe I	849.626	Fe I
239.827	Ca I	267.634	Mn I	532.857	Fe I	854.050	Fe I
240.335	Cu II	273.859	Na II	537.208	Fe I	856.662	N II
240.969	Na II	274.564	Fe I	541.031	Fe I	859.256	Fe I
241.222	Fe I	275.502	Fe I	553.604	K II	862.839	Cl I
242.297	Fe II	279.463	Fe I	558.914	Ca I	866.025	N II
243.370	Na I	280.157	Al II	568.837	Na I	867.923	Ne I
244.315	Na II	285.095	Na II	585.729	Ca I	870.169	Mn I
245.760	Fe I	288.046	Fe I	588.977	Na I	871.063	N II
246.074	Na II	301.973	Fe II	589.571	Fe I	871.777	Mg I
247.201	Mo I	308.113	C II	610.303	Fe I	881.872	Fe I
247.764	N II	309.167	Fe I	612.246	Ca I	904.486	Cl I
248.201	Fe I	315.809	Fe I	614.173	Fe I	909.387	Fe I
248.700	Fe I	317.849	Fe I	616.203	Ca I	911.025	O II
248.949	Fe II	334.884	O II	643.895	Ca I	926.633	O I
249.199	Fe I	343.986	Fe I	646.244	Ca I	938.644	Fe I
249.759	Fe II	358.042	Fe I	656.229	H I	940.518	Fe I
250.565	Fe I	359.263	Fe I	670.770	Fe I	945.993	N I
251.064	Fe II	360.478	Cl II	693.816	K I	962.024	C I
251.498	Fe I	361.792	Fe I	714.761	Fe I	965.786	C I
251.808	Fe I	363.048	Ca I	715.673	O I	1011.511	N I
252.304	O II	364.713	Mn I	720.172	Ca I	1053.882	N I
252.737	Fe I	371.923	Mo I	732.544	Ca I		

Table 2. The trace elements' specific wavelengths, average spectral intensity and the ratio

Trace element	Specific wavelength (nm)	Average spectral intensity		The ratio of air-drying to nature	The ratio of element to Fe in Air-drying leaves
		Air-drying	Nature		
Fe	422.743	10147.72	7485.75	1.36	1.00
K	234.212	3225.88	2050.70	1.57	0.32
Al	237.218	1563.05	1525.63	1.02	0.15
Mg	263.008	7307.31	4734.69	1.54	0.72
Na	588.977	62501.57	64445.97	0.97	6.16
Ca	643.895	3152.75	1782.66	1.77	0.31
Mn	870.169	2755.64	1692.33	1.63	0.27

4   Conclusion

In this paper, we used the Laser-Induced Breakdown spectroscopy to detect the trace elements in leaves and obtained the spectrum of Fe, Ca, Na, Mg, K, Cu, Al, Mn, C, H, O, N, Cl. Concertration of Na is the highest. And we compared the air-drying leaves with nature leaves and got the conclusion that the water will affect the detection of elements in leaves. Experiment shows LIBS can quickly analyse the relative content of trace elements in leaves.

References

1. Ferreira, E.C., Milori, D.M.B.P., Ferreira, E.J., Da Silva, R.M., Martin-Neto, L.: Artificial neural network for Cu quantitative determination in soil using a portable laser induced breakdown spectroscopy system. Spectrochimica Acta Part B 63, 1216–1220 (2008)
2. Sun, L.-X., Yu, H.-B., Xin, Y., Cong, Z.-B.: Quantitative Analysis of Mn and Si of Alloy Stseels by Laser-Induced Breakdown Spectroscopy. Spectroscopy and Spectral Analysis 30(12), 3186–3190 (2010)
3. Wu, J.-L., Fu, Y.-X., Li, Y.: Detection of Metal Ions in Water Solution by Laser Induced Breakdown Spectroscopy. Spectroscopy and Spectral Analysis 28(9), 1979–1982 (2008)
4. Liu, X.-Y., Wang, Z.-Y., Hao, L.-Q.: Application of laser induced breakdown spectroscopy technology in biom edicine field. Laser Technology 32(2) (April 2008)
5. Nunes, L.C., da Silva, G.A., Trevizan, L.C.: Simultaneous optimization by neuro-genetic approach for analysis of plant materials by laser induced breakdown spectroscopy. Spectrochimica Acta Part B 64, 565–572 (2009)
6. Pouzar, M., Cernohorsky, T., Prusova, M., Prokopcakova, P., Krejcova, A.: LIBS analysis of crop plants. Journal of Analytical Atomic Spectrometry 24, 953–957 (2009)
7. Samek, O., Lambert, J., Hergenroder, R., Liska, M., Kaiser, J., Novotny, K., Kukhlevsky, S.: Femotosecond laser spectrochemical analysis of plant samples. Laser Physics, Lett. 3(1), 21–25 (2006)

8. Zhang, Q., Xiong, W., Chen, Y.-Q., Li, R.-H.: Rapid Measurement of Trace Mercury in Aqueous Solutions with Optical-Electrical Dual Pulse LIBS Technique. Spectroscopy and Spectral Analysis 31(2), 521–524 (2011)
9. Xie, C.-L., Lu, J.-D., Li, J.: Matrix effect on Laser-Induced Breakdown Spectroscopy of fine coal. Journal of Engineering Thermophysics 29(2) (February 2008)
10. NIST, `http://www.physics.nist.gov/PhysRefData/ASD/lines_form. html`
11. Zhang, D.-C., Ma, X.-W., Zhu, X.-L., Bin, L., Zu, K.-L.: Application of laser-induced breakdown spectroscopy in analyzing micro elements in three kinds of fruit samples. Acta Physica Sinica 57(10) (October 2008)
12. Zhang, D.-C., Ma, X.-W., Zhu, X.-L., Li, B., Zu, K.-L.: Microelements in Potato and Lily Samples Studied by Laser-Induced Breakdown spectroscopy Technology. Spectroscopy and Spectral Analysis 29(5), 1189–1192 (2009)
13. Sun, D.-X., Su, M.-G., Dong, C.Z.: Quantitative analysis of element concentration in Al alloy by using laser-induced breakdown spectroscopy. Acta Physica Sinica 59(7) (July 2010)

Land Use/Land Cover Classification
Based on Multi-resolution Remote Sensing Data[*]

Yuechen Liu, Zhiyuan Pei[**], Quan Wu, Lin Guo, Hu Zhao, and Xiwei Chen

Chinese Academy of Agricultural Engineering,
MaiZidian Street 41, Chaoyang District, Beijing 100125, China
liuyuechen@agri.gov.cn, peizhiyuan@tom.com

Abstract. The paper summarized pre-existing research works relating to land use/land cover classification based on multi-resolution remote sensing data. According to the features of regions, we carried out of the land use/land cover classification of level Ⅲ classes in 148 group of Xinjiang agricultural reclamation eighth division. The land use/land cover classification system divided land in study area into 6 level Ⅰ classes, 16 level Ⅱ classes, and 22 level Ⅲ classes with multi-spatial-resolution remote sensing data. Thus we set up a set of land use/ land cover remote sensing classification and corresponding code system.

Keywords: remote sensing classification, land use/land cover, classification system, code system.

1 Introduction

The land resource is playing an important role in production development for both nation and region. The reasonable development and protection of land resource have became key issue for human to explore. The accuracy and timely update of land use/land cover classification are be of great significance to global change, environmental monitoring, yield estimation et al. Monitoring land use / land cover via remote sensing imagery has the advantage of macroscopic, fast, real-time, characteristics and so on, specially over a large area. Using remote sensing data to obtain and update information may be more effective and accurate. Using single remote sensing data source to carry on the land use/land cover classification has some limitations, such as the existence accuracy is low, versatility not strong, and so on. With rapidly development of the remote sensing techniques, various kinds of remote sensing datasets appears one after another, which used in land resource investigation with different spatial scale. This research employs the different remote sensing data imagery as a foundation, induces thedomestic and foreign projects and researches, related to land use/land cover classification with remote sensing solution. According

---

to the region features, we carried out of the land use/land cover classification and the code system in Xinjiang agricultural reclamation eighth division, for satisfying the need of project research and instruct the area agricultural production.

2   Overview of Land Use/Land Cover Classification with Remote Sensing

Using remote sensing data for establishing land use/land cover classification system is development tendency. Low spatial resolution data has the short visit cycle and has big coverage area. Low spatial resolution data is suited to big scale research of land use/land cover, such as region, intercontinental, country and so on. However, lower spatial resolution will cause the mixed pixel and be difficult to divide features of land. Medium-resolution data were more applied to establish land use/land cover classification, which satisfy the request of database and service structure. Along with the high spatial resolution data being applied, the accuracy of discretion about land types is enhancing greatly. But it has some defects, such as long visit cycle and taking massive storage space. High spatial resolution data was more used in accuracy test and training of supervises classifies in present projects and researches.

2.1   Land Use/Land Cover Classification Based on High Spatial Resolution [1]～[3]

On the high spatial resolution image, the spatial information is richer and it could reflect size, shape of the terrain features and its relations more definitely. It has been already to launch many researches in land use/land cover change with high spatial resolution image. Land classification item in "golden land" project, completed by Sichuan remote sensing center of China, which divided all land types in the study area to be 3 kinds of Ⅰ class land type, 9 kinds of Ⅱ class land type and 21 kinds of Ⅲ class land type. Land use investigation in the LongWan power station finished by Central South University of Forestry and Technology, which divided all land types in the study area to be 3 kinds of Ⅰ class land type, 9 kinds of Ⅱ class land type and 35 kinds of Ⅲ class land type. The project of land investigation of urban land use application in Tanggu district, Tianjin of China, which divided all land types in the study area to be 3 kinds of Ⅰ class land type, 8 kinds of Ⅱ class land type and 23 kinds of Ⅲ class land type.

The land classification in the three projects were used QuickBird data and based on "the Nation Land classification (Implementation)" issued by Ministry of Land and Resources of China. In general, the framework of the three classification system is basically the same, have adopted three-level classification. Regarding the division of water body, three projects carried on were according to its different purposes of use. They all have division of water body in their Ⅲ class land type, corresponding to agricultural land, construction land and unused land in Ⅰ class land type. In addition, compared with the other two projects, the classification system was more targeted and

its coverage types were more comprehensive generated by the project of land investigation in Tanggu district. For example, the division is more detailed for residential land and Independent mining land in its Ⅲ class land type. Urban land was split into towns and cities in the Ⅲ class land type, moreover, added salt pans and special land in it.

2.2   Land Use/Land Cover Classification Based on Medium Spatial Resolution [4]～[8]

At present, medium-resolution data is applied more universal on national scale study of land use. Chinese Academy of Sciences made a land resource classification system based on TM image with the resolution of 30 m，and utilized Spot image with the resolution of 20 m to interpret key areas. It divided all land types in the study area to be 6 kinds of Ⅱ class land type, 25 kinds of Ⅱ class land type. Forest land in the Ⅱ class land type was further divided to 3 types, such as needle forest land, broad-leaf forest land, conifer and broadleaf mixed forest, which were all included in the Ⅲ class land type. In this project, Garden and forest land, mining, transportation and other types of land were merged appropriately, which enhanced operability of remote sensing classification in the Ⅰ level class land type. According to the level of land use/land cover and integrating status of land use investigation, it obtained the land use/land cover types of class Ⅱ.

In the project, "HuangHuaihai agricultural land variation remote sensing investigation for ten years", the Chinese Academy of Agricultural Engineering combining with sample investigation and TM multi-spectrum data, divided land of 7 provinces in HuangHuaihai area to 6 kinds of Ⅰ level class land type and 16 kinds of Ⅱ class land type.

National Land Cover Data，NLCD established by America, which divided all land types in the study area to be 9 kinds of Ⅰ class land type, 37 kind of Ⅱ class land type. The Ⅲ and Ⅳ class land types could get basing on the class Ⅱ for requirements. Land types of class Ⅰ were extracted from TM images with the resolution of 30 m. From Ⅱ to Ⅳ class land types, the land types were obtained from Aerial Images with different height of photography.

2.3   Land Use/Land Cover Classification Based on Low Spatial Resolution [9]～[20]

Global Vegetation Monitoring Unit made use of SPOT／VGT data to execute the theme of Global Land Cover2000. Its classification system was finished combining with classification software of FAO/UNEP-Land Cover Classification System and a classification system basing on data with the resolution of 1 km of International Geosphere-Biosphere Program (IGBP). In this project, land type was divided to 22 kinds on a global scale. The IGBP combining unsupervised classification and NDVI data of 12 months of a year divided the type of land use/land cover for 17 kinds on a

global scale. However, the versatility of the classification system is not strong for it applying only to NOAA-AVHRR data. University of Maryland, using supervised classification, NDVI and the 41 time series images, generated by the 5 channels of NOAA-AVHRR, to produce a global 1km land cover product. Its category types is basically same as IGBP's, when it got rid of the wetland, farmland / natural vegetation mixed, snow and ice covered the three types.

2.4   Land Use/Land Cover Classification Based on Multi Spatial Resolution

Land use/land cover (LULC) based on remote sensing data was a Land use/land cover classification based on multi-resolution, which generated by U.S. Department of Interior Geological Survey（USGS）in 1976. The classification system is divided into four. The Ⅰ level class was divided with LandSat data; for the division, the data obtained at 12400m altitude or more, or less than 1:80000 scale;Ⅲ level class to got data at the orbital altitude 3100～12400m, at 1:20000 scale～1:80000; Ⅳ level class to got data at 3100m orbital altitude or less, or more than 1:20000 scale. This classification system included 9 kinds of Ⅰ level class land type, 35 kinds of Ⅱ level class land type and the Ⅲ, Ⅳ level class can be expanded flexibly, combining with airborne remote sensing data and ground survey and other data. [21]

Comparing to Anderson land classification system, it's more suitable for remote sensing data. The smallest unit of land cover was divided depending on the mapping scale and resolution of remote sensing data and so on. The classifications of Ⅰ and Ⅱ level class adapt to some researches of global or continental, while the classifications of Ⅰ and Ⅱ level class are suitable to some researches of regional or country wide and so on. At the same time, the system followed many features of the Anderson land classification. Such as, the first, its types named simple and easy to be accepted and used; the second, classified information can be converted or accessed any time; the third, categories can be refined downward or integrated upward. However, it took into account both land use and natural ecological background of land in Ⅰ level class of Anderson land classification system, which made the relationship between categories is too complex to distinction. For example, some types were divided according to the situation of land use, such as 1 - urban or building land, 2 - agricultural land, 4 - forest land and so on. While 3 - Mountain, 6 – wetlands and 8 – tundra，which were divided according to ecological background of land resources. [22-24]

3   The Land Use/Land Cover Classification in Study Area

There are some respective flaws, existing in variety of classification systems. On the one hand, the boundaries between the two categories were not clear, so that does not apply to land cover mapping; on the other hand, the formulation of classification system is not conducive to monitoring land cover change [25]. Responding to the above problems, the study develops a more realistic and clear classification system based on existing experience and achievements.

3.1   The Overview of Study Area

Xinjiang agricultural reclamation eighth division located at Tianshan north foothill center-section, south Songorine Basin, which is the counterpart of (Longitude): 84°58′-86°24′ / (Latitude): 43°26′-45°20′. This area is a typical temperate continental climate; it has distinct arid and semi-arid characteristics for its annual rainfall of 180-270 mm and annual evaporation is 1000-1500 mm. There are 334200 ha of agricultural land including 197400 ha of arable land, 2600 ha of garden land, 27100 ha of forest land, 73900 ha of grassland and 26600 ha of construction land and so on. 148 group of Xinjiang agricultural reclamation eighth division is located 80 kilometers west of the city of Shihezi city, which is an important production base of Xinjiang agricultural reclamation eighth division. The148 group is a large state-owned enterprise, dominated by agriculture, while combining with agriculture, forestry, animal husbandry and industry. In this group, cotton, wheat, corn, sugar beet, soybean etc were main agricultural planting. There are less land can be used in the study area, affected by drought, desertification and salinization. Low efficiency of land use, situation of the land use is incompatible with the characteristics of land resources, which is leading to poor performance of land use structure. Soil organic matter content is low and the heavy is low, however, it requires massive investment in the land development and utilization. The land use types with a vertical distribution of zonal,   affected by topography and climate.

3.2   Land Use/Land Cover Classification System in Study Area

3.2.1   The Bases and Principles of Land Use/Land Cover Classification

By comparing and analyzing domestic and foreign classification systems and according to the characteristic of multi-resolution, multi-time phase remote sensing data and features of district, we build the classification system for Xinjiang agricultural reclamation eighth division. The system mainly reflected the following principles: the first, the land type division need reflect consistency or similarity of land use/land cover and natural condition, as well the primary and secondary of Land utilization manner; the second, it is advantageous to the natural resource reasonable use and the protection (for example, Haloxylon, Tamarix and other vegetation with sand-fixing role should be divided into the type of shrub land rather than grass land, even if where has been grown many kinds of thick herbaceous plants below it.); the third, it is necessary to give attention to the types with special significance (for example, vegetation that has the great economic value and the species gene bank function; Special industrial crop and orchard, and so on).

3.2.2   Land Use/Land Cover Type Division

For meeting the requirements of desertification monitoring, sparse scrub and sparse grassland were distinguished from high canopy density of the bush and high coverage grassland by Wang Liwen. And she partitioned the land of Xinjiang into 14 kinds of

Table 1. Table of land use-cover classification system

Ⅰ level class		Ⅱ level class		Ⅲ level class	
Name	Code of land type	Name	Code of land type	Name	Code of land type
Arable Land	1	Paddy Field	11	Paddy Rice	111
		Arid Land	12	Cotton	121
				Corn	123
				Wheat	124
				Tomato	125
				Other Crops	126
		Leisure Land	13	Leisure Land	131
Garden and Forest Land	2	Forest Land	21	Forest Land	211
				Stocked Land and Shrub Land	212
		Garden Land	22	Vineyard	221
				Land for Peach	222
				Other Fields	223
Constructive Land	3	Residential, Industry and Mining Land	31	Residential Land	311
				Industry and Mining Land	312
		Transportation Land	32	Transportation Land	321
Water Body	4	Water Surface	41	Drainage and Ditch	411
				River, Lake, Basin, Pond	412
		Shallow Seas and Tidelands	42		
		Snow Cover	43		
Grass Land	5	Natural Meadow	51	Grass Land	511
				Wild Grass Ground	512
		Artificial Pasture	52	Artificial Pasture	521
Bare Land	6	Bare Field	61	Alkaline Land	611
				Other Naked Land	612
		Sand	62		
		Bare Rock	63		
		Gobi Desert	64		

type. [26]Liu Aixia who divided the land in Shihezi to be 6 kinds of land use type, including planter land for wheat, planter land for cotton, alkaline land, sand and residential land, water area, irrigable land. The classification system was built based on CBERS-1 image data, and full accounted of the local real condition, such as there

are large area of sand and alkaline land and less area of forest land and grass land in the study area; and most of the remaining land is cultivated. [27] Moreover, carries on the classification through the regionalization method regarding the vegetation scarce area, may reduce the influences regarding the precision of land use/land cover classification to a certain extent, because of the local climate, the terrain, the soil and so on is created regarding the land utilization/cover classification precision. Space position knowledge of ground features is benefit to raise classified precision. For example, there are no other types except farmland, water body, garden and forest land, residential land etc; small area bare land will be used by people, distributing around towns; sand with small area can turn other land type through the government and so on. [28]

We divided the land types of land use/land cover and built a classification system with three levels, referring to "Land utilization Present situation Investigation Technical Schedule", "The Nation Land classification (Implementation)", "Remote Sensing Investigation Technology and Method for Medium and Small scale land use change", etc. It is a comprehensive kind of system, which collect land utilization and the cover in a body to instruct agricultural production and agricultural resource management.

3.2.3   Analysis and Application of the Land Use / Land Cover Classification System

3.2.3.1   The Analysis. The Ⅰ level class division By the land cover concept to describe natural quality of biological cover types on land surface. Dividing the Ⅰ level class into 6 kinds of type, from the aspect of cover, such as the arable land, the garden and forest land, the constructive land, the water body, the grass land, the bare land, which combine with multi-time phase and low spatial resolution data, in order to reflect the land cover state in study area.

The Ⅱ level class division Considered the feasibility of remote sensing technique, the interpretation accuracy, the type of land use and the degree of utilization, merged or increased the land type of Ⅱ level class in former classification systems. The new system was divided into 16 kinds of type. (1)Increase the fallow land. Fallow land is frequent and it is defined that stop the cultivation more than one year of the land. Fallow land can be extracted by superposition of different time phase data. (2) Differentiates the field and the forest land. It is easily to pick-up types of the garden land and the forest land from remote sensing image, according to the local planter pattern and certain space position knowledge. (3)The combination of the rivers, the lake, the basin and the pond, may strengthen the feasibility about technical processing, simultaneously also satisfies the actual need in agricultural production. (4)Distinguish the artificial pasture and the natural pasture. The animal husbandry was one of important industrial, so artificial pasture is becoming a more necessary

land type in the study area. The artificial pasture is neater, can satisfied the accurate request of extraction.

The Ⅲ level class division Taken 148 groups as the key research region and considered that land type's use process since long, the agriculture resource protection and the planter management's need, detail classified the types of Ⅲ level class by high resolution remote sensing data into 22 kinds. In the sub-category of the garden land and the forest land, which were all based emphasizing respective function. That is to emphasize its products, application and so on.

3.2.3.2   The Application. Taking the new land use/land cover classification as a foundation, carry on the Ⅰ and Ⅱ level classes division to agricultural reclamation eighth division of Xinjiang and the Ⅲ level class division to 148 group of Xinjiang agricultural reclamation eighth division. Namely, take NDVI time series of MODIS separately as the classified characteristic to various terrain features category degree of membership function for gaining the Ⅰ level classification result; According to the different classified object and choosing the decision tree construction method which based on knowledge and samples, as well as the man-machine interactive way, gain the Ⅱ level classification result with TM data; gain the Ⅲ level classification result with Rapideye data. (Classified result as shown in picture 1, picture 2 and picture 3).
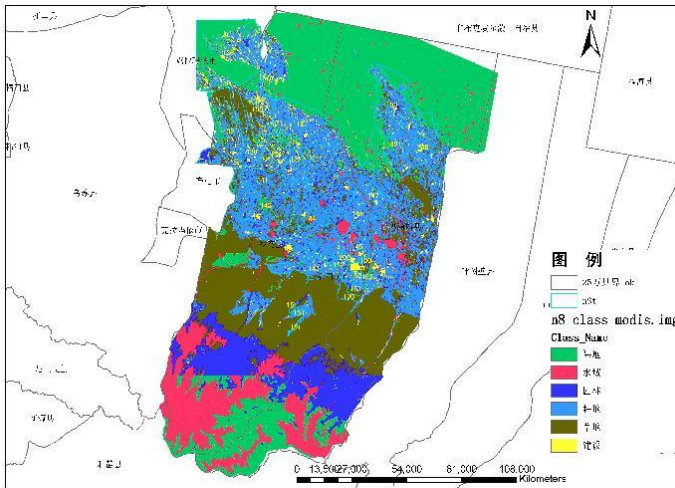
Picture 1. Agricultural reclamation eighth division land use/land cover remote sensing classification result in 2010 (Ⅰ level class)
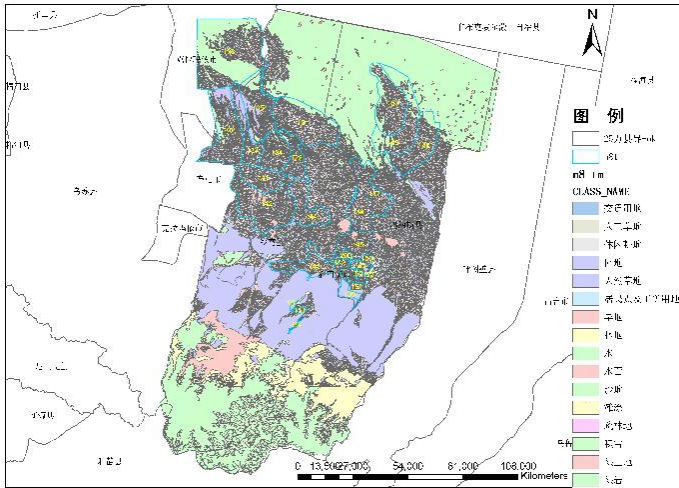
Picture 2. Agricultural reclamation eighth division land use/land cover remote sensing classification result in 2010 （Ⅱ level class）
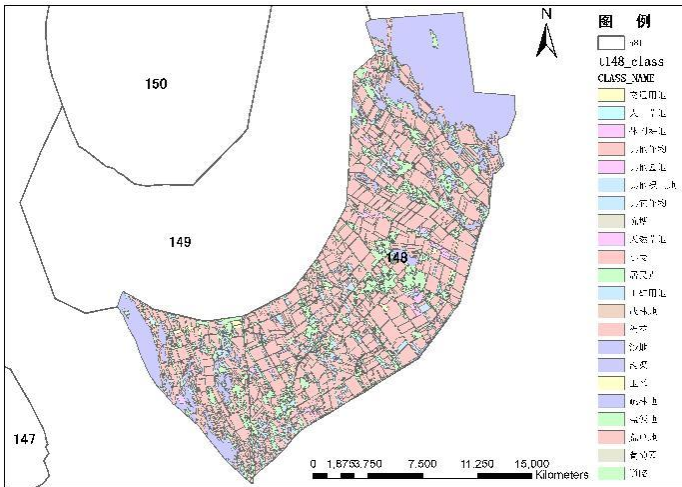
Picture 3. Agricultural reclamation eighth division land use/land cover remote sensing classification result in 2010 （Ⅲ level class）

4   Conclusions

(1) The multi-phase and multi-resolution data can display its superiority fully for land types changed along with season factors, the transition region among different cover types and classification in different scales to establish the classification system.

(2) The land use/land cover classification system was produced, which was based on cover status and land use condition in the study area. In this system, to a certain extent, avoided crossing and confusion categories of status, with clear relationship between classes, been helpful for subprime type of classification.

(3) Considering maneuverability of using remote sensing technology to discriminate land types and cover degree of land use types, reasonable merged or increased in previous classification system of Ⅱ level land type. Namely, divided the new system, needed to achieve the combination similarity and the distinguish diversity, by large to small, by senior to lower. Eventually, built a associated, logical and scientific classification system, facilitate land cover remote sensing monitoring and conventional data sharing between land use status investigation.

(4) The way of classification according to the land use/land cover status, was helped to control and supervision the main quantity about land use types and the implementation of the plans.

References

1. Xiao, X., et al.: The Application of the Quickbird Remote Sensing Images to the Regulation of Kinds of Land Use——By the Example of Anju District, Suining. Acta Geologica Sichuan 28(4), 331–334 (2008)
2. Xing, Y., et al.: Investigation on Land Use Based on QuickBird Image. Journal of Anhui Agricultural Sciences 35(10), 3001–3002, 3004 (2007)
3. Lv, X., et al.: Application Study on QuickBird Image in Investigation of Urban Land Use Status. Tianjin Agricultural Sciences 14(2), 12–15 (2008)
4. Ratanopad, S., Kainz, W.: Land Cover Classification and Monitoring in Northeast Thailand Using LandSat 5 TM Data. International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences 1(2), 137–144 (2006)
5. He, Y., et al.: FAO/UNEP-Land Cover Classification System (LCCS) and Use for Reference. China Land Science 19(6), 45–49 (2005)
6. Xiao, P., et al.: A Land Use/ Cover Classification System Based on Medium Resolution Remote Sensing Data. China Land Science 20(2), 33–38 (2006)
7. Clevers, J., Bartholomeus, H., Mücher, S., de Wit, A.: Land Cover Classification with the Medium Resolution Imaging Spectrometer (MERIS). EARSeL eProceedings 3(3), 354–362 (2004)
8. Xiao, P., et al.: Methods and Accuracy Assessment for Land Use and Cover Classification Based on Medium Resolution Remotely Sensed Data. Remote Sensing for Land & Resources 4, 41–45 (2004)
9. Mücher, C.A., van de Velde, R.J., Nieuwenhuis, G.J.A.: Mapping Land Cover for Environmental Monitoring on a European Scale. Pilot Project for the Applicability of NOAA/AVHRR HRPT data.Report 93, SC-DLO, Wageningen, The Netherlands. RIVM report No. 402001004 (1994)
10. Malmberg, U.: BALANS Land Cover and Land Use Classification Methodology. BALANS report Balans-utv-24 2.0, Stockholm (2001)
11. Agrawal, S., Joshi, P.K., Shukla, Y., Roy, P.S.: Spot-Vegetation multi temporal data for classifying vegetation in South Central Asia. Communicated to Current Science Journal (2003)

12. Anonymous, Forest Cover Assessment in Asia. In: Proceeding of the International Workshop Asian Forest Cover Assessment and Conservation Issues (2002)
13. Lu, L., et al.: The Mapping and Validation of Land Cover in Northwest China from SPOT4-VEGETATION. Journal of Remote Sensing 7(3), 214–220 (2003)
14. Li, J., Guan, J.-D.: The Vegetation Classification of the Return Farmland to Pasture or Forest Region in Shaanxi-Gansu-Ningxia Based on SPOT/Vegetation Data. Agricultural Science & Tchenology 10(5), 179–183 (2009)
15. Song, Y., et al.: Study on Vegetation Cover Change in Northwest China Based on SPOT VEGETATION Data. Journal of Desert Reseatch 27(1), 89–93 (2007)
16. Xu, W., et al.: China Land Cover 2000 Using SPOT VGT S10 Data. Journal of Remote Sensing 9(2), 204–214 (2005)
17. Gong, P., et al.: Land Cover Mapping with MODIS NDVI Time-series Data in Northeastern China. In: The 9th International Symposium on Physical Measurements and Signatures in Remote Sensing, Beijing, China, October 17-19, pp. 12–415 (2005)
18. Liu, A., et al.: Land Cover Classification Based on MODIS Data in Area to the North—west of Beijing. Progress in Geography 25(2), 96–104 (2006)
19. Liu, A., et al.: Method for remote sensing monitoring of desertification based on MODIS and NOAA/AVHRR data. Transactions of the Chinese Society of Agricultural Engineering 23(10), 145–150 (2007)
20. Gong, P., et al.: On the Classification System of Land Cover in China Based on MODIS. Journal of QINGDAO University of Science and Technology (SOCIAL SCIENCES) 23(2), 79–83 (2007)
21. Gong, P., et al.: Progress of the Research on Classification System of Land Vegetation. China Journal of Agricultural Resources and Regional Planning 27(2), 35–40 (2006)
22. Anderson, J.R., Hardy, E., Roach, J., et al.: A land use and land cover classification system for use with remote sensing data. U.S. Geological Survey Profession Paper, Washington, DC (1976)
23. Anderson, J.R.: Toward more effective method of obtaining land use data in geographic research. The Progressional Geographer 13, 15–18 (1961)
24. Yuan, X., et al.: Comparison and Analysis on Land Cover Remote Sensing Monitoring and typical Classification System. Journal of YUNNAN Polytechnic University 15(4), 8–10 (1999)
25. Sha, Z., et al.: Study on the Methods of Deriving the Information of Land Cover in the Arid Areas by Using TM Data. Arid Land Geography 28(1), 59–64 (2005)
26. Wang, L., et al.: Detecting the Areas at Risk of Desertification in XINJIANG Based on MODIS NDVI Imagery. Journal of Infrared and Millimeter Waves 26(6), 456–460 (2007)
27. Liu, A., et al.: Land use Classification in Arid and Semi-Arid Areas Using CBERS-1 Imagery. Journal of the Graduate School of the Chinese Academy of Science 20(3), 334–340 (2003)
28. Li, S.: A Synthetical Study on Land Use/Cover Change and Desertification in Arid and Semiarid Region. Lanzhou University (2006)

Study on Quick Identify of the Brand of Seabuckthorn Juice Based on PCA and SVM[*]

Zhipeng Liu and Shujuan Zhang

College of Engineering,
Shan xi Agricultural University,
Taigu, China
`sxaulzp@126.com, zsujuan@263.net`

Abstract. To achieve the Seabuckthorn juice brand non-destructively, we put forward a fast identify method based on Visible_near infrared reflectance (NIR) spectroscopy. We use the Field Spec 3 spectroradiometer to collect 40 sample spectra data of the three kinds of Seabuckthorn juice separately. the sample data was preprocessed by Using of average smoothing method and multiplicative scatter correction (MSC) method. Then principal component analysis (PCA) was used to process the spectral data after pretreatment. The samples were divided into 90 model samples and 30 prediction samples, the sample of eight modeling data as input variables of the support vector machine (SVM) to build SVM model, and to identification juice brands. 30 unknown samples of the three brands were predicted for classification, and the results showed that the SVM model on the identification seabuckthorn juice brand has achieved a 99.9% accuracy. Therefore, near infrared spectroscopy coupled with principal component analysis and SVM can be quickly and accurately distinguish the brand of seabuckthorn juice.

Keywords: Visible_NIR spectroscopy, Principal component analysis (PCA), Support Vector Machine(SVM), Seabuckthorn juice brand.

1   Introduction

Seabuckthorn (Hippophae Rhamnoides Linn.), also known as vinegar Liu, is   belong to deciduous shrubs or small trees [1]. In China, distribution of genus is the largest and the species is the most. It is Distributed in Shanxi, Shaanxi, and Inner Mongolia, 19 provinces and autonomous regions, the total area is 1800 Ten thousand mu. Seabuckthorn have strong adaptability and vitality, and is resistant to Sand wind, drought, water wet, salt, poor soil, heat and cold. Since the mid-80s of the 20th century, China has become to focus on comprehensive development and processing and utilization about seabuckthorn, the number of research datum and works about Seabuckthor is increasing [2]. In the recent years, With the implement of the grain project, seabuckthorn is

cultivated large areas. Seabuckthorn fruit contains protein, unsaturated fatty acids, vitamins for the whole, and is the valuable medicinal and edible plant resources. Currently, seabuckthorn juice is as a green drink, its internal quality such as taste, sugar content, acidity and vitamin content has received widespread attention. The Brand range is large, and the differences in taste and quality are more significant. So a simple, rapid and nondestructive identification of juice brand and technology are essential.

Near infrared spectroscopy (NIRS) technology developed rapidly in the late 80's last century can achieve non-destructive testing techniques of physical test. Almost the main structure of all the organic composition can be found the signals in near-infrared spectroscopy, the table and obtained easily spectrum, unpollution testing, low-cost detection, so Near Infrared Spectroscopy (NIR) has the reputation of the Giants[3].In the recent years, the application of near infrared spectroscopy technology in many areas has obtained a great development, it has played a significant role in the advancement of scientific research in the field of production and technological progress. Nowdays, using near-infrared spectroscopy, domestic and foreign scholars have identified the brands such as soy sauce[4], bayberry juice[5], soybean[6], white vinegar[7], yogurt[8], milk powder[9], coffee[10], tea[11], melon[12], chilli[13], but no identification on seabuckthorn juice brand.

PCA is a data mining techniques in the multivariate statistic, and is one of the means of cluster analysis. The aim is to select a smaller number of new variables to replace the original more variables under the premise of not losing the main spectral information, it resolve the difficulties which is not be analyzed because of the overlapping bands, and is widely used as a spectrum analysis mathematical method[14]. SVM (Support Vector Machine) approach is based on statistical theory, Huber robust regression and Wolfe duality theory[15][16], its advantage is strong generalization performance, high precision fitting, global optimization. To address prediction and identification in the case of the limited sample, it provided a strong theoretical basis and effective solution. SVM method is based on structural risk minimization principle rather than empirical risk minimization principle, it avoid learning phenomenon in the artificial neural network, has very good generalization performance, can find the global optimal solution, and has many advantages when solving the small sample, nonlinear, high dimension and local minima problems in identifying the performance of such models[17][18].

The ultimate goal of this study is to realize the identification on juice brand, using data mining method combining the visible - near infrared spectroscopy, principal component analysis and support vector machine.

2   Experiment Material

2.1   Experimental Device

The experimental equipment includes notebook computer, spectrometer, halogen light, and correction whiteboard. FieldSpec 3 spectrometer is produced by ASD (Analytical Spectral Device）in America, spectral sampling interval of 1nm, the sampling range of 350～2500nm, scanning frequency of 30 times, Probe field angle of 25°, diffuse reflectance sampling of spectrum, 14.5V halogen light source matching spectrometer.

Spectral data is exported in the form of ASCII code for processing, analysis software includes ASD View Spec Pro V5.0, Unscramble V9.7 and DPS (Data Procession System For Practical Statistics).

2.2   Source of Experimental Samples

Three brands of juice with the same date with the batch production purchased from the market, totaling 120 samples, were randomly divided set of 90 samples for modeling and set of 30 samples for prediction, Xiapusaier sea buckthorn juice, sea buckthorn juice and Wei Shijie Huiyuan juice respectively 40 samples. The juice is filled into the high-1.4cm, 6.5cm diameter of dish, the liquid height 10mm. Then spectrometer sample probe is placed in the top of sea buckthorn juice, sea buckthorn juice from the surface of 50mm, and the surface into a 45 °angle fixed. Sheng-like container is away from light source center 350mm, and the surface into a 45 °angle of illumination.

2.3   Spectral Data Pretreatment

In order to remove the effects from the high-frequency random noise, baseline drift, uneven samples, light scattering, we apply smoothing method for spectrum pretreatment, smooth points to 9.All of the pretreatment process is carried out in the Unscramble V9.7 software for greatly filtering the high-frequency noise generated by various factors, further multiplicative scatter correction MSC(Multiplicative Scatter Correction) processing [11].

3   Results and Analyses

3.1   Spectrum Curve Analysis

The typical curve of near-infrared diffuse reflectance spectra of three varieties of sea buckthorn juic is shown in Figure 1. Horizontal axis in Figure 1 for the wavelength range of 350～2500nm, the vertical axis for the spectral reflectance. Fig.1 shows various species of sea buckthorn juice is different spectrum, with some characteristics and fingerprints. Through application of ASD ViewSpec Pro V5.0 software, spectral reflectance data of samples averaged is measured, and converted into ASCII code to export, then through Unscramble V9.7 software for principal component analysis.

3.2   Main Component Qualitative Cluster Analysis

The purpose of PCA is data reduction, to eliminate the information overlapping part of the co-exist information, by converting lots of original spectral variables to change a smaller number of new variables into a linear combination of original variables, and to make new variables maximize the representation of data structure of the original variable[4][5]. Using Unscramble V9.7 software, we can obtain the PCA clustering of three juice 60 samples(Fig.2,A1:Huiyuan juice A2:Xiapusaier juice A3: Wei Shijie juice).
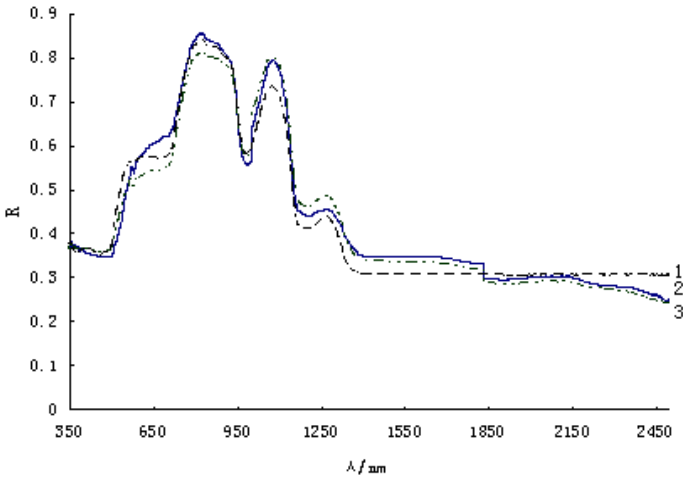
Note:1-Xiapusaier juice spectral curve; 2-Huiyuan juice spectral curve; 3-Wei Shijie juice spectral curve

Fig. 1. Visible_Near infrared reflectance spectroscopy of three different varieties of Seabuckthorn juice
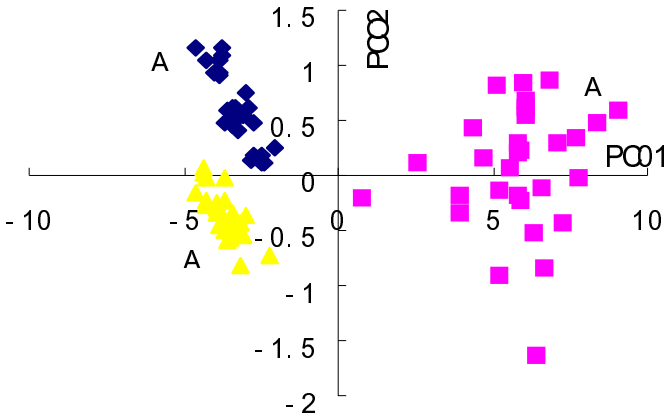
Fig. 2. PCA scores plots (PC1× PC2) for Seabuckthorn juice sample across the entire spectral region

Fig.2 shows PCA scores plots (PC1× PC2) for seabuckthorn juice sample across the entire spectral region, Horizontal axis represents each sample of the first principal component scores, the vertical axis represents each sample value of the second principal component scores. "Xiapusaier juice", "Wei Shijie juice" and "Huiyuan juice" are clearly divided into three categories(Fig.2),to show the main components 1,2 well

effect the three seabuckthorn juice.Fig.2 also shows 20 Huiyuan juice samples with good degree of polymerization is better distributed in the second and third quadrant and aggregate near the X-axis. The degree of polymerization of 20 Wei Shijie juice samples is better, and is mostly distributed in the aggregate in the first quadrant and near the Y axis. The degree of polymerization of Xiapusaier juice is inferior than the first two samples, but also is distributed in the fourth quadrant. The three seabuckthorn juice is little overlaped and can be distinguish. Analysis showed that the first two principal components has some clustering effect on the three kinds of juice, can qualitatively distinguish between different brands of juice, but can not give quantitative identification model. so we consider the establishment of brand identification model based on a number of principal components and support vector machine.

3.3   The Principal Component Analysis Results

Sample spectral band from 350～2500nm total 2150 points, when using the whole spectrum, calculation amount is large, because of weak spectral information of some samples of the region, and lacking of correlation with sample composition or its nature. Table 1 shows the accumulative reliabilities of eight principal components. The accumulative reliabilities shows the explaination extent of the principal component to the original variables. The accumulative reliabilities of the first eight principal components is to 99.99%, it shows that the eight principal components can explain 99.99% of the original wavelength.

Table 1 shows that the principal component analysis is a very effective data mining method, which has compressed more than 2000 wavelength variable into 8 new variable each other perpendicular, the eight new variables do not affect each other, but also are on behalf of the information contained in the most of original variables.

Table 1. Accumulative reliabilities of the first 8 PCs

No.of PC	PC01	PC02	PC03	PC04	PC05	PC06	PC07	PC08
Accumulative reliability /%	97.997	99.315	99.720	99.877	99.965	99.981	99.986	99.992

3.4   Support Vector Machine Model

The first thing is to resolve the kernel function selection, when applicating the identification model of support vector machine, since the effection of predictive power to sample data is different according to applying different kernel function in SVW algorithm. The most common is the radial basis function. In the paper, considering input and output is highly nonlinear in the most forecasts, we select the exponential radial basis function as the SVM kernel function within the product.

The common kernel functions are as follows:

1) Polynomial kernel function:

$$K(x, x_i) = (x \cdot x_i + 1)^d \tag{1}$$

2) Gaussian kernel function:

$$K(x, x_i) = \exp\left[-\frac{\|x - x_i\|^2}{2\sigma^2} \right] \tag{2}$$

3) Sigmoid function:

$$K(x, x_i) = \tanh(v < x, x_i > + c) \tag{3}$$

We make the principal component score of 90 (30 per brand) samples as the training set of the support vector machine algorithm, while the remaining 30 (10 for each brand) sample principal component score as the prediction set of the support vector machine, and classifying juice brand with a class on the remaining class support vector machine method. A class on the remaining class support vector machine algorithm is as follows.

We suppose the training set of known samples (Dn) contains n samples,

$$D = \{(x_i, y_i) | i = 1, 2, \cdots, n\} \in (X \times Y)^n, x_i \in X = R^n, y_i \in Y = \{1, \cdots, M\}, i = 1, \cdots, n \tag{4}$$

The choice of kernel function K(x,xi)determines the structure of feature space, Kernel function value is equal to the inner product of the vector X and the vector Xi in the respective feature space $\Phi(x)$ and $\Phi(x_i)$.

Radial basis function:

$$K(x, x_i) = \exp\left[-\frac{\|x - x_i\|^2}{2\sigma^2} \right] \quad (\sigma \text{ is the parameter}). \tag{5}$$

So we select the appropriate parameters σ according to each output indicator, then fit the sample by SVM, establish the appropriate fitting model $SVM_i (i = 1, 2, \cdots, 30)$ for respectively fitting $y_i (i = 1, 2, \cdots, 30)$ in table 2.

Table 2 shows the support vector machine model predicts unknown samples of sea buckthorn juice with below ±0.1 deviation and 100% recognition ratio, this study establishes the SVM near infrared spectroscopy classification model with more accurate identification, compared with the conventional model of discriminant analysis.

Table 2. SVM model predictions for unknown samples

Sample number	Real value	Predicted value	Bias	Sample number	Real value	Predicted value	Bias
1	1	0.99319	0.00681	16	2	1.95086	0.04914
2	1	1.00822	-0.00822	17	2	1.99182	0.00818
3	1	0.99265	0.00735	18	2	1.98151	0.01849
4	1	0.99086	0.00914	19	2	1.98210	0.01790
5	1	0.99323	0.00677	20	2	1.99062	0.00938
6	1	0.99323	0.00677	21	3	3.01081	-0.01081
7	1	1.00735	-0.00735	22	3	2.88968	0.11032
8	1	0.99332	0.00668	23	3	2.99442	0.00558
9	1	0.99134	0.00866	24	3	2.95287	0.04713
10	1	0.982303	0.017697	25	3	2.96318	0.03682
11	2	2.01765	-0.01765	26	3	2.97322	0.02678
12	2	1.97105	0.02895	27	3	2.99949	0.00051
13	2	1.98735	0.01265	28	3	2.95702	0.04298
14	2	2.00447	-0.00447	29	3	3.01189	-0.01189
15	2	1.96037	0.03963	30	3	3.00159	-0.00159

Note: Variety value 1-Huiyuan juice; 2-Xiapusaier juice; 3-Wei Shijie juice.

4   Conclusion

In this paper, on the base of the spectral data acquisition and analysis of the three kinds of juice on the diffuse reflectance, we establishes brand juice identification model, identifying three seabuckthorn juice brands by combining the principal component analysis with the support vector machine method. The experiments and calculations indicate that the model predicted good results, with below ±0.1 prediction deviation of the unknown juice sample and the identification rate of 100%, the near-infrared spectral identification model established in the paper is more accurate than the conventional model of discriminant analysis. It is suggested that the support vector machines has a good application performance in the near infrared spectroscopy juice brand identification.

References

[1] Ma, J.: Progress in the chemical constituents of sea buckthorn. Heilongjiang Traditional Chinese Medicine 14(3), 208–209 (2001)
[2] Li, F.: Seabuckthorn Research. Research Advances 6(1), 7 (2009)
[3] McClare, W.F.: Anal. Chem. A 66(1), 43 (1994)

[4] Tong, X., Bao, Y., He, Y.: Application of near infrared spectroscopy technique for detection of the brand of soy sauce. Spectroscopy and Spectral Analysis 28(3), 597–601 (2008)

[5] Cen, H., Bao, Y., He, Y.: Bayberry juice based on spectroscopy of rapid identification method for species. Spectroscopy and Spectral Analysis 27(3), 503–506 (2007)

[6] Turza, S., Toth, A., Varadi, M.: In: Near Inf rared Spect roscopy: Proceedings of the International Conference, p. 183. NIR Publications, Chichester (1998)

[7] Wang, L., Liu, F.: Application of visible - near infrared spectroscopy technology white vinegar brands and rapid detection of pH values. Spectroscopy and Spectral Analysis 28(4) (2008)

[8] He, Y., Feng, S., Li, X.: Application of Near Infrared Spectroscopy Identification of species of yogurt. Spectroscopy and Spectral Analysis 26(11), 2021 (2006)

[9] Min, H., Yong, H., Cen, H.: Application of visible - near infrared spectroscopy Fast Discrimination of varieties of infant milk powder. Spectroscopy and Spectral Analysis 27(5), 916 (2007)

[10] Wang, Y., He, Y.: Based on Visible - Near Infrared Identification of the coffee brand. Spectroscopy and Spectral Analysis 27(4) (2007)

[11] Li, X., He, Y., Qiu, Z.: A visible - near Infrared Spectroscopy new method for identification of tea varieties. Spectroscopy and Spectral Analysis 27(2), 279–282 (2007)

[12] Seregely, Z., Deak, T.: Gyorgy Denes Biszt ray. Chemometrics and Intelligent Laboratory Systems 72, 195 (2004)

[13] Yu, W., Yong, K.: Fluorescence characteristics of capsaicin. Food Science 24(11), 105 (2003)

[14] Wang, L., Liu, F.: Application of visible - near infrared spectra for white vinegar brands and rapid detection of pH values. Spectroscopy and Spectral Analysis 28(40), 813 (2008)

[15] Cui, Q., He, Z., Cui, F.: Support Vector Machine Prediction Model of Coke Quality. Chemical Engineering 1(1), 28–31 (2006)

[16] Du, X., Zhang, J.: Sales of agricultural products based on support vector machine. Management Science 4(8), 129–134 (2005)

[17] Deng, N., Tian, Y.: A new method of data mining towels a support vector machine. Science Press, Beijing (2004)

[18] Li, G., Wang, M., Zeng, H.: Introduction to the shoe support machine. Electronic Industry Press, Beijing (2004)

Research and Application
of Space-Time Evolution of Soil Fertility Data Mining
Based on Visualization[*]

Wei Dong, Guifen Chen[**], Jian Jiang, and Guowei Wang

Institute of Information Technology, Jilin Agricultural University, Changchun 130118, China
{dong.wei86,guifchen}@163.com

Abstract. Soil fertility is a comprehensive reflection of soil properties, and its variation has been a central issue in the field of soil science. Most of current research on Variation of the soil fertility are most for a single time and place, and using expertise, simple mathematical model approach that can not effectively simulate the temporal and spatial variation of soil nutrients law. Therefore, this paper presents a spatial fuzzy clustering mining algorithms based on visualization to investigate the law of spatial and temporal evolution of soil fertility mining.

Keywords: soil fertility, spatial fuzzy clustering, temporal GIS, visualization, space-time evolution.

1 Introduction

Soil fertility is a comprehensive reflection of soil properties, it is changing in a vary period of time and space, it can fully reflect the natural factors and impacts of human activities on soil, and plays an important role on crops growth[1]. However, the interaction of the water, fertilizer, gas, heat and urbanization, environmental pollution and other man-made factors constitute complex space objects and relationships between the soil fertility itself and the corresponding the environment that hinders people's precise control to the soil spatial and temporal evolution fertility comprehensive, systematic analysis and intuitive [2-3]. Therefore, investigate the temporal and spatial variation of soil fertility is significance on the division of the farm management area, the implementation of variable rate fertilization and other precision agriculture technology.

Over the years, there are a lot of researches on soil fertility. Yu Yang etc[4] conducted a typical analysis about spatial variation of soil nutrient factors on black

soil areas in Northeast ; Zhou Xu etc[5] proposed a approach of " based on GIS and AHP evaluation of arable soil fertility fuzzy", and has been applied in Pu an County, Gui zhou Province .Liling etc[6]take the long-term experiment in 1980-2006 which made by Hei long jiang Academy of Agricultural Sciences as a platform to compare the affect of different fertilization system on soil fertility of black soil region; Zougui Mei etc[7] suggests that urbanization is not only changes the land and geographical space, but also changes the soil fertility. Most of these researchers above focus on the soil itself, or only on single time or space relationship, mainly using the expertise or the simple mathematical model of data mining and conventional methods can not meet the vast amounts of spatial data, and then causes an inefficient procedure and large bias result. Comprehensive consideration of various factors study is very rare, and can not generate an accurate and effective simulation of space-time evolution of the law of soil fertility [8-9].

Thus, this paper integrates data mining, temporal GIS and visualization technology to study the evolution of soil fertility for many years, and improve the accuracy and efficiency of data mining and provide a reliable basis for the effective analysis and evaluation of soil fertility and fertilization of corn accurater decision- making. The result has been applied in the implementation of the national "863" project, "corn precise operating system research and application", the variable rate fertilization improves spatial diversity of soil fertility and enhances the ability of the agricultural management for agricultural production capacity which receives significant economic and social benefits.

2   Materials and Methods

2.1   Study Area

The experimental site is the national "863" model town , the No.3 13th District, Gong peng Village Yu shu City, locates northwest of Yu shu City and 26km from the city, in 126.315738-126.317017 longitude and latitudes between the 44.999859-45.002761. As a semi-humid temperature mild climate zone, the annual average temperature is 4℃, frost-free period has 135 days or so, the average annual precipitation is 500-700 mm. Plots of the total area of approximately 375 acres, the main crops are maize.

2.2   Data Acquisition and Processing

In this study, using GPS, GIS, RS technologies to obtain information of experimental corn cropland. Useing DGPS to divide the cropland into 40m×40m grid cell, A1~L10 as the sampling points. In this grid sampling unit, sampling depth is 25 cm, the sampling method is five-point plum sampling approach, which takes the four corners and centers of the grid on the soil samples mixture as the grid soil samples [10], shows in Figure 1. The experiment collects maize soil organic matter, available nitrogen, available phosphorus and potassium and other nutrient content data, production data and other related information, laboratory tests on soil samples to obtain a number of attribute data for analysis then acquire on the field of knowledge.
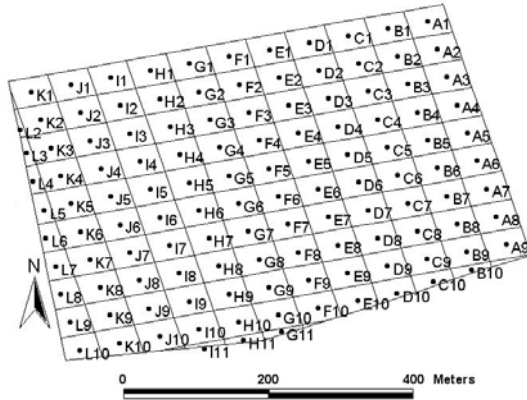
Fig. 1. Fertilization trellis

2.3   Method Description

2.3.1   Spatial Interpolation and VTK

Spatial interpolation is the procedure to estimate the value of other points from the known points. Commonly used spatial interpolation methods include Thiessen polygon method, Kriging method, the triangle method, Sibson natural neighbor method, inverse distance function method and so on. Kriging method becomes the preferred method of spatial interpolation, because of its compatible on integrity and locality[11]. In this paper, using the Kriging method to map the spatial variation of soil nutrient. The result can be displayed by the visualization toolkit VTK(Visualization Toolkit) as three-dimensional visualization.

VTK is an open source software system that includes computer graphics, image processing and visualization all in one and provides strong support to visual development tool for the study [12]. VTK uses pipeline mechanism to achieve visual process, and according to obtain the original data and the result to display to choose the appropriate algorithm . Source is the beginning of line by reading the source data files, etc. to produce source data. Filter handles a number of input data, and produces a number of output data and transform data various. Mapper is the interface between the visualization pipeline and the graphics model that converts processed Filter data into graphics data . Actor is the entity in the display window to receive the data attributes delivered from Mapper. Then Renderer window shows the results come out in[13-14].

2.3.2   Algorithm for Mining Fuzzy Weighted Space

Weighted spatial fuzzy clustering analysis includes determining the weight coefficient and the establishment of the weighted fuzzy clustering model two aspects [15]. Main steps as shown in Figure 2.
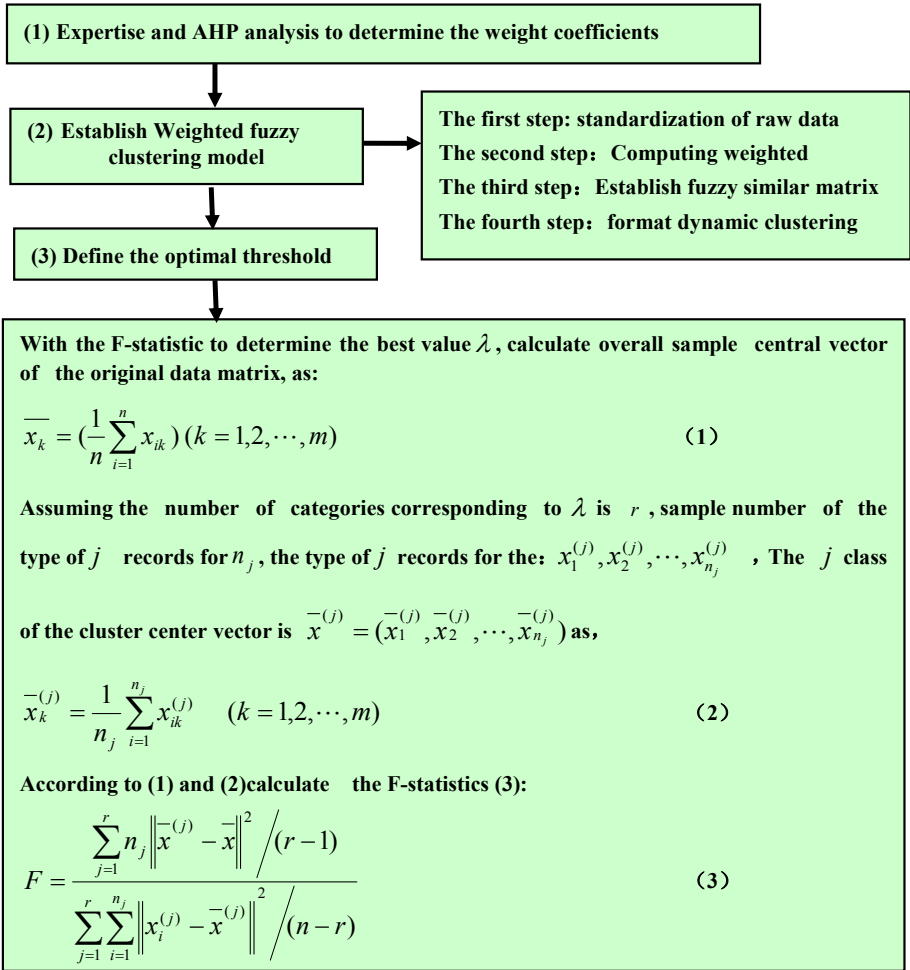
Fig. 2. Flow chart of the weighted fuzzy clustering

2.3.3 Temporal GIS Technology

Temporal GIS (T-GIS) was first proposed by Thrift in 1977, it indicates that time can be regarded as an external independent dimension apart from two-dimensional or three-dimensional space. Temporal GIS is also known as 4D-GIS, which means the three-Dimensional pace plus time dimension. Since the 1990s, With the further development of GIS applications, cadastral management, resource management, environmental monitoring, underground engineering, geology, mining, marine and

other applications, temporal GIS has become a central research issue [16]. Temporal GIS technology can truly effectively manage the time of spatial data(reconstruction of historical status, track changes, predict the future, etc.), display the evolution of temporal and spatial data's variation , implementation historical and trend analysis of the different periods to achieve the purpose of dynamic monitoring.

3   Results and Analysis

3.1   Three-Dimensional Spatial Variation of Soil Nutrient Spatial Map

Takes soil available phosphorus of experimental base as an example three-dimensional spatial variability of soil nutrient map generation process.

(1) Convert the attribute information of soil available phosphorus in 2003 and 2008 into spatial information, using the Kriging method of interpolation them to form a raster layer. Then employ the VTK system to calculate the three-dimensional space within the area of each point of the element values with Kriging interpolation as VTK points data source, using VTKDelaunay2D function to generate DEM, which is VTK Unstructure Grid, use VTKdemreader to read the DEM, using VTKImage Shrink3D render three-dimensional effects and VTKElevationFilter set the output of z-axis, until adopted rule voxel model VTKPoly-Data Mapper that you can draw three-dimensional spatial variability of soil available P (Figure 3, Figure 4).
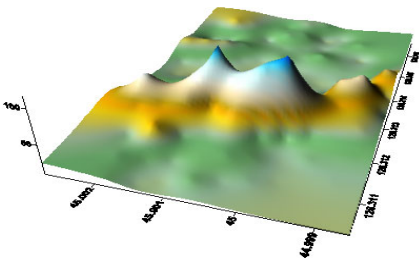
Fig. 3. Three-dimensional spatial
variability map of Available P 2003
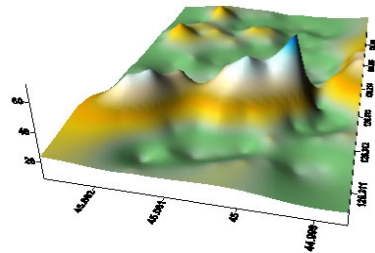
Fig. 4. Three-dimensional spatial
variability map of Available P 2008

Size of the z-axis data of the available phosphorus in the three-dimensional spatial variability map can fully demonstrate in 2003 (the year before fertilization variable z is the range of 50ppm-100ppm) and significant effect of variable fertilization in 2008 (the first five-year variable-rate fertilization range of zis20ppm-60 ppm) that contribute to the proper evaluation of soil fertility. While the three-dimensional spatial variability map for arbitrary translation, rotation, subdivision, extract isosurface and other operations to help us dig out the hidden rules behind the data.

(2) Analyze the soil organic matter, available nitrogen, available phosphorus and potassium 4 planting of 2008 combined with GIS data to obtain the three-dimensional

spatial soil nutrients variability map of nitrogen, phosphorus, potassium No. 3 region in 2008 (Figure 5), according to historical data for acquiring soil nutrient spatial evolution of history, provide an objective, image, reliable decision support tools to verify the result of variable rate fertilization process and prediction.
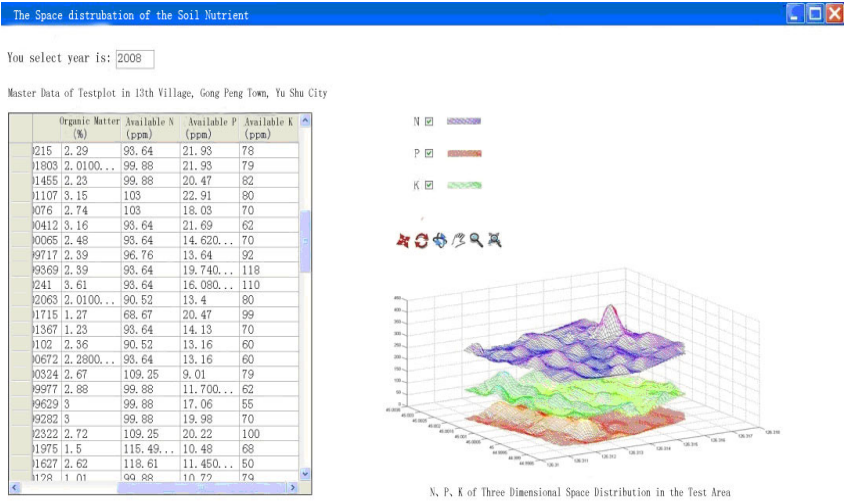
Fig. 5. Nitrogen, phosphorus and potassium distribution of three-dimensional space in 2008

3.2   Visualization of Spatial Data Mining

(1) Based on the principle of Weighted spatial fuzzy clustering algorithm, respectively, process the soil nutrient such as: the year before variable rate fertilization (2003), consecutive two-year variable rate fertilization(2005), consecutive 5-year variable rate fertilization (2008) and for seven consecutive years (2010) soil organic matter, available nitrogen, available phosphorus and potassium   data, the first use of expertise and   AHP to determine the weight coefficients to obtained weight in soil organic   matter,   available   nitrogen, available phosphorus and   potassium respectively are 0.1053, 0.2105, 0.1579, 0.5263; then establishing the weighted fuzzy clustering model; Finally, according to the results of clustering to determine the best threshold calculate F-distribution value to determine the optimum number of categories, F value is greater, indicating the distance between classes large categories, the better.

   (2) According to the spatial relationship of the grid, clustering results will then be analyzed in a GIS then visualization. (Figure6).
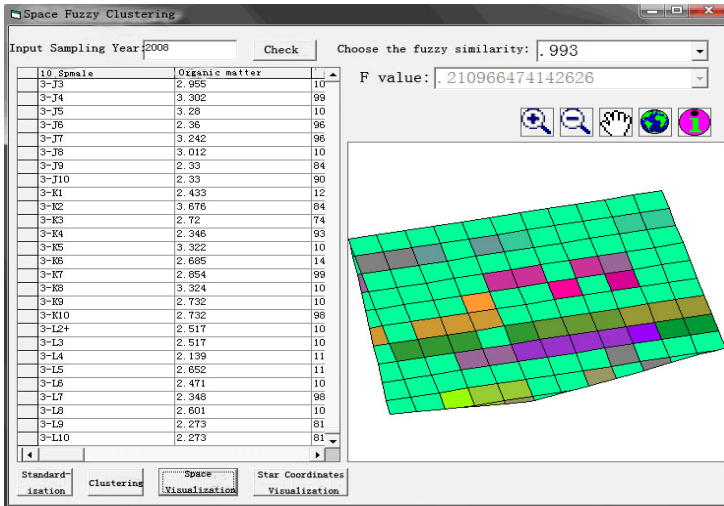
Fig. 6. Maize field soil nutrient spatial fuzzy clustering map in 2008

3.3   Space-Time Evolution of Soil Fertility

Temporal GIS technology add in the time dimension on the basis of the traditional relational model ,to expand the relational mode and relational algebra to handle temporal data, which directly or indirectly support spatial and temporal data storage, representation and processing based on the relational model . Based on this idea, this paper extends the traditional relationship model to achieve dynamic display the spatial and temporal evolution of soil fertility.

Analysis and determine the soil fertility dynamic evolution of spatial fuzzy clustering of the basic spatial pattern is the basic idea of this model which is spatial variation in soil fertility as background information, then follow the appropriate time step to map soil fertility clustering results of the state of each time slice data. Also expressed as: F (Si, T)-F (Si+1-i, T); which (S) for the space system (T) for the time system. Consists of two parts: (1) obtain the same region of the same type of clustering data over different time periods; (2) display continuous segment that the difference between the nutrient. Use the index to resolve the object description space objects and their properties in the time information retrieval, display and overlay, better perform dynamic presentation and analysis of spatial information between the query and visualization, easy, fast access to soil fertility dynamics simulation of the dynamic evolution of the information, and seize its essential character and the law (Figure 7).
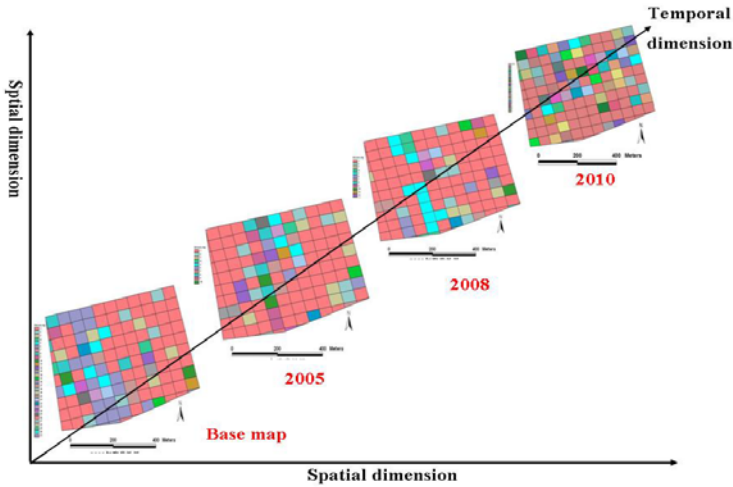
Fig. 7. Map of clustering variation

Figure 7 shows

(1) Accumulate the soil nutrients gradually through the continuous variable-rate fertilization, the entire block of soil organic matter, available nitrogen, available phosphorus and potassium, indicating differences in the overall decrease year by year, the similarity increase year by year.

(2) The results show that variable rate fertilization on soil nutrient spatial differences in a significant improved over years, indicating the temporal GIS model is able to describe the dynamics of the spatial variation of soil fertility.

4   Conclusion and Discussion

This paper discusses the use of spatial data mining, temporal GIS technology combined with visualization competent for many years of comprehensive analysis and comparison of soil fertility and effective evaluation of spatial and temporal evolution of fertility.

(1) Employ VTK and Kriging interpolation method to map three-dimensional distribution of soil nutrients is more natural, clear and intuitive than the traditional two-dimensional graph to represent the data variation.

(2) By visualization of mining weighted spatial fuzzy clustering algorithm applied comprehensive cluster to soil available nitrogen, available phosphorus, potassium and organic matter on 2003, 2005, 2008 and 2010, from the data analysis results, the continuous variables in the same block fertilization on soil nutrient spatial differences could obtain a significant improvement. Mainly reflected in: With the continuous increase in the year variable rate fertilization, differences in soil nutrients gradually reduce the overall data, data distribution is more concentrated and integrated similarity increased every year.

(3) Using the temporal GIS technology to display the weighted spatial fuzzy clustering results which can dynamic descript the spatial time variation of the variable

fertilization on soil nutrient, and indicating that the temporal GIS technology is a good new method for the soil fertility evaluation .

The effect of variable rate fertilization is a function that closely integrated with time and space, the text   analysises the data of time and space in only 7 years , so the evaluation of the effect of variable rate fertilization is only a preliminary, It   needs continuous research in order to obtain variable rate fertilization of economic and ecological benefits of a comprehensive evaluation.

References

1. Guo, X.D., Fu, B.J., Ma, K.M., et al.: Spatial variability of soil nutrients based on geostatistics with GIS:A case study of Zunhua City of Hebei Province. Chinese Journal of Applied Ecology 11(4), 557–563 (2000)
2. Karimipoura, F., Delavara, M.R., Frankb, A.U.: A simplex – based approach to implement dimension dependent spatial analyses. Computer & Geoscience 36 (2010)
3. Zhang, Q.L., Shi, X.Z.: Characteristics of spatio-temporal changes of soil fertility in JinTan County, JiangSu Province. Acta Pedologica Sinica 41(2), 315–319 (2004)
4. Yu, Y., Liu, J.P., Xu, Y.Y.: Spatial Variability Characteristics of Soil Nutrient Contents in the Black Soil Region of Northeastern China. Research of Soil and Water Conservation 16(5), 66–69 (2009)
5. Zhou, X., An, Y.L., Xu, W.C., Deng, M.L.: Fuzzy Evaluation on Soil Fertility of Cultivated Land Based on GIS and Improved AHP——A Case of Pu'an County in Guizhou Province. Chinese Journal of Soil Science 40(1), 52–55 (2009)
6. Li, L., Sun, Y.K., Zhou, B.K.: Effect of different fertilization systems on soil fertility in black soil of Heilongjiang Province. Journal of Northeast Agricultural University 41(4), 53–58 (2010)
7. Zou, G.M., Huang, M.Y., Su, D.R., Wang, Z.Y.: Characteristics of Temporal-Spatial Variability of Soil Fertility in City Greenbelt Soil of Coastal Saline Soil. Chinese Agricultural Science Bulletin 26(5), 110–115 (2010)
8. Lv, Z.L., Wang, X.B., Zhong, X.C.: A comprehensive evaluation on the fertility of salty soil based on geographic information system. Systems Engineering-Theory & Practice (9), 166–171 (2008)
9. Sun, R.J., Wang, D.J., Lin, J.H.: Evolution of Soil Fertility in Taihu Region and Its Causes. Soils 38(1), 106–109 (2006)
10. Yu, H., Chen, G., Bi, C.: The Study of Database Modeling as an Example on Maize Precision Fertilization. Journal of Maize Sciences 16(4), 184–188 (2008)
11. Chang, K.J.: Introduction to Geographic Information Systems. The McGraw-Hill Companies, Inc. (2002)
12. William, J.S.: The Visualization Toolkit User's Guide, 4th edn. Kitware Inc., New York (2001)
13. Hu, Y.J., Zhi, M., Xiao, W.S.: A VTK-Based Algorithm for the Triangulation Reconstruction from Scattered Points. Computer Engineering & Science 30(1), 69–72 (2008)
14. Yu, S.: The Study and Implemention of Visualization Based on VTK. Xi'an University of Architecture and Technology, Xi'an (2009)
15. Chen, G.-F., Cao, L.-Y., Wang, G.-W.: Application of Weighted Spatially Fuzzy Dynamic Clustering Algorithm in Evaluation of Soil Fertility. Scientia Agricultura Sinica 42(10), 3559–3563 (2009)
16. Liu, S.J., Li, T.F., Chen, H.L., et al.: Research on the Simulation of NDVI Dynamic Monitoring Based on Temporal GIS. Research of Soil and Water Conservation 13(4), 165–169 (2006)

Spatial Information Sharing Technology Based on Grid

Hong-bin Zhang[1,2,3], Bao-rui Chen[1,2,3], Gang Li[2,3], and Xiao-ping Xin[1,2,3,*]

[1] Hulunber Grassland Ecosystem Observation and Research Station, Beijing 100081, China
[2] Key Laboratory of Resource Remote Sensing and Digital Agriculture,
Ministry of Agriculture, Beijing 100081, China
[3] Institute of Agricultural Resources and Regional Planning,
Chinese Academy of Agricultural Sciences, Beijing 100081, China
`xinxp@sina.com`

Abstract. This paper analyzes the status of spatial data sharing technology. The grid technology based spatial information sharing platform has been adopted to solve the problem of "Isolated Information Islands". A Grid technology-based information sharing system is proposed and analyzes key technologies and processing mechanism.

Keywords: Grid, Information Sharing, metadata, geospatial information.

1 Introduction

Sharing of Geospatial information has great positive effect to economy and social development. But the data management systems built by different technique systems can not be accessed by each other, thus a lot of "isolated information islands" come into being. How to integrate these heterogeneous data sources distributed in different sites and to realize management and sharing quickly have become research focus at present.

2 Present Situation and Analysis of Geospatial Information Sharing

2.1 Present Situation of Sharing Geospatial Information

Under the promotion of application demands based on Geospatial information and with the development of Internet technology, relative research organizations and institutions promote one after another information publishing platforms and sharing systems. The main typical systems are as follows.

TerraServer: is an aerial images and satellite images publishing system designed and developed by Microsoft. Users can browse and download images in designated area by Internet.

---

* Corresponding author.

TerraShare: is a production of Z/I imaging company. It is modularized and client/server mode geography images management and publishing system. It can deal with remote sensing images and altitude modules and digital grid images.

Google Earth: is invented earth software developed by Google Company. It can publish images and vector data based on three-dimensional model and provide popular GIS functions such as inquiry, measurement, path query and so on. The systems of the kind include World Wind, Visual Earth and so on.

The research work in spatial information sharing of our country includes two aspects: Scientific data sharing engineering: Science and Technology Department began to implement scientific data sharing engineering from 2002 and in 2003 started up to build sharing centre of mapping and earth system scientific data and sharing service experimental unit and to provide data query and download service to public.

Chinese Spatial information Network: Chinese Remote Sensing Centre has already taken advantage of advanced information technique fully to built large scale and comprehensive spatial information professional website-"Chinese Spatial Data Network"based on spatial information sharing work and experience in recent years, emphasized to set up data sharing mechanism, focused on sharing spatial information. The network realized spatial information directory query and content query and download orders.

2.2   Existed Problems

Compared with advanced countries in the world, our country fall behind in spatial information development, utilization, obtaining, management, updating and services.

① Lack of integration of spatial information resources
② Short of standardized state spatial information resources
③ Backward management and low sharing level
④ Low level of automatic obtaining spatial information
⑤ Backward infrastructure of spatial information resources

2.3   Traditional Geospatial Information Sharing

At present traditional spatial information sharing methods includes: ①data format conversion mode②data direct access mode ③data mutual operation mode ④Web GIS data sharing mode. But the traditional methods can not realize mutual operation between heterogeneous spatial data and different Web GIS platforms.

2.4   Analysis of Implementation Technology

Comprehensive analysis of features and techniques in spatial information publishing systems based on Internet, there are two sharing methods.

①Users obtain needed spatial information through spatial meta data query and share them by storing in physical mediums or downloading from internet. This method can be called off-line manual sharing mode. The technique matter is to publish and browse spatial information with the aid of web technique and to provide downloading service through HTTP and FTP agreements, such as TerraServer and Chinese Spatial information Network.

②Publishing remote sensing images and providing simple applications of query and browse by special clients or browse plug-ins. The technique essence is that clients adopt special agreements to obtain data from servers. The servers usually adopt HTTP service or web services such as Google Earth.

Analysis above two methods, we can see that sharing methods and their technique means in present systems have two aspects shortages.

①Present sharing methods can not eliminate "isolated information islands" truly

The development of internet technique already can resolve information publishing easily, but can not resolve information sharing well. Storage systems used incompatible access methods however, thus remote sensing resources become isolated to users. With the building of network mapping service systems in different technique structures, the isolated information islands are not only not decreased but also increased.

From the technical means to analyze, that is because an easy sharing environment is wanting in internet based on TCP/IP agreement to resolve sharing issues involved in distributed and heterogeneous system, such as resources management, service publishing and safety problems.

②Can not share stored resources

Spatial information has the characteristic of mass memory. In a variety of application systems providing high speed and stability of the stored data service have a larger proportion of cost. Obviously, browse, download off the line of the way not only can not share storage resources, but also create the great wastage of storage resources in data circulation renew.

From the technical means to analyze, mainly because the technical level has not yet been able to direct support for remote sensing data owners to provide a summary way, that will be able to at a lower cost to wan other users to provide real-time online, stable and reliable data access service. On the other hand, internet will not be able to provide convenient integration service directly for data users. Difficulty lies in how to publish the service of information, how to realize data query, how to integrate services in application systems and how to control the data access and so on.

Emerge of grid technique proved chance to resolve above questions.

3   Grid Technology

Grid is the third generation of internet application following by Internet and www. Traditional internet realizes union of computer hardware. Web realizes union of web pages. But grid tries to realize complete union of all resources on the Internet and

provides weldless and compositive calculation and cooperation environment and provides fictitious and limitless calculation and distributed data resources and realizes resources sharing of fictitious organization and question solution.

Based on the research work and experience of key theories and techniques on grid calculation including resources management, information safety, information services, data management and application development environment, America Argonne national lab developed Globus Toolkit. It is grid calculation tool software which can running in every platform. At the same time, it is a grid application environment and development platform. Globus toolkit of development based on the grid application has provided on several areas of support.

①Grid safety. It can assure grid calculation environment to run regularly. Globus mainly combines sophisticated distributed safety technologies and expands to some extent in order to adapt to features of grid calculation environment.

②obtain and publish of grid information. How to publish, query and search resources information in grid calculation environment are preconditions for utilization resources effectively.

③Grid resources management. Because resources in grid environment mainly located in wan wide area network environment, frequently-used lan local area network resources management technologies can not be used effectively. So Globus realizes higher level resources management. Under support of information services, it can manage resources in wan effectively.

④Grid long distance data transmission. To realize high speed and reliable data transmission in wan and transparent long distance file I/O access of applications.

To sum up, the core of grid conception is resources sharing and cooperation work. Grid users can use all resources including calculation, storage and data conveniently, but it is not need to know who provides the services. On demands is the aim that grid tries to realize. Some person sum up the function of grid technology is to provide the most important resources for the most needed persons in the most appropriate time.

Advance and development of grid technology raise the hopes to solve difficulties in spatial information sharing and services fundamentally. Introduction grid technology to spatial information utilization and application has become important task at present for us.

4   Spatial Information Sharing Service Based on Grid

4.1   Technical Line

Spatial information grid technology is based on basic grid technology. The basic frame and function component of spatial information sharing platform is set up based on the technical frame, resources management, information services, safety architecture provided by basic grid laying the foundation for spatial information grid application system. The spatial information grid system architecture is as follow.
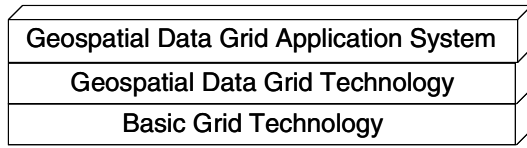
Fig. 1. Spatial information grid

4.2   Sharing Platform Frame

Grid middleware is adopted to eliminate isolated information islands. Spatial information sharing platform frame based on grid middleware is designed to realize transparent integration of spatial information resources. The grid middleware is the score to realize sharing spatial information taking the charge of transparent and dynamic access distributed spatial information resources. In the frame, there are several parts.
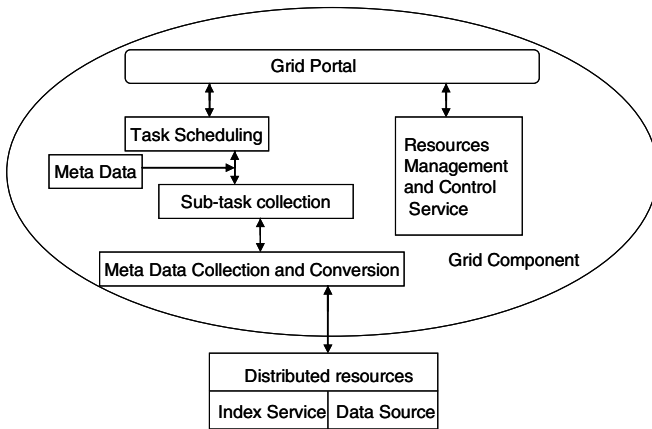
Fig. 2. Grid-based Geospatial sharing frame

(1)Grid data resources. That is rock-bottom distributed spatial information resources which can be accessed by internet. Mainly include two parts: ① space source, that is the basic unit of information stored and it can be database or distributed heterogeneous spatial data source; ② information search service, mainly mean through spatial information source to establish an index and to provide index service to upper.

(2)Task scheduling service, this part is mainly responsible for transferring request from spatial information portal, decomposition calls for resources, processing tasks and management of scheduling tasks.

(3)Meta data collection and conversion. According to the results of task scheduling, information collecting task service is localization converted to meet distributed spatial information resources format.

(4)Spatial information resources control service. This part is the control centre of the whole frame and makes every grid component in the reunification of the standard and includes system configuration and index services.

4.3   Key Technologies

(1) Presentation of information and meta data
Presentation of information reflects in two aspects: the first presentation is storing information in database or other storage medium; the second presentation is interface presented to users. Meta data abstracts description of data objection in order that information can be expressed by relationship of elemental attribute and value. We call the mate data as middle layer of data presentation. Generally speaking, data presentation can set up several middle layers and every application program logic layer has corresponding data presentation middle layer. Presentation of meta data can take advantage of semantics web technology fully.

(2) Task scheduling and resources management
Grid resources is complicated, numerous, varied and large. So how to manage all resources in the grid effectively is a painstaking task. Task scheduling mainly take the charge of resolving requirements transferred from spatial information portal, and decomposing of resources scheduling demands and scheduling tasks and management scheduling tasks. Task scheduling can dynamic decompose and schedule tasks according to every site disposition situation in grid in order to improve running efficiency. So scheduling algorithm will affect system efficiency directly.

(3) Safety technique
The aim of grid is to connect all the information service sites in the internet in order that all users can use information services. Because sites of grid located in different places, how to share data resources safely between sites and how to assure integrity of sharing data are questions to be solved at present. Grid safety authentication mechanism mainly includes identification authentication, information access authentication, audit mechanism. When setting up perfect safety mechanism, it should be avoided that safety authentication will waste too much processing capacity. If system resources used up by safety authentication is more than calculation resources saved by system, grid will lose itself true means. Fault-tolerant and automatic crash recovery should also take into consideration in grid system.

5   Results

Grid technique is a new generation of internet. Resources sharing is the most important mode in grid technique application. With the aid of grid, combined with spatial information sharing situation, study spatial data sharing and services based on grid in order to perfect spatial information sharing and application environment.

References

1. Qiang, W., Jin, H., Shi, X., Zou, D.: A Novel-Based Access Control Model for Grid. In: Jin, H., Pan, Y., Xiao, N., Sun, J. (eds.) GCC 2004. LNCS, vol. 3251, pp. 293–300. Springer, Heidelberg (2004)
2. Cao, H., Spooner, D.P., Jarvis, S.A., Nudd, G.R.: Grid load balancing using intelligent agents. Future Generation Computer System 21(1), 135–149 (2005)
3. OGC 07-006rl. OpenGis catalogue services specification version 2. Open Geospatial Construction (2007)
4. Wang, L.: Geospatial data sharing based on grid. Journal of Computer Application 30, 60–63 (2010)
5. Feng, H., Fan, H., Li, S.: Study on West Mapping Spatial Information sharing and Service Platform based on grid. Journal of Geomatics 35(3), 56–72 (2010)

A Robust Graph Based Learning Approach to Agricultural Data Classification

Baojie Ji[1], Caili Su[2], and Wanzhong Lei[3]

[1] Mechanical & Electrical Engineering College,
Hennan Agricultural University, Zhengzhou 450002, China
[2] Institute of Environmental Science and Engineering, North China University of Water
Resources and Electric Power, Zhengzhou 450011, China
[3] Department of Electrical Information Engineering,
Henan Institute of Engineering, Zhengzhou 451191, China
hnkfjybj@163.com

Abstract. This paper proposes a novel graph based learning approach to classify agricultural datasets, in which both labeled and unlabelled data are applied to the classification procedure. In order to capture the complex distribution of data, we propose a similarity refinement approach to improve the robustness of traditional label propagation. Then the refined affinity matrix is applied to label propagation. Thus, the traditional pair-wise similarity is updated with scores using median filter of its neighbors in manifold space. And the proposed classification approach can propagate the labels from the labeled data to the whole dataset. The experiments over agricultural datasets have shown that embedding information fusion approach in manifold space is beneficial in classification.

Keywords: semi-supervised learning, graph based learning, label propagation.

1 Introduction

With the development of precision agriculture, probabilistic modeling and machine learning method has attracted great research attention in the agriculture domain. Many representative classifiers have been used in agriculture classification, such as the two-dimensional multi-resolution hidden Markov models [1], support vector machine [2],[3], Bayes Point Machine [4], and Mixture Hierarchical Model [5], but the classification result can not be satisfied. In this paper, we propose novel graph based learning approach to classify agriculture dataset. The pair-wise similarities of dataset are first computed in attribute space. To improve the robustness of traditional graph based learning methods, we then refine the pair-wise similarities by embedding the neighbor information in manifold space.

The rest of the paper is organized as follows. Section 2 describes the theory of graph based learning. Section 3 introduces the proposed classification scheme. Section 4 represents the key algorithm. Section 5 reports and discusses the experimental results and Section 7 concludes the paper.

2   Graph-Based Semi-Supervised Learning

2.1   Graph-Based Learning Theory

Semi-supervised learning (SSL)[6],[7],[8],[9],[10],[11],[12], which attempts to learn from both labeled and unlabeled data, is a promising approach recently. As a major family of SSL, graph-based learning[13] algorithms have been the focus of much recent machine leaning research. The key to which learning is the two basic assumptions: 1) neighborhood assumption: nearby points are likely to have the same label; 2) structure assumption: points on the same structure (such as a cluster or a sub-manifold) are prone to have the same label. Note that the first assumption is local, while the second one is global. The cluster assumption implies us to consider both local and global information during learning.

In graph-based learning, data points $(1,\cdots,l$ labeled points and $(l+1,\cdots,n$ unlabeled points) are arranged in a weighted undirected graph. The graph is characterized by a weight matrix W , whose elements $W_{ij} \geq 0$ are similarity measures between vertices i and j , and by its initial label vector $Y_L = (y_1,\cdots,y_l), y_i \in \{1,\cdots,C\}$,that defines labels for the first l points. If there is no edge linking nodes i and j $W_{ij} = 0$ .Other than that, applications have considerable freedom in choosing the edge set and the W_{ij} weights. The commonly used weight is defined by a Gaussian kernel: $W_{ij} = \exp(-d(x_i,x_j)^2/(2\sigma)^2)$ where $d(x_i,x_j)$ is the estimated distance between feature vector x_i and x_j , σ is a bandwidth hyper-parameter. Various distance measures can be used, e.g. Cosine distance, Euclidean distance or Jensen Shannon.

2.2   The Robustness of Graph-Based Learning

Although graph based learning, which attempts to learn from both labeled and unlabeled data, is a promising approach to deal with classification, one of the key problems is the robustness. This can be easily understood from the label propagation framework, which states that predicting the labels of the unlabeled data on a graph is equivalent to propagating the labels of the labeled data along the edges to the unlabeled ones. Clearly, the existence of the bridge points on the graph will cause the labels to wrongly propagate to different classes. The two moon example in [14] shows that the bridging points can bias the final classification results severely.

3   The Proposed Classification Approach Based on Graph-Based Learning

3.1   Graph Construction

Suppose the dataset is $X = (x_1,\cdots,x_l,x_{l+1},\cdots,x_n) \subset R^m (1,\cdots,l$ labeled dataset and $l+1,\cdots,n$ unlabeled dataset), the initial similarity matrix W , is constructed as

$$W_{ij} = \exp(-d(x_i, x_j)^2 / (2\sigma)^2) \tag{1}$$

Where $d(x_i, x_j)$ is the distance in feature space. L1 distance is used in our implementation. σ is a bandwidth hyper-parameter, Note that $W_{ij} = 0$ because there are no loops in the graph. We use the KNN approach to construct the graph.

To alleviate the bridge points in traditional graph based learning to improve robustness in classification, we propose a pair-wise similarity refinement scheme to refine the traditional approach, in which the average Hausdorff distance[15] of the neighborhood information of each data is used to refine the traditional pair-wise matrix W,

$$aveH(A, B) = \frac{\sum\limits_{a \in A b \in B} \min \|a - b\| + \sum\limits_{b \in B a \in A} \min \|b - a\|}{|A| + |B|} \tag{2}$$

Where $|\bullet|$ measures the cardinality of a set. In other words, $aveH(\bullet, \bullet)$ averages the distances between neighbor information of the data and that in the other data. So

$$\tilde{W}_{ij} = \exp(-d(A, B)^2 / (2\sigma)^2) \tag{3}$$

Thus the similarity becomes the fusion information of its neighborhood instead of the initial pair-wise similarity. In this way, the bridge point problem can be alleviated to some extent. Then symmetrically normalize $\tilde{W}$ by $S = D^{-1/2} W D^{-1/2}$. where D is a diagonal matrix , $D_{ii} = \sum\limits_{j=1}^{n} \tilde{W}_{ij}$ and S is the initial pair-similarity symmetrical matrix.

3.2   Label Propagation

After the graph has been constructed, we predict the classification labels of the unlabeled data in the dataset with the data similarities.

Suppose there are C classification labels, and the classification label set is $L = \{1, \cdots, C\}$. Let M be $n \times c$ matrices with nonnegative real valued entries. F denote the set of label functions defined on X, initialize a $n \times c$ matrix T with binary vectors encoding the known labels for the first l rows, $F_i = \delta_C(y_i) \forall i \in \{1, 2, \cdots, l\}$, in each propagation step, we let each data fuse a fraction of label information with its similar data, therefore, the label of X at time $m + 1$ becomes

$$F^{m+1} = aSF^m + (1-a)T \tag{4}$$

where a is the parameter and $a \in (0,1)$ We will use (4) to update the labels of each image until convergence. The theorem in [14] guarantees that the sequence $\{F^m\}$ converges to

$$F^* = (1-a)(I-aS)^{-1}T \tag{5}$$

4   The Proposed Key Algorithm

Step1. Compute the k nearest neighbors of each data with L1 distance.

Step2. Construct the semantic similarity matrix $\tilde{W}$ as $\tilde{W}_{ij} = \exp(-d(A,B)^2/(2\sigma)^2)$, average Hausdorff distance is used in our implementation.

Step3. Symmetrically normalize $\tilde{W}$ by $S = D^{-1/2}WD^{-1/2}$, where D is a diagonal matrix and $D_{ii} = \sum_{j=1}^{n} \tilde{W}_{ij}$.

Step4. Do iteration according to (4) $F^{m+1} = aSF^m + (1-a)T$ until convergence.

Step5. According to (5) decide the label for each unlabeled image based on the convergent matrix of F^* .

5   Experimental Results and Discussion

5.1   Experimental Data Sets

To evaluate the proposed techniques, we use the UCI[15] agricultural dataset on mushroom and soybean. The first one includes descriptions of hypothetical samples corresponding to 23 species of gilled mushrooms in the Agaricus and Lepiota Family. The attribute information is 22D, such as cap-shape, cap-surface, gill-color, stalk-root and so on. The second dataset has 19 classes, there are 35 categorical attributes. The construction of each data set for our experiments is done as follows: Firstly, we randomly select 40% of the data from the each category, and put them into test set to evaluate the performance of classifier. Then, the rest are used to create training sets.

5.2   Performance Measures

To analyze the performance of classification, we adopt the popular F1 measure. F1 measure is combination of recall (re) and precision (pr), $F_1 = 2.re.pr/(re + pr)$.

Precision means the rate of documents classified correctly among the result of classifier and recall signifies the rate of correct classified documents among them to be classified correctly. The F1 measure which is the harmonic mean of precision and recall is used in this study since it takes into account effects of both quantities.

5.3   Experimental Results and Discussion

The comparison between the traditional SVM method and the proposed scheme are showed in figure1 and figure2, from the result, it can be seen that the method is effective. Our approach yields a higher performance compared to traditional methods over all categories. For mushroom, our approach yields values of 86.4% average classification result, whereas the SVM yields the F1 values of 84.2%. For figure2, our approach also yields higher average classification performance of 21.2% over the SVM methods.
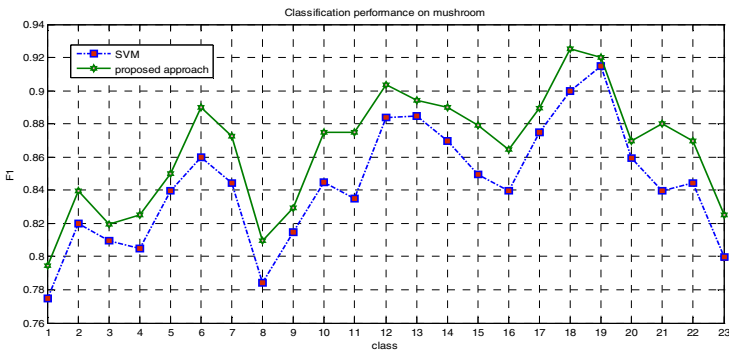
Fig. 1. Comparison of classification performance on mushroom
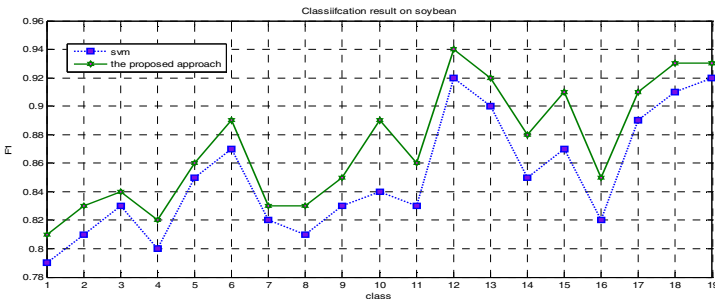
Fig. 2. Comparison of classification performance on soybean

6   Conclusion

In this paper, a novel approach to agricultural data classification based on a novel graph based learning approach is proposed. First the initial pair-wise similarities in dataset are

computed based on the distance in attribute space, then it is refined by the fusion of its neighborhood information. Finally the updated similarities are adopted to propagate the class labels from the labeled dataset to the whole dataset. Experimental results show that the proposed approach can propagate more accurate classification results and improve the robustness of the traditional scheme.

References

1. Li, J., Wang, J.Z.: Automatic Linguistic Indexing of Pictures by a Statistical Modeling Approach. IEEE Trans.Pattern Anal. Mach. Intell. 25(9), 1075–1088 (2003)
2. Cusano, C., Ciocca, G., Schettini, R.: Image Annotationusing SVM. In: Proceedings of Internet imaging IV, vol. 5304. SPIE (2004)
3. Gao, Y., Fan, J., Luo, H., Xue, X., Jain, R.: Automaticimage annotation by incorporating feature hierarchy andboosting to scale up SVM classifiers. In: Proceedings of the 14th Annual ACM International Conference on Multimedia, Santa Barbara, CA, USA, October 23-27 (2006)
4. Chang, E., Kingshy, G., Sychay, G., Wu, G.: CBSA:content-based soft annotation for multimodal image retrieval using Bayes point machines. IEEE Trans. on CSVT 13(1), 26–38 (2003)
5. Carneiro, G., Vasconcelos, N.: A database centric view of semantic image annotation and retrieval. In: Proceedings of the 28th Annual international ACM SIGIR Conference on Research and Development in information Retrieval, Salvador, Brazil, August 15-19 (2005)
6. Lippmann, R.P.: Pattern classification using neural networks. IEEE Communications 27(11), 47–64 (1989)
7. Chapelle, A., Zien, B.: Scholkopf. Semi-supervised learning. MIT Press (2006)
8. Shahshahani, B., Landgrebe, D.: The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon. IEEE Transactions on Geoscience and Remote Sensing 32(5), 1087–1095 (1994)
9. Zhu, X., Ghahramani, Z., Lafferty, J.: Semi-supervised learning using Gaussian fields and harmonic functions. In: Proceedings of the 20th International Conference on Machine Learning, pp. 912–919 (2003)
10. Zhou, D., Bousquet, O., Lal, T.N., Weston, J., Schollkopf, B.: Learning with local and global consistency. In: Thrun, S., Saul, L., Scholkopf, B. (eds.) Advances in Neural Information Processing Systems, pp. 321–328 (2004)
11. Belkin, M., Niyogi, P., Sindwani, V.: On manifold regularization. In: Proceedings of 10th International Workshop on Artificial Intelligence and Statistics, pp. 17–24 (2005)
12. Blum, A., Chawla, S.: learning from labeled and unlabeled data using graph mincuts. In: Proc.18th International Conf. on Machine Learning, pp. 19–26 (2001)
13. Zhou, D., Weston, J., Gretto, A., Bousquet, O., Scholkpf, B.: Ranking on data manifolds. In: Proc of NIPS 2003 (2003)
14. Wang, Zhang, C.: Label propagation through linear neighborhoods. IEEE Transactions on Knowledge and Data Mining, 55–67 (2008)
15. Zhou, D., Weston, J., Gretto, A., Bousquet, O., Scholkpf, B.: Ranking on data manifolds. In: Proc of NIPS 2003 (2003)

Wheat Grain Protein Content Estimation Based on Multi-temporal Remote Sensing Data and Generalized Regression Neural Network

Cunjun Li[*], Qian Wang, Jihua Wang, Yan Wang,
Xiaodong Yang, Xiaoyu Song, and Wenjiang Huang

Beijing Research Center for Information Technology in Agriculture,
Beijing, 100097, China
lcj963@163.com

Abstract. Monitoring grain protein content in large areas by remote sensing is very important for guiding graded harvest, and facilitates grain purchasing for processing enterprises. Wheat grain protein content (GPC) at maturity was measured and multi- temporal Landsat TM and Landsat ETM + images at key stages in 2003, 2004 growth stages were acquired in this study. GPC was estimated with multi-temporal remote sensing data and generalized regression neural network (GRNN) method. Results show that the GPC prediction accuracy of the GRNN model is higher, with the average relative deviation of self-modeling, average relative deviation of cross-validation as 0.003%, 0.321%; 4.300%, 7.349% for 2003 and 2004 respectively. GRNN method proves to be reliable and robust to monitoring GPC in large areas by multi-temporal and multi-spectral remote sensing data.

Keywords: wheat, grain protein content, multi-temporal, Landsat, generalized regression neural network.

1   Introduction

Grain protein content (GPC) is a key indicator of wheat quality, which determines the usage of wheat for making bread, cookie or noodle. In world wheat trade, the higher the GPC is, the higher price grain production manufactures will pay for wheat farmers. Monitoring GPC aids to harvest wheat according to grain protein content, and facilitates processing enterprises to purchase graded grains. It becomes possible to monitor the quality of crops using quantitative remote sensing technology development in recent years.

Many scholars have monitored wheat protein content with remote sensing technology. Wright D.L et al. (2003) found that a good relationship existed between measured wheat organ nitrogen content and vegetation indices derived from airborne near infrared image and Quickbird image at heading stage, which can be applied to

---

[*] Corresponding author.

guide the fertilization at anthesis. The GPC was improved after twice fertilization for wheat under nitrogen stress. But poor relationships were found between vegetation indices derived from only the image derived from heading stage. Badri etc (2003) found that, vegetation indices derived from Landsat satellite images were greatly significantly correlated with wheat GPC (r > 0.75) two weeks before anthesis. Reflectance of near-infrared wavelengths of Aster image were significantly related to GPC three weeks after flowering with lower correlation coefficient (r < 0.5). Preliminary explanations were made for monitoring mechanism. Mieke Reyniers et al. (2006) used color infrared aerial image and Cropscan geophysics spectrometer to predict wheat GPC a month ahead of wheat harvest with prediction accuracy as 90%. Petterson (2007) got satisfied prediction GPC using TCARI index in planting. Huang et al.(2004) studied correlations between total wheat leaf nitrogen and grain quality components during late maturation period. The results showed that total nitrogen can predict GPC and gluten content. Wang et al. (2004) used PPR to monitor GPC based on field plots. Hansen et al. (2003) conducted plot experiment and measured ground-based multi-spectral, wide band (visible - near-infrared wavelengths) data. Wheat yield and protein partial least squares models were established. The high yield prediction accuracy is in contrast to low protein prediction accuracy. Song et al. (2006) monitored wheat protein content based on nitrogen translocation in wheat plant and Aster images at anthesis. Zhao et al. (2005) predicted winter wheat grain protein content with water vegetation indices derived from TM image acquired at anthesis. Analyzed relationships between GPC and nitrogen transport process and moisture stress.

Monitoring winter wheat grain protein content at field scale experiment with different level of nitrogen fertilization at the key growing stage is very common. However, at large scale wheat GPC will be affected by soil, weather conditions, management and other factors. These factors along with cultivars together contribute spatial variability of GPC. Only one remote sensing image at certain period can't untangle all the factors aforementioned. Multi-temporal images overcome the shortcomings of single image and can monitor the major factors that cause the spatial variety of wheat grain protein content. Therefore, in this paper we attempt to use multi-temporal remote sensing images and GRNN method to monitor GPC.

2   Material and Method

2.1   Research Region

More than twenty wheat fields, distributed randomly, flat, uniformly grown, with areas more than 50 acres, were selected for this study in Beijing rural area, China in 2003, 2004. Regular management were adopted on these fields. Local popular varieties were planted on the fields, such as Jingdong 8, Jing 9507 and so on. The fields are shown in figure 1.
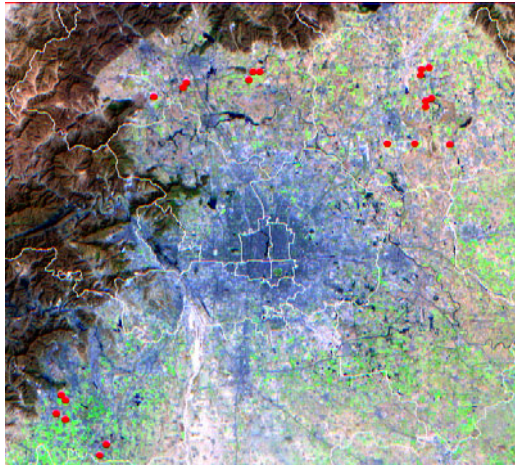
Fig. 1. Study area and the distribution map of wheat plot surveyed

2.2   Data Acquisition and Preprocessing

One month after wheat maturity, GPC was tested with NIR (Foss1241, Foss Tecator, Sweden).

The time and image type of the satellite data obtained include Landsat TM and Landsat ETM+ in wheat growing season of 2003 and 2004, which are shown in table 1.

Table 1. The satellite data Obtained

YEAR	DATE	TYPE
2003	April 7,	ETM+
	May 9,	ETM+
	May 17,	TM
	May 25	ETM+
2004	April 1	TM
	April 17,	TM
	May 18	TM

The radiation correction adopts Empirical experience linear method (Liang, 2003). High radiation body and low radiation body with flat land were determined. Through experience linear spectral reflectance method and remote sensing image transformation, DN values of target true reflectance were calculated. Low radiation body and high radiation body were Ming tombs reservoir and drying yard of Xiaotangshan precision agriculture demonstration base respectively.

Geometric correction of satellite data includes the following steps. First of all, 1:100,000, topographic map in the study area was used as base map. Second, the April 7th, 2003 Landsat image was corrected with control points of the base map.

Finally, the error should be controlled within a pixel. Other images were also geometrically corrected in the same way. Beijing 54 coordinate system was adopted for projection with the center longitude as 117 °.

2.3   Data Analysis

2.3.1   Extraction of Remote Sensing Spectral Index

According to Landsat band set and previous studies, seven multi-spectral vegetation indices were selected to monitor wheat nitrogen content, biomass, leaf chlorophyll density, and leaf equivalent water thickness. NDVI 21, NDVI 42 are sensitive to chlorophyll density, nitrogen content. NDVI47, RVI43 are sensitive to the biomass and leaf equivalent water thickness.

Table 2. Senven muti-spectral vegetation indices

Abbreviation	Index	Formula
NDVI42	NDVI of 4 and 2	NDVI42=(B4-b2)/ (B4+b2)
NDVI43	NDVI of 4 and 3	NDVI43=(B4-b3)/ (B4+b3)
NDVI47	NDVI of 4 and 7	NDVI47=(B4-b7)/ (B4+b7)
NDVI21	NDVI of 2 and 1	NDVI21=(B2-b1)/ (B2+b1)
NDVI57	NDVI of 5 and 7	NDVI57=(B5-b7)/ (B5+b7)
RVI43	RVI of 4 and 3	RVI43=B4/B3
RVI54	RVI of 5 and 4	RVI54=B5/B4

*1，2，3…7 refer to band of LANDSAT TM/ETM+.

2.3.2   Generalized Regression Neural Network

Neural network is a modeling method in recent years. Generalized regression neural network (GRNN) is a memory-based feed forward network.

The GRNN consists four layers: input layer, pattern layer, summation layer and output layer (Specht, 1991; Chtioui, 1999; Haidar, 2011). Figure 2 shows a schematic diagram of generalized regression neural network architecture. GRNN has a special property in which no iterative training of the weight vectors is required. That is, any input–output mapping is possible by simply assigning the input vectors to the centroid vectors and fixing the weight vectors between the radial basis function units and outputs identical to the corresponding target vectors. This training algorithm is much better than the back propagation training, which involves long and iterative training as well as facing the problem of local minima. Moreover, a special property of GRNN is that it enables users to flexibly configure the network suitable for real hardware implementation, by adjusting only two parameters which are the center and width. Since the radial basis function acts as a detector for different input vectors, the weight vectors are computed accordingly and there is no need to train the network. Thus, the GRNN is more straightforward and does not require a training process.
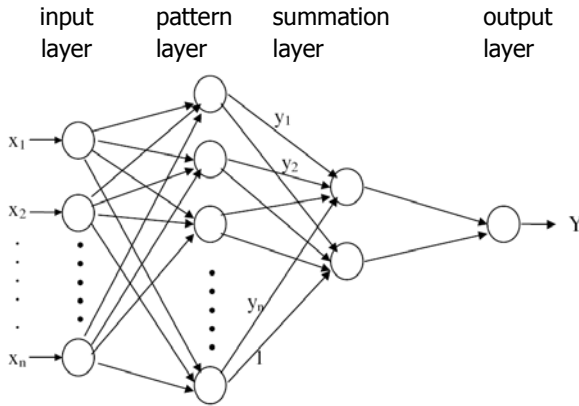
Fig. 2. Schematic diagram of GRNN architecture

In GRNN, If $f(X, y)$ represents the known joint continuous probability density function of a vector random variable, X, and a scalar random variable, y, the conditional mean of y given X (also called the regression of y on X) is given by

$$E[y \mid X] = \frac{\int_{-\infty}^{\infty} y f(X, y) dy}{\int_{-\infty}^{\infty} f(X, y) dy} \tag{1}$$

Where f (x,y) is the joint density and can be estimated by using Parzen's nonparametric estimator. Substituting Parzen's nonparametric estimator for f (x,y) and performing the integrations leads to the fundamental equation of GRNN:

$$\hat{Y}(X) = \frac{\sum_{i=1}^{n} Y^i \exp(-\frac{D_i^2}{2\sigma^2})}{\sum_{i=1}^{n} \exp(-\frac{D_i^2}{2\sigma^2})} \tag{2}$$

Where σ is the smoothing parameter (sigma weight) and $D_i^2 = (x - x_i)^T (x - x_i)$. Training the GRNN involves finding the optimal values for the σ parameters in eq(2), whose optimal value is often determined experimentally (Kim,2003).

In this research, Matlab was used to train GRNN for protein estimation using muti-temporal Landsat TM/ETM+ data.

2.3.3   Model Validation

The small number of selected sampling fields was about 20 in the study due to the large area, leading to some verification problems. Cross-validation method is used in this study. .Assume the number of all the samplings is n. Each time a sampling is left out to validate the models established on the other n-1 sampling. In this way, all the n samplings are used to validate corresponding models. Deviation is calculated as: bias = | predicted value- actual value| / actual value× 100%. So according to the deviation the predictive power of modeling can be compared.

3   Results

Multi-temperal Landsat TM/ETM+ image from 2003 and 2004 were chosen as source data to estimate wheat grain protein content by GRNN method. Figure 3 shows the measured grain protein content versus predicted values derived from multi-temporal, multi-spectral parameters of all samples in 2003 with GRNN method.. Table 3 is the validation results.
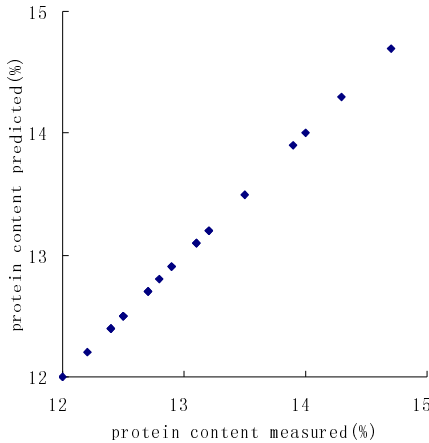
Fig. 3. Measured GPC versus predicted GPC in 2003

Table 3. Cross validation of predicted wheat GPC using muti-temporal airborne Landsat mutispectral data and GRNN method in 2003

Group	Deviation of cross-validation	Average deviation of self-modeling	Group	Deviation of cross-validation	Average deviation of self-modeling
1	0.995	0.002	12	6.459	0.001
2	2.336	0.001	13	6.733	0.001
3	7.919	0.001	14	7.541	0.001
4	2.291	0.001	15	1.676	0.001
5	6.429	0.002	16	2.613	0.001
6	6.041	0.001	17	3.914	0.007
7	1.643	0.001	18	0.688	0.001
8	11.567	0.001	19	0.355	0.003
9	9.794	0.001	20	1.078	0.013
10	0.692	0.001	21	3.788	0.015
11	5.752	0.001	Average	4.300	0.003

As shown in figure 3 and table 3, the GRNN method has potential to simulate GPC for multi-temporal and multi-spectral parameters of all samples in 2003, with high consistency and little deviation between measured and simulated ones. The average relative deviation of self-modeling and cross-validation were 0.003% and 4.300% respectively. The low deviation shows that the GRNN method has powerful ability to construct models and is a robust method to predict grain protein content.

Figure 4 shows the measured GPC versus the simulated values in 2004. 22 samples were collected in 2004, table 4 shows 22 cross-validation results for corresponding models.
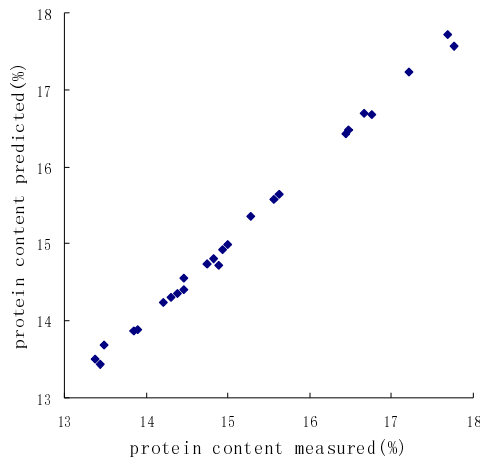
Fig. 4. Measured GPC versus predicted GPC in 2004

Figure 4 and table 4 shows that GRNN method has the potential to predict GPC with multi-temporal and multi-spectral parameters in 2004. The average relative deviation of self-modeling and cross-validation are 0.321% and 7.349% respectively. The high prediction accuracy shows that GRNN is suitable to simulate GPC.

By comparing the prediction results in 2003 and 2004, the former was better than the latter. This is perhaps because that only one image (May 8th during late period of wheat growth) was acquired in 2004, while three images including May 1st, May 9th and May 25th, were acquired during wheat growth in 2003. This is consistent with the previous conclusions that images for wheat anthesis and grain filling periods are of great importance to predict GPC (Wang, 2004).

Table 4. Cross validation of predicted wheat GPC using muti-temporal Landsat mutispectral data and GRNN method in 2004

Group	Deviation of cross-validation	Average deviation of self-modeling	Group	Deviation of cross-validation	Average deviation of self-modeling
1	6.660	0.085	13	10.188	1.044
2	0.659	0.152	14	13.359	0.491
3	22.515	1.502	15	3.697	0.098
4	2.628	0.126	16	10.975	0.043
5	6.703	0.005	17	7.446	0.13
6	7.241	0.036	18	15.625	0.613
7	2.433	0.034	19	0.987	0.134
8	13.256	0.026	20	6.014	0.51
9	10.467	0.008	21	1.879	0.228
10	0.135	0.047	22	1.992	0.375
11	5.589	0.467			
12	11.233	0.904	Average	7.349	0.321

4   Conclusion and Discussion

In this study, GRNN method proves to accurately predict GPC in large areas. Although wheat GPC were affected by soil, weather conditions, management and other factors, multi-temporal images acquired during critical growth development can overcome the low accuracy problems caused by single image.

However the distinguished weather conditions between 2003 and 2004 make it difficult to apply the models established on one year's data to predict the other year's grain protein content. Besides, it is necessary to combine meteorological data to establish prediction model in the future studies.

References

1. Badri, B.B., Apan, A.A., Kelly, R.M., Jensen, T.A., Strong, W.M., Butler, D.G.: Relating satellite imagery with grain protein content. In: Proceedings of the Spatial Sciences Conference, Canberra, pp. 22–27 (2003)
2. Chtioui, Y., Panigrahi, S., Francl, L.: A generealized regression neural network and its application for leaf wetness prediction to forecast plant disease. Chemometr. Intell. Lab. Syst. 48, 47–58 (1999)

3. Haidar, A.M.A., Mustafa, M.W., Ibrahim, F.A.F., Ahmed, I.A.: Transient stability evaluation of electrical power system using generalized regression neural networks. Applied Soft Computing Journal 11(4), 3558–3570 (2011)
4. Hansen, P.M., Jorgensen, J.R., Thomsen, A.: Predicting grain yield and protein content in winter wheat and spring barley using repeated canopy reflectance measurements and partial least squares regression. Journal of Agricultural Science 139(3), 307–318 (2002)
5. Huang, W.J., Zhao, C.J., Wang, J.H., Wang, J.D., Ma, Z.H.: Application of red edge variables to nutrition diagnosis and grain quality forecast of winter wheat. Transactions of The Chinese Society of Agricultural Engineering 20(6), 1–6 (2004) (in Chinese)
6. Kim, B., Kim, S., Kim, K.: Modelling of Plasma Etching using a Generalized Regression Neural Network. Vacuum 71, 497–503 (2003)
7. Liang, S.L.: Quantitative Remote Sensing of Land Surfaces. Wiley & Sons, Hoboken (2004)
8. Reyniers, M., Vrindts, E., Baerdemaeker, J.D.: Comparison of an aerial-based system and an on the ground continuous measuring device to predict yield of winter wheat. European Journal of Agronomy 24(2), 87–94 (2006)
9. Song, X.Y., Wen, J.H., Wang, J.H., Liu, L.Y., Li, C.J.: Preliminary application of ASTER images in winter wheat quality monitoring. Transactions of the Chinese Society of Agricultural Engineering 9, 148–154 (2006)
10. Specht, D.F.: A General Regression Neural Network. IEEE Transactions on Neural Networks 2(6), 568–576 (1991)
11. Wang, Z.J., Wang, J.H., Liu, L.Y., Huang, W.J., Zhao, C.J., Wang, C.Z.: Prediction of grain protein content in winter wheat (Triticum aestivum L.) using plant pigment ratio (PPR). Field Crops Research 90(2-3), 311–321 (2004)
12. Wright, D.L., Ritchie, G., Rasmussen, V.P., Ramsey, R.D., Baker, D.: Managing Protein in hard red spring wheat with remote sensing. In: The 6th Annual National Wheat Industry Research Forum, Hyatt Regency Albuquerque, New Mexico (2003)
13. Zhao, C.J., Liu, L.Y., Wang, J.H., Huang, W.J., Song, X.Y., Li, C.J.: Predicting grain protein content of winter wheat using remote sensing data based on nitrogen status and water stress. International Journal of Applied Earth Observation And Geoinformation 7(1), 1–9 (2005)

Study on the GPS Data De-noising Method Based on Wavelet Analysis[*]

Debao Yuan, Ximin Cui, Guo Wang, Jingjing Jin, Dongli Fan, and Xiaogang Jia

College of Geoscience and Surveying Engineering, CUMT, Beijing,
100083, Beijing, China
{yuandb,cxm}@cumtb.edu.cn, wangguo0123@foxmail.com

Abstract. Signal de-noising is one of the classical problems in the field of signal processing. As a new signal processing tools, wavelet analysis, which has excellent noise performance, has caused growing concern and attention. The wavelet threshold de-noising has been researched systematacially, and the wavelet de-noising method is used on the GPS signal, which has achieved very good results.

Keywords: wavelet, threshold, de-noising, GPS signal.

1   Introduction

Analysis of wavelets theory is part of the branch of mathematics-----functional analysis, which is developed on the basis Fourier analysis theory, and it can provide more information that Fourier analysis theory can not provide. So in terms of theoretical study and engineering application it is of great value and practical significance. The past ten years, wavelet analysis has rapidly developed, and it has been widely used in graphics, image processing and analysis, CT Imaging, radar, seismic exploration, edge phase cycle slips detection, error analysis, waveform decomposition and combination (refactoring), data compression and so on. Wavelet analysis not only in time domain but also in frequency domain has good localization properties, its window size does not change its shape, and it is a time-frequency localization analysis method that can be changed by time window and frequency window. It has the characteristics of multi-resolution analysis, and the density of different frequency components spatial sampling in time-domain as the signals in the process of thinning can be adjusted automatically. Therefore, the wavelet analysis has the properties that it

can reflect the details of observation function and analysis. Based on this characteristic, it was known as mathematics microscope.

In the process of testing signal, Signals are often under the influence of interference and noise signal, such as data collection and transmission may introduce noise. Therefore in practice, de-noising must be carried out before measurement data analysis.

At present, there are traditional methods of achieving noise filtering and wavelet de-noising methods, in the actual test different de-noising method should be chosen depending on the different noise and signal. De-noise filter is only suitable for stationary random signals and signal separation, and entropy of a signal will be higher after changes, thus signal correlation can not be get. Wavelet analysis method has developed nearly a dozen years; it has an obvious advantage in de-noising because it has the characteristics of multi-resolution analysis, it can be used for non - stationary signal de-noising, it is not only effective to remove the noise, but also better to retain signal of mutation section.

2   Principle of Wavelet Threshold De-noising

General principle of signal de-noising is that the frequency distribution of noise and signal are different. In the traditional signal de-noising method based on Fourier transformation, the band overlapped part of the signal and noise is always made as small as possible, thus in the frequency domain by time - invariant filter, it can be distinguished between signal and noise. But if the two overlapping area is large, it is unlikely to make a noise removal effect.

With white Gaussian noise signal Wavelet transform, known by the characteristics of Wavelet transform, Gaussian noise by Wavelet transform is still a Gaussian distribution, and it is evenly distributed in parts of the frequency scale space. Due to their limited nature, wavelet coefficients of signal are just concentrated in limited parts of the frequency on the scale-space.

Principle of threshold de-noising: After wavelet decomposition, each coefficient model larger or smaller than a threshold should be processed separately, then process wavelet coefficients to inverse transform, and reconstruction after the signal de-noising, thus to achieve the purpose of de-nosing. The main steps are as follows:

(1) Decompose signal using Wavelet to obtain Wavelet coefficients;
(2) Process the threshold of wavelet coefficient to get new Wavelet coefficients;
(3) Reconstruction de-noised signal by the new Wavelet coefficient;

2.1   Select Threshold Function

Donoho has divided threshold into the threshold function into soft threshold and hard threshold, Assume that w is the size of wavelet coefficients, ^w is the size of wavelet coefficients, λis the threshold value.

(1) Hard threshold

When the absolute value is less than the given threshold Wavelet coefficient value, assume Wavelet coefficient as 0, otherwise, the Wavelet coefficient is invariant, which is showed in Fig 1(a).

$$\varpi = \begin{cases} w, |w| \geq \lambda \\ 0, |w| < \lambda \end{cases} \tag{1}$$

(2) Soft threshold

When the absolute value is less than the given threshold Wavelet coefficient value, assume Wavelet coefficient as 0, otherwise, minus the threshold, which is showed in Fig 1(b).

$$\varpi \equiv \begin{cases} \text{sgn}(w)(|w| - \lambda), |w| \geq \lambda \\ 0, |w| < \lambda \end{cases} \tag{2}$$

(a) Hard threshold   (b) Soft threshold
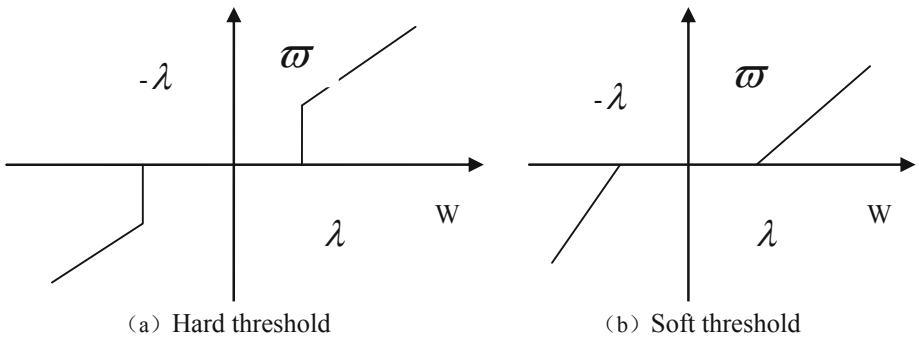
Fig. 1. The soft and hard threshold method of estimated Wavelet coefficient

2.2   Select Threshold

How to choose a threshold value in the threshold function is a key problem that is related to de-noising directly. If the threshold is chosen too small, some noise will be saved after the signal de-noising; If the threshold is chosen too big, it will cause distortion. So it is a key problem to select threshold while de-noising. Donoho has given the threshold value based on orthogonal Wavelet transform $\lambda = \sigma\sqrt{2\ln^N}$, where n is the length of the signal, σ is the noise standard deviation, which can be evaluated by the following equation.

$$\sigma = \frac{\text{median } (|d_j(k)|)}{0.6745} \tag{3}$$

Where dj(k) is the high frequency coefficient after the Wavelet Decomposition, median is calculating intermediate values of operation command in MATLAB. The threshold selection method results are unsatisfactory in the practical application, the following guidelines can be chosen:

(1) Fixed threshold
The selected algorithm is:

$$\lambda = \sqrt{2\lg N} \tag{4}$$

Where N is the signal length.

(2) Stein's unbiased likelihood estimation for threshold values (rigrsure threshold)
This calculated SURE threshold is based on Stein's unbiased maximum likelihood estimator, for each threshold value, find out its corresponding value-at-risk, and then choose a threshold to make the risk threshold values to be smallest.   Specific algorithms are:

$$P = [P_0, P_1, ...P_{n-1}], P_0 \prec P_1 \prec ... \prec P_{n-1} \tag{5}$$

The elements of P are square of the wavelet coefficients, which is from small to large order. Risk algorithm is:

$$R(K) = [N - 2K - (N - K)P_K + \sum_{I=1}^{K} P_i]/N \tag{6}$$

Where k=0,1,…, N-1. According to the resulting risk curve R(k), Mark their minimum corresponding value to be kmin,Then threshold is defined as:

$$\lambda = \sigma \sqrt{pk_{min}} \tag{7}$$

(3) Heuristic threshold
It is synthesized of the former two thresholds, what is chosen is optimal prediction variable threshold. When the ratio of the signal and noise is small, a fixed threshold is adopted; otherwise, using rigrsure norm.

(4) Max and minimum threshold
Its principle is to make the estimated maximum risk minimization, the threshold selection algorithm is:

$$\lambda = \begin{cases} \sigma[0.3936 + 0.1829(\ln N / \ln 2)]N \succ 32 \\ 0 N \leq 32 \end{cases} \tag{8}$$

(5) Penalty threshold based on wavelet packet transform

Wavelet packet analysis on low-frequency signal decomposition, on high-frequency signal decomposition, it has a more accurate analysis capability. Principle of de-noising based on wavelet packet analysis more or less agrees with Wavelet analysis. The key problem is to select threshold, the following are penalty threshold value method.

Wavelet packets coefficients sort in the order from small to large, $C = [C_1, C_2, ...C_n]$. Supposed function: $\text{crit}(t) = -\sum_{k \prec t} c_k^2 + 2\sigma^2 + (\alpha + \log(n/t))$,

which t = 1 ,2 , ···, n, n is the number of wavelet cofficents, α is experience coefficient whose value must be greater than 1,and the typical value is 2. Take t as variable, calculate the minimal value of crit (t). Set the crit(t) for the minimum value of t for t 0, the function λ= | ct0 |

3   The Application of Wavelet Analysis in GPS Signal Noise

Each GPS satellite launched signal which can be distinguished, including carrier signals, P-code (Y-codes), C/A code, data code (D-code) and other signal component, that all these signal components are produced on the basis of the same fundamental frequency f0 which is 10.23MHz. P-code and the C/A code collectively are known as the ranging codes, whose carrier signals include L1 and L2 carrier, frequency 1575.42MHz and 1227.60MHz respectively. On the carrier frequency L1, the modulation includes C/A code, P-code and D-code, however on the frequency L2 the modulation P-code and D-code only. Both L1 and L2 carrier is a sine wave, and the frequency of P-code is 10.23MHz, the sequence of which the length of an element is 100ns, corresponding to 30m distance. When there is no other error, the measurement accuracy on P-code can achieve over a meter, and whose cycle is 267d. C/A code is much simpler than the P-code, which frequency is 1.023MHz, corresponding to the wavelength of 300m. Due to the C/A code cycle of 1ms, it could be locked by receiver soon. The work procedures of P-code also firstly lock the C/A code by receiver, worked out the system information, then switched to P-code precision pseudo observations [1-2].

Ranging codes and D-code of GPS satellites, are modulated to a carrier wave with the technology of phase modulation, and the amplitude of modulation code only is 0 or 1. If the value of code is 0, take the code corresponding to the State as +1. And the value of code is 1, the corresponding status codes get -1. Then, after the carrier multiplied by the appropriate status code, it enables carrier modulation, that is, the code signal is onto a carrier wave.

GPS technology is inseparable with the pseudo code. GPS receivers use pseudo code received signal, to identify the GPS satellites. In precision GPS phase measurements,

solve phase ambiguity with the use of pseudo code. There are many errors source in pseudo measurement, such as ionosphere refraction, troposphere refraction, multipath effect, time difference effect, receiver noise and so on. On the ionosphere refraction and troposphere refraction, there already are many technologies or models to eliminate and correct. Time difference effect also can eliminate well with certain model. But it is different to multiple paths effect, because it gets different multiple paths effect due to receivers or locals. Multipath effects in GPS data processing cannot be ignored, which is also an important error source in static GPS precise positioning besides receiver noise.

Because measurement noises of GPS receiver and multipath delay errors are noise signals of GPS phase observations, from wavelet detection signal analysis, they can consider on the amplitude of the mutation, are stationary signals. In GPS signals, the useful frequency of signal could include signal launch frequency, data sampling frequency, data solutions frequency and so on. As which kind of signal is the useful signal what we need of, it depends on we prepared to extract the information from data, so the frequency range of the information is just the interested and useful frequency. The information which go beyond the above frequency range, can be considered useless information or noise. From the perspective of pure frequency, various measurement noise and multipath effect signals are in a certain frequency range. The time-frequency characteristics between useful signal in the GPS observation sequence and noise are usually not same. In time domain and frequency domain, the useful signals are localized, characterized in low frequency or stationary signal. But, observation noise and multipath effects distribute globally in time-frequency spaces, existing everywhere in the whole observation time domain, and in frequency domain performance as high frequency signal. In observation of GPS relative positioning, access to the useful signal a steady signal, appeared in the low-pass filtering of results, and various noises are reflected in the high-pass filtering results.

GPS signal noise processing with Wavelet analysis, firstly should get one dimensional sequence carrier signal. Due to the different data formats designed by GPS receiver manufacturers, GPS raw data should be transferred from the receiver to the computer, and then take original observation data file to form GPS RINEX observation files in standard format with the random software. This study exploits a GPS signal processing software with MATLAB and VC++ language. The software firstly extracts the original carrier signal from observation data in the format of the RINEX, and makes the difference process on one dimension original carrier sequence signal to form one-dimensional single and double observation sequence signal, as the input signal of Wavelet decomposition. This article takes Wavelet Decomposition for GPS signals by the Wavedec function. The decomposition should reconstruct high-frequency signal by the specified wavelet functions, that is to say that it could de-compound the GPS signal into low frequency and high frequency. On account of the high frequency characteristic of the noise, analysis the part of high frequency decomposition to research the characteristics of GPS signal noise.

Figure 1 has shown GPS signal of the monitoring ground surface in Datong mining area, apparently the signal affected by noises. Using quadratic spline Wavelet, make

Wavelet transform on the four-level of binary scale, and the details reflected GPS signals changes at all levels.

Figure 2 is the curves of modulus maxima resulting from wavelet transform of GPS signal in all scales, which you can distinguish the modulus maxima formed by the effective GPS signal and noises. From Figure 2, as the scale increases, the modulus maxima of amplitude and thickness of noise wavelet transform rapidly decrease. When in large scale, the singularity of effective signal is gradually emerging, and the modulus maxima of amplitude about wavelet transform increases as the scale increases.

Figure 3 is the contrastive curve between the original GPS signal and the reconstructed GPS signal after de-noising which following the above process. The de-noising GPS signal is used in line adjustment of controller parameters, which ensured data accuracy finally.
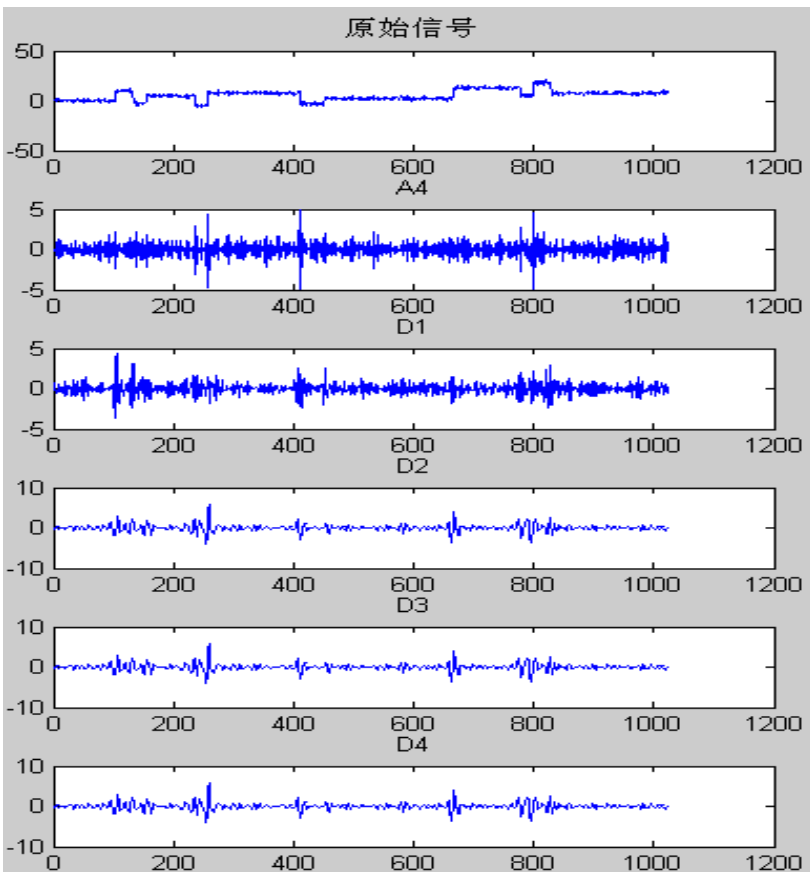
Fig. 2. GPS signal including chirp and wavelet transformation under each criterion
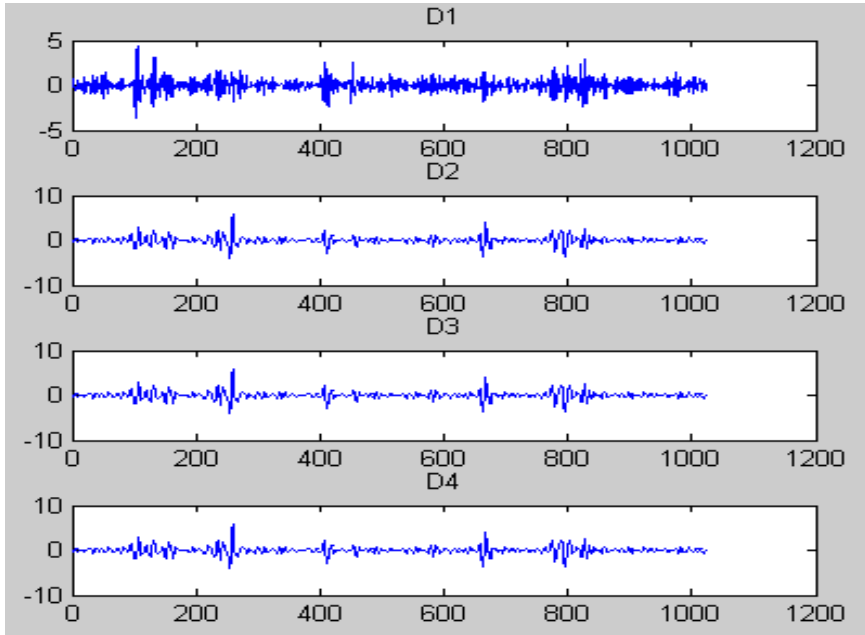
Fig. 3. GPS signal and mold maximum value curve under each criterion wavelet transformation
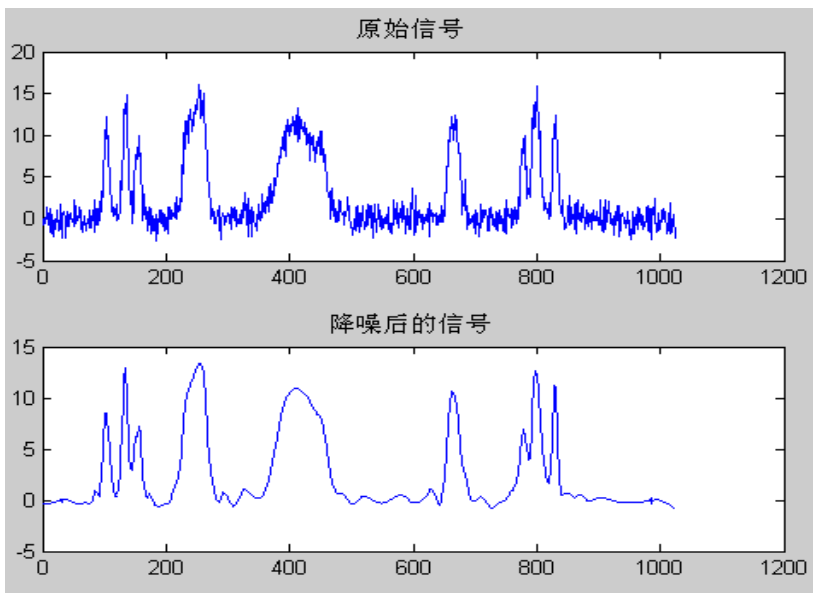
Fig. 4. Original signal and denoising signal of GPS

4   Conclusion

In the occasion of small signal-to-noise, this paper gives an de-noising algorithm, which provides the precise position of singular point of the original signal in strong noise background, this is the traditional spectrum analysis method cannot match. Besides, the impromptu algorithm of searching the great value saves calculation time but sometimes it is not too accurate. If certain part of original signal has min shaped structure, what means that there are many singular points in subsection of signal, extreme point in j level become j-1 level, and the candidate extreme points are more than one. So in the case, this method is not fit. Summary out some experiential criterion by specific debugging. Based on wavelet transform, the algorithm of alternating projection of modulus Maxima is an approximate recovery to the original signal. Most of the signal information is available and the recovery and reconstruction is in very high precision. In generally GPS signal processing, refactoring error is tolerable. The computing time of alternating projection algorithm is a little longer, so it could not meet real-time requirements. If a more efficient algorithm is established, it is of great significance.

References

[1]   Zhou, Z.: Principle and application of GPS satellite measurements. Press of Surveying and Mapping, Beijing (1997)
[2]   Liu, J.: Principle and method of GPS satellite navigation and positioning. Science Press, Beijing (2003)
[3]   Wang, X., Wang, D.: Discrete - time signal filtering. Publishing House of Electronics Industry, Beijing (2002)
[4]   Wen, L., Liu, Z.: Several methods of wavelet denoising. Journal of Hefei University of Technology(Natural Science) 25(2), 167–172 (2002)
[5]   Deng, D., Lin, C.: Wavelet Threshold Denoising Method Based on Neyman——Pearson Criterion. Signal Processing 19(3), 281–283 (2003)
[6]   Chen, Y., Wu, Z., Wu, Z.: Deformation analysis and prediction. Press of Surveying and Mapping, Beijing (1998)
[7]   Cui, X., Chen, L.: Dynamic subsidence monitoring and mechanical analysis of large deformation. Coal Industry Publishing House, Beijing (2004)
[8]   Liu, L., Xu, H., Sun, H.: Wavelet analysis method for gravity tide parameter determination precision. Science in China (Series D) 30(4), 442–448 (2000)
[9]   Lou, S., Shi, Y.: Systems analysis and design based on MATLAB –Neural Network. Xi'an Electronic Science and Technology University Publishing House, Xi'an (1999)
[10]   Daubechies, I.: The wavelets transform . Time-frequency localization and signal analysis. IEEE Tranns. on IT 36(5), 961–1005 (1990)
[11]   Liu, L.: Wavelet theory and its application in the fields of Geodesy. Chinese Academy of Sciences PhD thesis (1999)
[12]   Deng, N.: Spatial displacement of dam partial regression models and neural network model. Wuhan University PhD thesis (2001)

[13] Guo, J.G.: GLONASS combined measurement and research on deformation monitoring data processing. Wuhan University PhD thesis (2001)
[14] Guo, H.: Iterated Extended Kalman Filter Application to Real time GPS Data Processing. Journal of Wuhan Technical University of Surveying and Mapping (WTUSM) 30(4), 442–448 (2000)
[15] Tao, B.: Free Network Adjustment and Deformation Analysis. Press of Surveying and Mapping, Beijing (1984)
[16] Yang, Y.: Kalman Filtering for Dynamic System. Journal of PLA Institute of Surveying and Mapping (2) (1997)

Forest Cover Classification from Multi-temporal MODIS Images in Southeast Asia Using Decision Tree

Sijie Wu[1], Jianxi Huang[2], Xingquan Liu[1], and Guannan Ma[2]

[1] School of Geosciences and Info-Physics, Central South University, 410083, Changsha, China
[2] College of Information and Electrical Engineering, China Agricultural University, 100083, Beijing, China
{vr_hjx,dx01wusijie}@163.com

Abstract. MODIS data is of significant for the classification of regional forest cover due to its high temporal resolution and high spectral resolution. Forest cover is an important parameter for forest ecosystem. The objective of this preliminary study is to mapping forest cover from multi-temporal MODIS data with decision tree. The classification forest samples were selected from four global land cover datasets with specific rules. The selected samples were used to generate rules of the decision tree for the classification of forest cover. The study results show that multi-temporal remote sensing data with decision tree method have great potential to improve the regional forest cover mapping.

Keywords: Multi-temporal, forest cover, decision tree.

1 Introduction

Identification of types of forest has significant for forestry source monitoring and management. Because of the capability of acquiring regional surface information, remote sensing has become a reliable tool for identifying types of forest in regional and global scales. Currently, most of the applications of remote sensing classification are the traditional statistical pattern recognition method, such as minimum distance, parallelepiped, maximum likelihood, and mixed-distance method, cyclic cluster method and other supervision or unsupervised classification method. Because of the existence of spatial resolution of remote sensing image itself and "same object with different spectrum", "different objects with same image" phenomenon, misclassification and leakage of points occur more frequently. These factors lead to the low classification accuracy. New methods of pattern classification are as follows: fuzzy classification, classification based on texture description of Markov random field model, classification of wavelet analysis, fractal texture method, neural network and expert system classification, etc [1-3]. Currently, remote sensing information composite technique is widely used [4, 5]. In recent years, researchers primarily utilize satellite remote sensing with vegetation surface temperature, terrain elements and other non-remote sensing of forest vegetation on the ground to identify sub-categories [6].

In the past decades, applied research results on the large area forest cover mainly using AVHRR data has achieved great success[9,10]. Due to limitations AVHRR data

for land cover mapping applications, there are still many uncertainties [11]. Thus, with the launch of Terra satellite, the use of MODIS data in regional scale studies of forest cover has been developed. Using MODIS global supervised classification model, Muchoney [12] classified the vegetation and land cover in central United States. On this basis, with IGBP classification system and the STEP global plots database and MLCCA (MODIS land cover classification algorithm method. Friedl [13] carried out a global land cover classification with a total of five months of MODIS data in 2000. Based on pattern decomposition method (PDM), Cen[14] conducted a study of land cover classification using MODIS data of the Kii Peninsula of Japan in 2001. Using MODIS 8 day composite reflectivity products, Carrão [15] evaluated the efficiency of MODIS hyper spectral data and relative land cover classification for a long time. On the basis of studying and comparing the abroad classification algorithm that having good application effects, Wu [16] achieved quantitative judgments on continuous coverage of MODIS data and generated the land use status classification. Liu [17] proposed a classification method using MODIS data to select and extract classification feature and do large area land use/cover classification combined with the multi-temporal characteristics. Classification test was conducted in Shandong Province in China.

As for classification data, MODIS data can provide more data products. Visible data and near infrared data of the MODIS can response the growth characteristics of different vegetation types in different periods well, it is perfectly suited for forest classification. Taking into account of close contact of the distribution of vegetation with the climate and soil, coupled with influence of the climate by the altitude, slope and other terrain factors, it is necessary to add the soil and terrain data in the classification of forest types.

2   Study Area and Data

There are 11 countries in Southeast Asia: Vietnam, Laos, Cambodia, Thailand, Myanmar, Malaysia, Singapore, Indonesia, Brunei, and the Philippines, Timor-Leste. The special geographical location makes the Southeast Asia having hot and humid climate and lush tropical forests. Wet equatorial climate and a tropical monsoon climate are two types of the Southeast Asia. The main natural vegetation here is the tropical rain forest and tropical monsoon forest. Figure1 show the study area of Southeast Asia.

It can be divided into two sub-areas:

1. Indochina area: The climate here is tropical continental monsoon climate. The climate of Malay Peninsula is wet equatorial climate. Annual rainfall of the Malay Peninsula and the rainy coast of Indochina are tropical rain forest landscape. Indochina with dry and wet season is Tropical monsoon forest landscape. Less rainfall Interior plains and valley are savannah landscape. Indochina base is mountain mixed forest. Coast of North Bay and the Gulf of Siam is filled with mangroves.

2. Southeast Asian island district is also called the Malay Archipelago area. It belongs to maritime equatorial rainy climate. The Philippine islands belong to maritime tropical monsoon climate, mainly for tropical rain forest landscape. Southeast Asia

with the Indonesian forest area for first is the world's second-largest rain forest (after Brazil). Papua new several second is in the western Pacific after Indonesia's second-largest forest resources garden.

In Southeast Asia, land surface is mainly composed by woodland, grassland, residential areas, etc., among which forest cover is the most typical one. In order to improve classification accuracy in the region, a lot of land surface sample data is needed. Because the large scope of the study area and too many type and the far distance, it is difficult to achieve the field sampling of vegetation types. Through data analysis on the existing global surface types, we found the land cover of the typical areas changes less, and many classification products can separate the type of surface area of typical areas very well. Therefore, in areas where vegetation integrated we reference map and interactively select the sample data, and using ROI to define the surface type.

Fig. 1. Study area (Southeast Asia)

In the establishment of classification algorithms, including supervised and unsupervised classification are in common use. In order to improve the accuracy of classification, decision-making regression tree algorithm of supervised classification was used in this research. It is characterized by the better use of sample data and extracting decision to create classification algorithm according to the sample data. The user can understand the decision-making rules well.

Decision tree is a tree structure similar with binary tree or multi-branches tree. The decision tree algorithm is one of the inductive learning algorithms. Induction is getting

the regular conclusion by contrast, analysis, summary and summarization from characteristics and attributes of examples that known and seemingly with no specific order and rules. This conclusion was used to generate decision tree as the rule of interpretation, then classification was carried out on cases out of samples. Early famous decision tree algorithm was ID3 posed by Quinlan [19] in 1993. This is one of the most basic decision tree algorithms. C4.5 algorithm was posed by Quinlan [20] in 1993 as Subsequence of ID3 and became basis of later decision tree algorithms. SLIQ (supervised Learning in Quest) algorithm is a fast scalable decision tree algorithm posed by IBM researchers that suitable for processing the data of large scale. CART decision tree, also called classification regression tree, was a data survey and prediction algorithm posed by Breiman, Friedman, Olshen and Stone in 1984.

QUEST is a new decision tree construction algorithm is formulated by Loh and Shih in 1997 The algorithm separates the variable selection and split point selection, not only suitable for continuous variables, but also for discrete variables, and it overcome the other decision tree algorithms that prefer to choose those predictor variables with more potential split point for classification. So basically it is unbiased on variable selection. And through hyper plane constituted by more than one predictor variable, this decision tree can distinguish between multiple members and non-class members of the class in the feature space. Studies have shown that QUEST decision tree classification computation's speed and accuracy are better than other decision tree construction method.

3   Material and Method

3.1   Data

MODIS standard products which produced by this project were using to do forest classification. Taking into account the classification of forests and other land cover classification using vegetation status, MODIS 13Q products were downloaded, including visible light reflectivity. Considering the region's distribution of the forest influenced by the soil type and altitude, we collected global 10km soil data and STRM-DEM data and extracted the slope information. In order to maintain the consistency of the above data, the data project coordinate system and projection were normalized. Although the MODIS data was standardization, there are still clouds of pollution. Using the SG remove cloud algorithms, NDVI-SG Index data was generated. As the greenness of forests changed little during the growth in the 16 days, the index data was synthesized, that is, two 16-day month synthetic data was made into the index data. The 12th monthly indices data was instead by 23rd layer data. By integrating the above data, the input data for classification software was ready. When carrying out the classification of the data by adding soil analysis, marked polygons were found which classification results were not the same size as same as grid soil data. After analysis, it was found that this is due to their resolution (the MODIS 1/10) was inconsistent with the image data. In the later experiments, the soil data were excluded. Taking into account the amount of data and storage limits, in order to support the Bit types of data files, when storing data, using altitude / 20

which altitude below 5,000 m in the region. Degree of slope data used (Degree) units. NDVI data was set to 0 when less than 0; Greater than 0, the draw ratio is 250 times.

3.2   Forest Classification Method

Based on four global land cover datasets, the samples with pure pixels were extracted. MERIS with 300 meters was used for validate the accuracy of percentage of each sample. And then the overall classification accuracy was discussed.

3.3   The Choice of Sample Set of Forest Types

Five global Land Cover datasets, including UMD, IGBP-DIS Cover, MODIS, GLC2000 and MERIS 2005 Global Land Cover data were collected in our study. 1 km land cover map of the first four class of Southeast Asia was encrypted cut for classification of 500 meters. Combined with GLOBCOVER, a standard reference data for classification in this region was formed. Because these data in time is quite

Table 1. A brief description of 5 global land cover datasets

No	Data name	Data source	Data explanation
1	UMD Global land cover data	University of Maryland (http://www.geog.umd.edu/landcover/1km-map.html)	spatial resolution 1km the 1990s
2	IGBP-DISCover Global land cover data	International Geosphere Biosphere Program (http://edcsns17.cr.usgs.gov/glcc/globe_int.html)	spatial resolution：1 km the 1990s
3	MODIS Global land cover data	Boston iversity (http://duckwater.bu.edu/lc/datasets.html)	spatial resolution：1 km The early 21st century
4	GLC 2000 Global land cover data	European Commission, Joint Research Centre (http://www-tem.jrc.it/glc2000/)	spatial resolution：1 km The early 21st century
5	MERIS 2005 Global land cover data	CNES, CNRS, IRD, Météo-France, and INRA (http://postel.mediasfrance.org/en/DOWNLOAD/)	spatial resolution： 300m 2005-2006

different and system of classification is different, classification system of forest mapping in Southeast Asia which was provided by Chinese Academy of Forestry was using for combining land cover types, encoding conversion and forming a comprehensive map of forest types. When producing the integrated forest type's graphs, the types with relatively large probability of appearance were on behalf of the forest types in the region, after the adoption of uniform codes. In order to analyze differences in forest types, using red color to represent the type inconsistent, the white for the same type. Types can be seen change little from on higher ground area; types change more in the flat terrain and more human activities places. Using ENVI-ROI function, the sample data were collected in each typical area. Based on the sample analysis and the decision-making algorithm, regression analysis of data for decision-making was conducted to form the classification decision in the region. Figure2, figure3, figure4 and figure5 and figure 6 represent IGBP-DIS, MODIS, GLC 2000, UMD and MERIS forest cover map respectively. And figure7 show the integrated forest cover results.
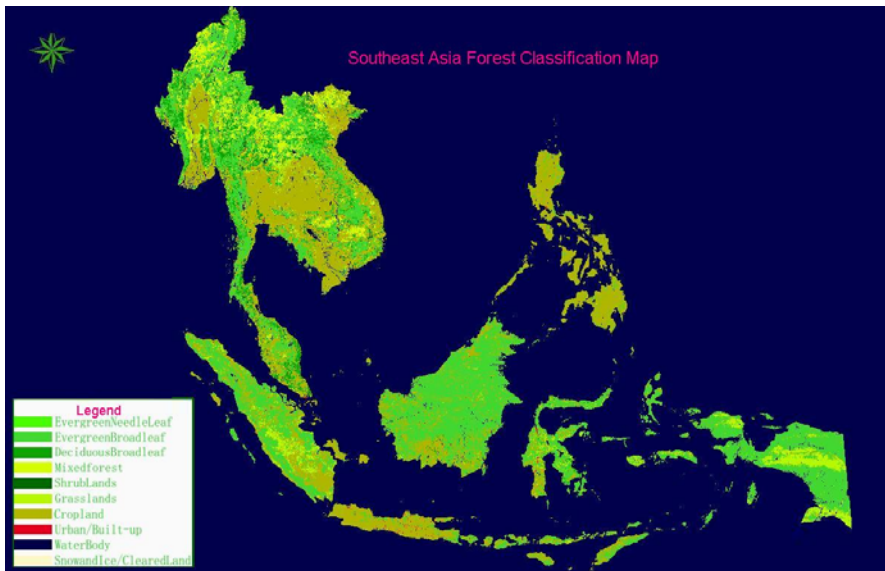
Fig. 2. IGBP-DIS Cover Global land coverdata
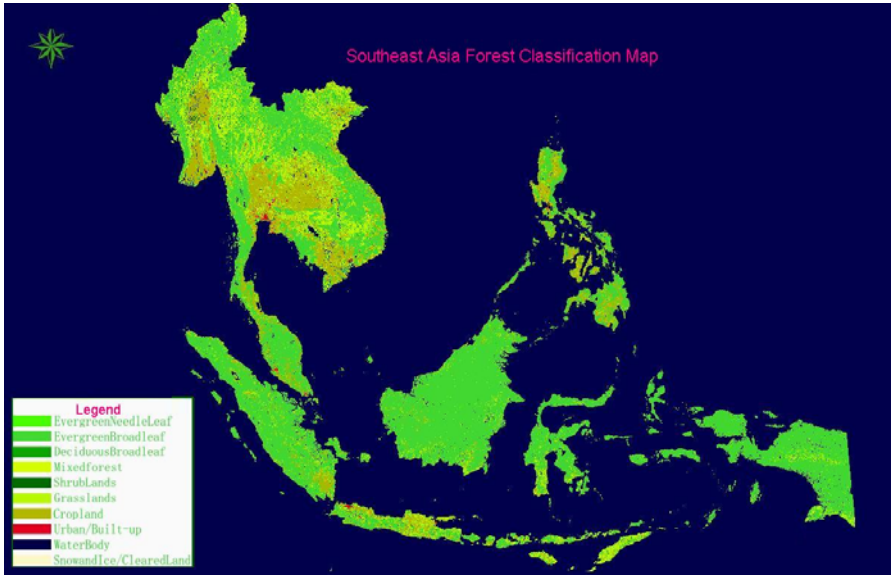
Fig. 3. MODIS Global land cover data
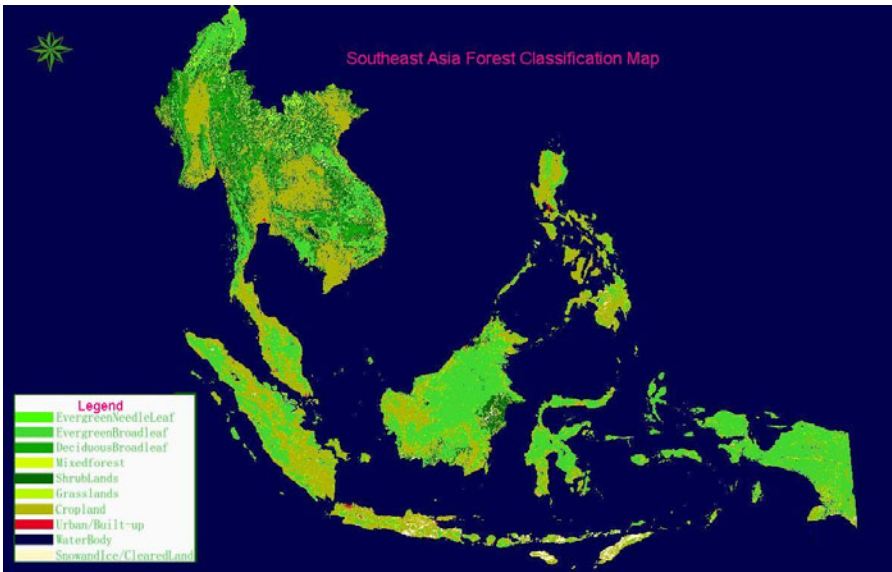
Fig. 4. GLC 2000 Global land cover data
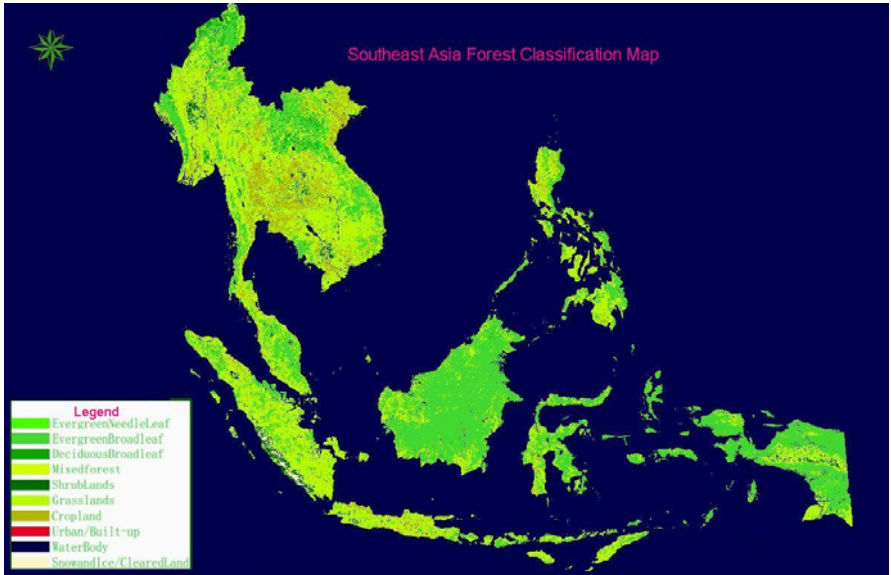
Fig. 5. UMD Global land cover data
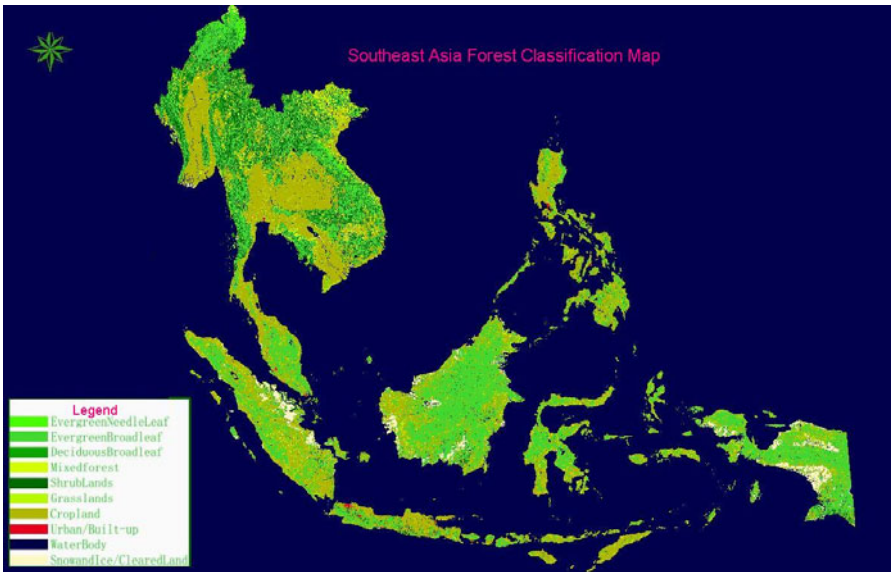
Fig. 6. MERIS 2005 GLOBCOVER
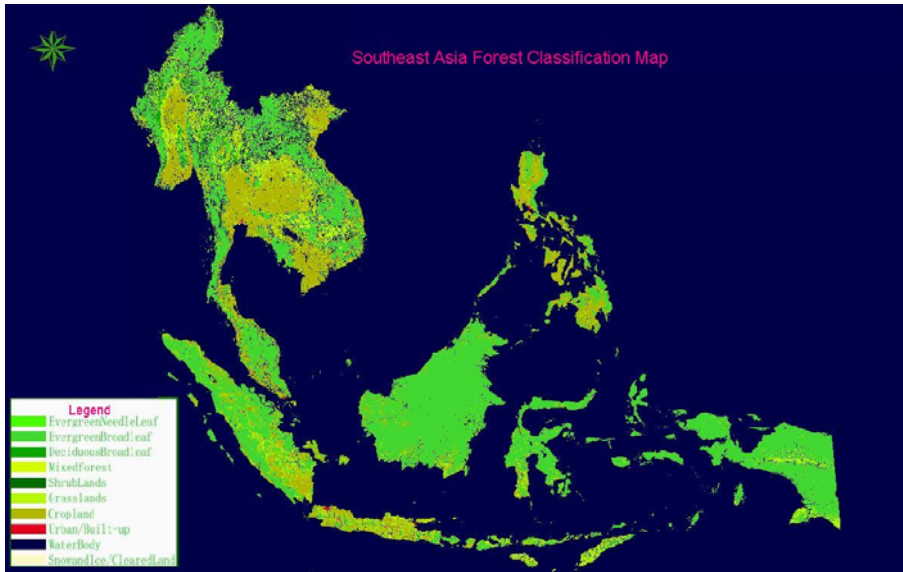
Fig. 7. Integrated results

3.4   Decision Tree Classification Algorithm

The key of the decision tree algorithm is to establishment decision rules through the sample points. There are several decision-making algorithms to establish the rules, but algorithms for massive remote sensing data classification have deficiencies. Especially there are few algorithms suitable for application in forest cover classification. Through the relevant literature and data on foreign inquiries, Quest algorithm has been selected to generate the decision tree of classification. The basic principle is as follows:

1)   First, predict variables are selected. Correlation between all predictor variables X and target variable Y is analyzed in turn. If X is discrete, the association strength between X and Y will be calculated using the chi-square test respectively, and find the P value; If X is ordered or continuous variables, P value will be calculated using analysis of variance.

2)   All the P values are compared to the value preset boundary (the default is 0.05). If they are less than the critical value, the minimum value of the variable P is selected as a branch variable; If they are all larger than the cutoff value, when X is continuous or ordered variables, P value will be calculated using Levene homogeneity test of variances, and when P value is less than the cutoff value, the minimum value of the variable P will be chosen as a branch of variables. If P values of homogeneity of variance test are large than critical value, the minimum value of the variable P in the first step will be the direct selection as the branch variable.

3) If the elected Branch variables are of discrete categorical variables, through the transformation, the goal variable Y will be maximized when X values are different. And largest discriminate coordinate will be calculated.

4) If Y is more classified categories, the mean number of X will be calculated for each type of Y value. These categories will eventually merge into two categories using the cluster analysis algorithm. This will simplify the multiclass discrimination problem to two types of discrimination.

5) Split point position is finally decided using quadratic discriminate analysis. And the selected original value of predictor variable X is got for building classification rules.

Based on the QUEST algorithm, the system combines the selection of the sample, decision analysis and classification. Firstly, using the sample data, samples were selected based on ROI selection tools in ENVI software. Then sampled were changed into data formats supported by QUEST, and QUEST algorithm interface was called by automated programs developed by IDL for establishing the decision tree automatically (suitable for established rules for ENVI decision data). Finally, classification was carried out using ENVI decision tree analysis tools to achieve judgments of the types of classification of forests in Southeast Asia. Classification accuracy and speed are greatly improved.

3.5   Sample Precision Analysis

Table 2. Accuracy analysis of selected sample points

Types	Evergreen needle leaf forest	cities and towns	Water body	Snow and ice + Bare land	Evergreen broadleaf forest	Deciduous needle leaf forest	Mixed forest	Shrub lands	Grass-lands	Crop-land
Evergreen needle leaf forest	59	0	0	0	0	0	0	2	5	4
Urban and built-up	1	75	2	0	0	0	0	0	1	2
Water body	0	0	28	0	0	0	0	0	0	0
Snow and ice + Bare land	0	2	0	48	0	0	0	3	0	0
Evergreen broadleaf forest	3	1	0	0	118	1	0	2	2	4
Deciduous needle leaf forest	0	0	0	0	0	62	0	0	2	0
Mixed forest	1	0	0	0	0	0	43	0	0	0

<div align="center">Table 2. (<i>Continued</i>)</div>

Shrub lands	0	0	0	6	3	0	1	80	0	1
Grasslands	7	1	0	0	1	6	0	2	119	4
Cropland	6	3	0	0	6	1	0	1	6	160
Classification accuracy	76.62%	91.46%	93.33%	88.89%	92.19%	88.57 %	97.73 %	88.89 %	88.15 %	91.43 %
Overall Classification accuracy	83.02 %									

Using QUEST algorithm, regression accuracy analysis of sample points established by decision tree is showed as follows. The accuracy is better than 75%, and the average accuracy reaches 83.02%. The above results only show that the regression accuracy of samples is high, but cannot represent the actual accuracy. The classification results are needed to be analyzed.

4   Results

In order to fully use the data features in the data processing, four schemes have been taken into account: scheme 1 implement the SG cloud-remove processing, and
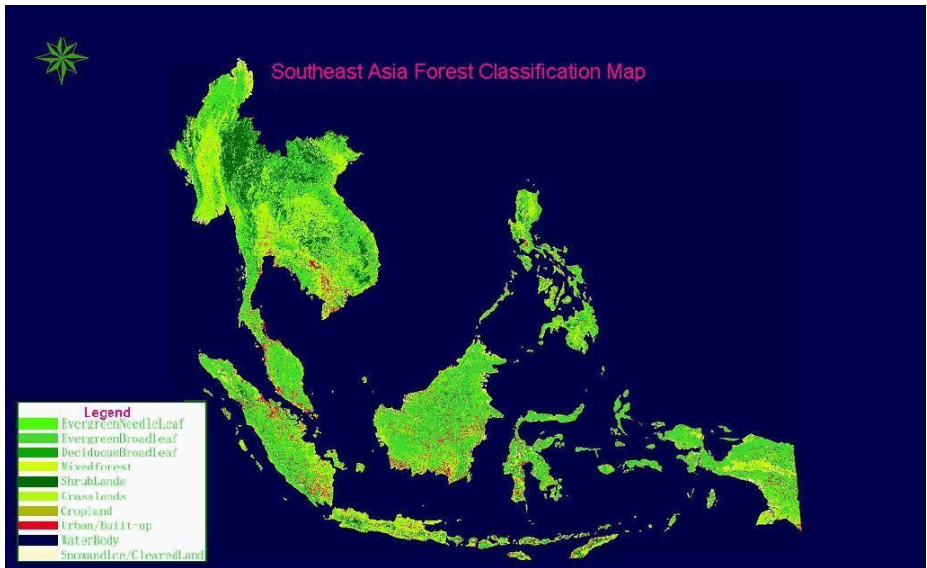
Fig. 8. Forest classification result with SG cloud-remove processing and months synthesis using maximum NDVI values
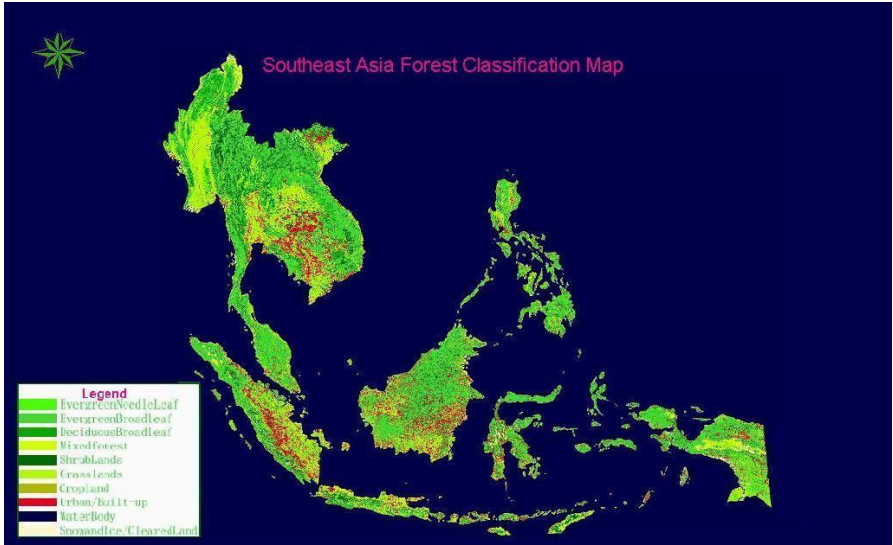
Fig. 9. Forest classification result with   SG cloud-remove processing and months synthesis using average NDVI values
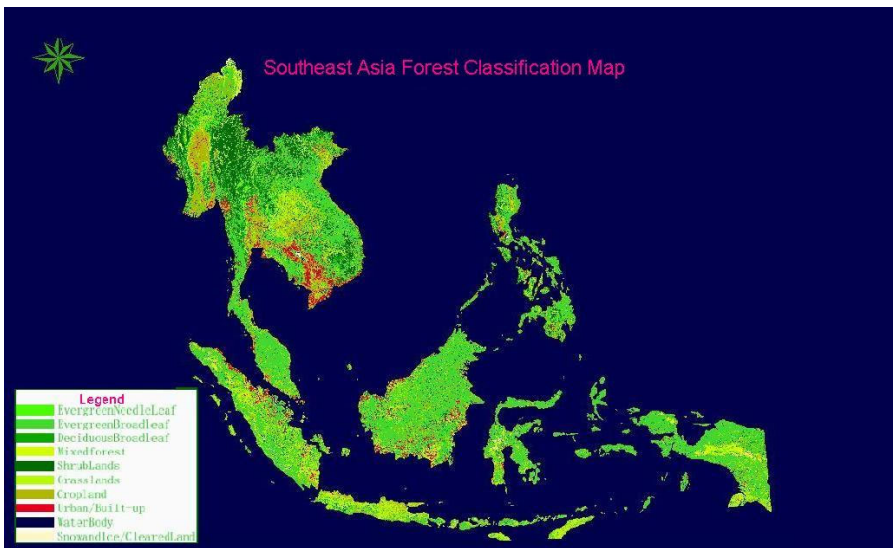
Fig. 10. Forest classification result with month synthesis using maximum NDVI values
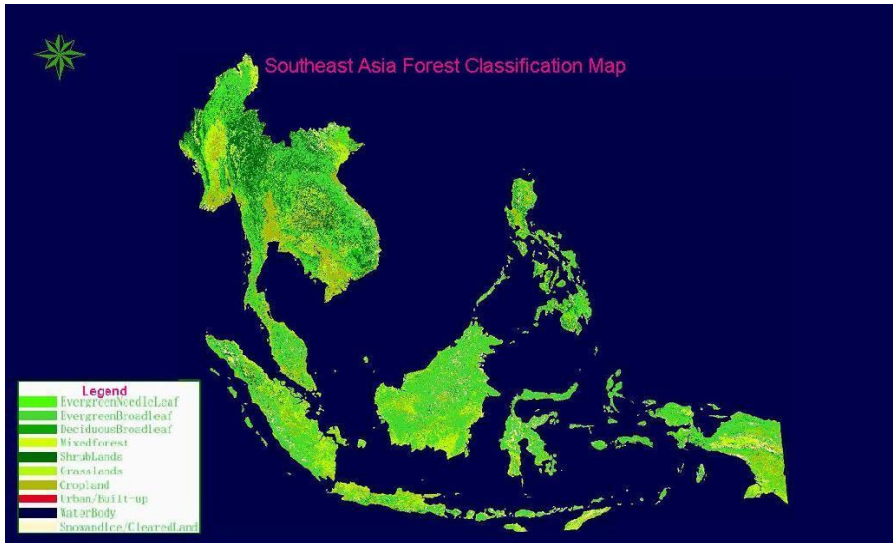
Fig. 11. Forest classification result with month synthesis using average NDVI values

month-synthesis used maximum NDVI values (SGMaxLANDTYPE); scheme 2 carry out the SG cloud-remove processing, and month-synthesis used average NDVI values (MeanLANDTYPE); scheme 3 without implementing SG cloud-remove processing, and month-synthesis used maximum NDVI synthesis (SGMaxLANDTYPE); scheme 4 without implementing SG cloud-remove processing, and month-synthesis used average NDVI values (MeanLANDTYPE). Data samples were collected in the same position of sampling pointon the above to construct different trees, and carry out the forest cover classification with quest decision tree algorithm. Classification results are shown in figure 8 to figure11:

Since 2005 global land classification data (GLOBCover2005) is the update data and its spatial resolution is better than 500m, it can be used as reference data for the accuracy verification of classification results of this project. In order to evaluate the classification accuracy, the results of 4 categories of data were compared with the GLOB Cover 2005. The accuracy of MaxLANDTYPE, MeanLANDTYP, SGMaxLANDTYPE ,SGMeanLANDTYPE are 65.05%,66.19%,76.66%, 78.14% respectively. It can be seen that high precision can be reached using SG cloud-remove processing and month-synthesis using average NDVI values (more than 78%). Meanwhile, the proportion of each category and the proportion of surface vegetation in each country were statistically. Table 3 indicates the results. With analysis on the proportion of different vegetation in different countries, the results showed SGMean LANDTYPE classification has achieved the highest accuracy.

Table 3. The distribution table of the proportion of different vegetation in Southeast Asia

No.	Types	SGMax-LANDTYPE	SGMean-LANDTYPE	Max-LANDTYPE	Mean-LANDTYPE	GLOBCover2005
1	Evergreen needle leaf forest	4.10	7.44	8.89	5.37	0.97
2	Evergreen broadleaf forest	40.63	38.61	35.73	37.29	35.00
3	Deciduous broadleaf forest	4.95	3.13	2.75	4.87	1.80
4	Mixed forest	1.91	1.53	1.96	1.79	6.41
5	Shrub lands	12.40	14.00	12.07	11.28	10.69
6	Grasslands	14.42	14.43	30.34	24.76	0.33
7	Cropland	16.22	15.14	0.00	0.00	39.56
8	Urban and built-up	4.46	0.00	5.11	11.31	0.15
9	Water body	0.31	0.31	0.20	0.21	2.06
10	Snow and ice + Bare land	0.58	5.39	2.96	3.12	3.02

5 Discussions

MODIS provide the multi-temporal forest cover information. Decision tree has advantage of understandable in structure, interpreted by its rules, calculating and precise in getting result, which lead to its high development in the field of forest cover classification. Multi-temporal MODIS-NDVI data with SG filtering have been used to the forest cover classification. We compare forest cover classification accuracy with the four schemes for time series MODIS-NDVI processing using SG cloud remove algorithm using Quest decision tree method. The results show that SG Mean LANDTYPE classification has achieved the highest accuracy. It can be seen that the decision tree algorithm with multi-temporal MODIS has great potential in regional forest cover mapping.

References

1. Huang, N., Liu, X., Zhu, M., Zhang, S.: Remotely Sensed Image Classification Technology Review. Journal of Test and Measurement Technology 15(2), 86–92 (2001) (in Chinese)
2. Li, S., Ding, S., Xu, S.: Study of Remotely Sensed Image Classification Technology. Journal of Henan University(Natural Science) 32(2), 70–73 (2002) (in Chinese)

3. Wu, J., Yang, X.: Purification of Training Samples in Supervised Classification of Remote Sensing Data. Remote Sensing for Land & Resources 26(1), 36–41 (1996) (in Chinese)
4. Chen, S., Zhao, Y.: Remote Sensing Analyst. Surveying and Mapping Press, Beijing (1990)
5. Zhao, Y.: Principles and methods of analysis of remote sensing applications. Science Press, Beijing (2003)
6. Liu, X., Zhang, X.: Research Advances and Countermeasures of Remote Sensing Classification of Forest Vegetation. Forest Resources Management 1, 61–64 (2004) (in Chinese)
7. Hansen, M., Dubayah, R., Defries, R.: Classification trees: an alternative to traditional land cover classifiers. International Journal of Remote Sensing 17(5), 1075–1081 (1996)
8. Defries, R.S., Hansen, M., Townshend, J.R.G., et al.: Global land cover classifications at 8 km spatial resolution: the use of training data derived from Landsat imagery in decision tree classifiers. International Journal of Remote Sensing 19(16), 3141–3168 (1998)
9. Hansen, M., De Fries, R., Townshend, et al.: Global land cover classification at 1km resolu-tion using a decision tree classifier. International Journal of Remote Sensing 21, 1331–1365 (2000)
10. Kressler, F.P., Steinnocher, K.T.: Detecting land cover changes from NOAA 2AVHRR data by using spectral mixture analysis. International Journal of Applied Earth Observation and Geo information 1(1), 21–26 (1999)
11. Loveland, T.R., Zhu, Z., Ohlen, D.O., et al.: An analysis of the IGBP global land cover characterization process. Photogrammetric Engineering and Remote Sensing 65(9), 1021–1031 (1999)
12. Muchoney, D., Borak, J., Chi, H., et al.: Application of the MODIS global supervised classification model to vegetation and land cover mapping of Central America. Remote Sensing 21(6&7), 1115–1138 (2000)
13. Friedl, M.A., Mclver, D.K., Hodges, J.C.F., et al.: Global land cover mapping from MODIS: algorithms and early results. Remote Sensing of Environment 83, 287–302 (2002)
14. Yi, C., Zhang, L., Furumi, S., et al.: Land cover classification in Kii Peninsula of Japan using Terra/MODIS data. In: Proceedings of the SPIE Third International Symposium on Multi-spectral Image Processing and Pattern Recognition, vol. 5286, pp. 692–695 (2003)
15. Carrão, et al.: Contribution of multispectral and multi-temporal information from MODIS images to land cover classification. Remote Sensing of Environment 112, 986–997 (2008)
16. Wu, F.: MODIS Data Land Use Classification. China Academy of Forest (2005)
17. Liu, Y., Niu, Z.: Regional Land Cover Image Classification and Accuracy Evaluation Using MODIS Data. Remote Sensing Technology and Application 19(4), 217–224 (2004)
18. Quinlan, J.R.: Induction of decision trees. Machine Learning (1), 81–106 (1986)
19. Quinlan, J.R.: C4.5 Programs for Machine Learning. Morgan Kaufmann, San Mateo (1993)
20. Mehta, M., Agrawal, R., Rissanen, J.: SLIQ: A Fast Sealable Classifier for Data Mining. In: Apers, P.M.G., Bouzeghoub, M., Gardarin, G. (eds.) EDBT 1996. LNCS, vol. 1057, pp. 18–32. Springer, Heidelberg (1996)

Large-Scale Microwave Remote Sensing of Retrieving Surface Multi-parameters Using Active and Passive Satellite Data: In the Tibetan Plateau Region of Maqu

Ruofei Zhong[1], Jianxi Huang[2,*], Jiangxia Wei[1], Qin Li[1], Jiao Guo[1], and Wei Su[2]

[1] Capital Normal University Key Lab of 3D Information Acquisition and Application, MOE, Beijing 100048, China
zrfgis@263.net
[2] College of Information & Electrical Engineering, China Agricultural University, Beijing, P.R. China, 100083

Abstract. To conduct with these land surface parameters inversion using microwave observations in the bare soil surface, it is needed to estimate soil moisture (SM), surface temperature (ST) and surface roughness changes with microwave observations. High-frequency passive microwave radiometer sensitivity of the roughness is very low, traditional ground truth can't provide an accurate large-scale roughness of pixel information, and active radar and scatterometer data for roughness of the high sensitivity, active and passive joint inversion of soil volumetric water content is the current hot research. The main objective of this research is to develop a suitable method instead of some traditional methods for the retrieval of SM and other parameters at large-scale, which is based on the synergistic use of AMSR-E and Quikscat/SeaWinds observations. Quikscat / SeaWinds can provide large-scale scatterometer data, first used in this study to establish AIEM simulation and backscattering coefficient of roughness between, and then estimate the roughness of the known information as auxiliary inversion surface AMSR-E temperature and soil volumetric water content. The retrieval results show that this proposed method is helpful to achieve a higher accuracy in the study region of Maqu.

Keywords: AMSR-E, QuikSCAT/Seawinds, Surface temperature, *Soil moisture*, Roughness.

1   Introduction

Bare ground or sparse vegetation in surface conditions, AMSR-E inversion of soil surface roughness and time-sharing need to consider the effect of temperature. Traditional methods for estimating the roughness of the signal for the major impact of microwave radiation is used calibration method or empirical method, but the drawback of the calibration methods influenced by the sample with the measured roughness no definite link, although the empirical method related to the measured

---

roughness, but the local applicability. AIEM model simulation analysis can be used to know: when the frequency similar to the initiative of the roughness of the sensitivity of the microwave signal was significantly higher than the passive microwave signal, active and passive with a more accurate estimate of roughness, which could improve the accuracy of soil moisture. Inversion of a variety of active and passive sensors SMAP soil moisture is the core mission plan.

The current study of active and passive joint inversion of soil moisture remote sensing data sharing, targeted active and passive sensors are mostly equipped with a satellite, such as TRMM, PALSAR, etc. (Wu, etc., 2006; Chen, etc., 2008), mainly in order to obtain synchronous data. Equipped with different sensors on the satellite joint inversion of the current post less, mainly due to different satellites through the same area of the transit time differences (in hours), in this time interval the surface with a pixel parameter information has changed. In addition, despite recent high spatial resolution of radar data inversion for soil water into a hot spot, but the large-scale spatial and temporal distribution of soil moisture monitoring for global change research still has an irreplaceable position. ZHONG et al (2006) with the AMSR-E and Quikscat / SeaWinds joint inversion of soil moisture research work directly to the scattering of dollars to a VV polarization channel radiometer with two low-frequency channels in the multi-input inversion parameters of the inversion, ignoring the two sensors is not at the same time to obtain the surface information, and the results show that soil moisture sensor for two different sensitivity in the AMSR-E on the basis of observations by adding Quikscat / SeaWinds of channel has not improved with the iterative inversion accuracy. Some researchers are using experimental data analysis Quikscat / SeaWinds on soil moisture sensitivity, they noted that: Quikscat / SeaWinds Only in very dry and very wet soil conditions are more sensitive to soil moisture, and the results of extracting soil moisture information still not ideal (Scipal etc., 2002; Mladenova, etc., 2009; Zyl, etc.). These studies show that: the above two use Quikscat / SeaWinds collaborative way and can not get a good retrieval of soil moisture inversion.

Not a long period of time, except for special circumstances such as rainfall, that the surface roughness remains constant, this study based on this proposed new program: First, use the AIEM model simulations Quikscat / SeaWinds two different incidence angles different polarized backscattering coefficient and the roughness of the relationship, and then extract the roughness of the information and provide it to AMSR-E, further inversion of soil moisture and surface temperature.

2   Study Region and Data

This study area latitude and longitude of 33 ° 30'-34 ° 15'N, 101 ° 38'-102 ° 45'E, ground-based data acquisition time for July 2008 to August, on-board data selection is Quikscat / Seawinds and AMSR-E passive microwave sensors to specific areas of Qinghai-Tibet Plateau to obtain the joint surface soil moisture data, Maqu mostly open grassland areas(Fig.1), vision, set in Maqu region collected 20 soil sites, (Fig.2) collecting data, including temperature , soil moisture, soil texture and so on.

As the sensor is relatively large field of view, this research work will be on a pixel is within the scope of the average number of sites, just remove the anomaly is equal to

the minimum or maximum value, the last for the entire study area for all the pixels as a whole is to the simulation, choose the model chosen for the Qp model of radiation scattering model AIEM model, based on data measured on the ground to complete the forward simulation, and compared with actual observations.

Fig. 1. Typical landscape Maqu region: mountain (upper left), grassland and bare soil (upper right), Yellow (medium), Yellow (bottom left), wetlands (bottom right)
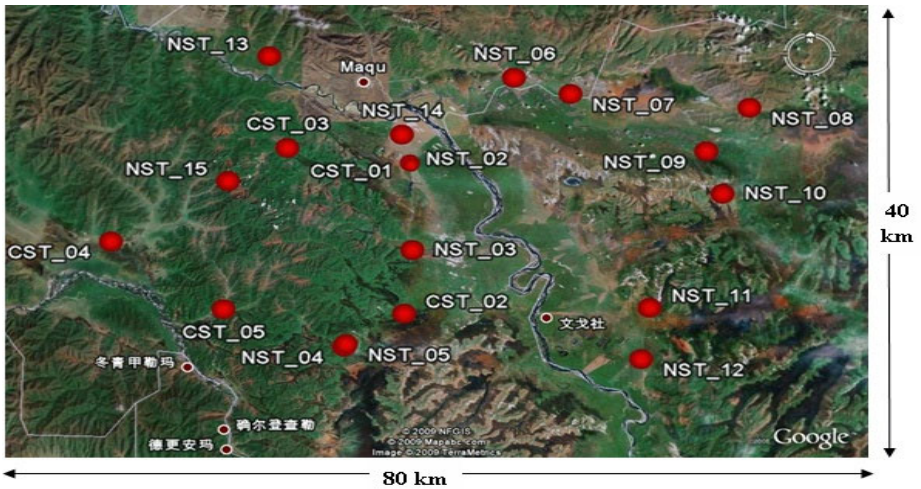
Fig. 2. Maqu region twenty maps of soil moisture collection site

According to this article, active and passive joint inversion scheme, first, using simulation databases established the relationship between the ratio of RMS height and the scattering coefficient and the difference of scattering plan two channels decibels, and then use Maqu region HH and VV scattering meter data to estimate the joint roughness slope information, as known parameters input to the microwave radiation model, which complete inversion of soil moisture. Here in order to eliminate measurement noise, the use of AMSR-E data in the 6.6GHZ frequency and 10.9GHZ

two four-channel inversion, inversion of surface parameters to soil moisture and surface temperature inversion iterative algorithm for the LM .

3   Physical Models and Retrieval Algorithm

3.1   Applicability of the Radiation Model

Used in this study taking into account the characteristics of active and passive sensors, and according to Shi Jiancheng's research team conducted in recent years in the Tibetan Plateau soil moisture inversion of the published research literature and related information, choice of radiation model for the Qp model, scattering model AIEM model.

AIEM model does not consider part of coherent radiation or scattering, but because of Quikscat / Seawinds the angle of incidence of 46 and 54.1, relatively large, even in large areas of undulating terrain, it is hard for the local angle of incidence close to 0 °, so the coherent scattering from formula to analyze, you can not consider it part of coherent scattering. Therefore, the model can be used to establish AIEM Quikscat / Seawinds analog observation database.

In the inversion of passive microwave soil moisture for bare surface of a variety of semi-empirical model, Shi and other articles will be in 2006, divided into three categories: Q/H model (1), Hp model (2) and Qp model (3). They form are:

$$R_p^e = [1 - r_{op} + Qr_{op}] \cdot H \tag{1}$$

$$R_p^e = r_{op}H_p \tag{2}$$

$$R_p^e = [\left(1 - Q_p\right)r_{op} + Q_p r_{op} \tag{3}$$

You can see, Hp model is Q / H model is based on the hypothesis Q = 0 obtained, Qp model at Q / H model based on the hypothesis H = 1 to get. Qp model of Hp compared with Q / H model, the advantage is that they considered the different polarization by roughness on the reflectivity of the impact.

In AMSR-E parameter conditions, with AIEM model to bare landmark Fresnel reflectivity research, there are two points found:

1. The effective reflectivity of surface with the frequency changes are caused by water in different frequency dielectric constant differences ,but the roughness at different frequencies to the effective reflectivity of the image is relatively stable, does not change with increasing frequency and strong .

2. Compared to the flat case, H polarization in the rough surface will reduce the reflected energy, or the emission energy increases. In other words, the rough surface effective reflectivity H polarization is always lower than the corresponding smooth surface of the Fresnel reflectivity. But V polarization is opposite.

So it can be a direct result of the two as follows:

1.   You can use a frequency the roughness of the parameters of the other frequency rough parameters.

2. Polarization of the surface effective reflectivity of the rough calculation there are effects, not on H, V polarization of the two kinds of roughness parameters equated.

Then, Shi and others with AIME, respectively, and Q / H model, Hp model and the Qp model 10.6GHz, 55-degree angle under the conditions of the simulation, and compare the results.

In the three semi-empirical model simulation results, Qp model and the best fit model results AIME, Q / H model followed, Hp model the worst. Investigate is reason, Hp model did not consider the error from the incoherent part, and in the case of high-frequency, non-coherent part of the effective reflection on the surface of the dominant. Q / H model error from the H, V two polarization roughness parameters under the same treatment.

Therefore, in this study, the use of parameters and roughness parameters Qp (s / l) relationship with the Qp model to simulate AMSR-E forward radiation reflectivity, enabling researchers to use a simple, single parameter roughness of the surface effective reflectivity model simulation and inversion.

3.2 Scattering Parameters of the Model Sensitivity Analysis

AIEM model can be used to simulate the exposed surface of the scatter meter data, but also to simulate bare surface radiometer data. Through model simulations, can be found: in the soil moisture, surface temperature, surface roughness and other parameters, the scatter meter data for roughness of the most sensitive, the sensitivity of other surface parameters is small; radiometer is most sensitive to soil moisture, surface temperature of the second, less sensitive to the roughness. Soil type and other soil parameters affecting radiation scatter meter data, but little impact.

Also, because the number of surface parameters, changes over time, soil moisture, surface temperature was greater than that roughness is essentially the same, in this study the use of active and passive microwave sensors to obtain data in the study area are not synchronized, Therefore, with the actual situation, the study of passive joint inversion method of soil moisture: the first with a scatter meter to estimate roughness slope, then the inversion of radiometric information with other surface roughness parameters, including soil moisture.

4 Datainversion

4.1 Forward Modeling and Calibration of Radiation

According to previous analysis of the applicability of the model, this study used a direct simulation of the model AIEM simulate the observations of scatter meter, the inversion may not process the general operation of the "model parameter calibration" work.

Simulated on a single site, there is insufficient data to use, and the method that points represent the surface of the simulation is not science. The research work will be on a pixel is within the scope of the average number of sites, just remove the anomaly equal to the maximum or minimum value, and finally the entire study area for all

pixels whole forward simulation by the entire study area all the pixels in the four channels on the forward simulation results (Fig.3).

Fig. 3. The study area all pixels in the four channels on the forward simulation results
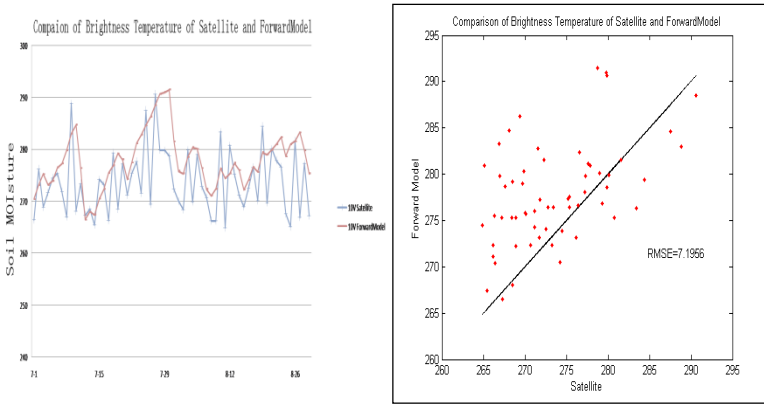
Fig. 3. (*Continued*)

4.2   Scatter Meter to Estimate Roughness

Quikscat / SeaWinds which was launched by FAA work in the Ku band, is different from operating frequency of ESA's wind scatter meter (ESA C-band).In this study, AMSR-E and Quikscat / SeaWinds combined inversion process, the bare surface conditions in the reflectivity of the passive and active backscattering coefficient of each surface parameter sensitivity analysis and found that Quikscat / SeaWinds on the surface the sensitivity of surface roughness parameters was significantly higher than the other, and AMSR-E on the surface roughness of the sensitivity is very low, so try to combine such a passive way: the use of Quikscat / SeaWinds inversion of surface roughness, and then AMSR-E soil moisture and surface temperature inversion. Here is AIEM model simulation Quikscat / SeaWinds bare surface conditions in the backscattering coefficient of a database to analyze its relationship with surface roughness, trying to build a function.

AIEM model simulation with Quikscat / SeaWinds scattering database parameters in Table 1:

Table 1. Input parameters setting for AIEM

Sensor	Frequency/(GHz)	Incient angle and polarization	bd/(g/cm^3)	Surface correlation function
QuikSCAT/ Seawinds	13.4	46°HH, 54°VV	1.3	Gaussian
ST/(°C)	**SM/(m^3/m^3)**	**cl/(cm)**	**sig/(cm)**	**Soil texture**
Range: 0~40, Interval：5	Range：0.02~0.44, Interval：0.02	Range：2.5~30, Interval：2.5	Range：0.25~3.5, Interval：0.25	Sand：47.4%, clay：5%

This study analyzed the surface temperature (st), soil volumetric water content (sm), soil bulk density (bd), roughness (RMS height (sig) and surface correlation length (cl) changes in the value of the latter to scattering coefficient.

The results showed that: the temperature from 0 ° to 30 °, backscattering coefficient of variation less than 0.05dB; soil volumetric water content from 0.02 to 0.50, the backscatter coefficient of variation less than 2.00dB; soil bulk density from 0.2 to 1.2, backscattering coefficient of variation less than 1.00dB; and can change the roughness backscattering coefficient of variation even up to tens of dB.

Based on the above analysis of the results, you can ignore the temperature change caused by changes in the backscattering coefficient, where the fixed temperature of 15 ° C, soil bulk density of 0.8 g/cm3, a second simulation. The simulation data according to different surface RMS height and correlation length of the configuration group, each change in soil moisture from 0.02 to 0.50, backscattering coefficient, including a 46 ° angle of incidence for the HH polarization and incidence angle of 54 ° the VV polarization, a total of 168 groups. Analysis have shown that, in most cases, the decibel value of two scattering data curves under different soil moisture conditions remain parallel. Due to space limitations, the following list only one of the two-group analysis chart (Fig.4).

Use the second simulation analysis of law get, by two different incidence angles and different polarization of the decibel value of the difference to eliminate the impact of soil moisture and roughness changes to inversion. According to our understanding of the Qinghai-Tibet Plateau region Maqu ground measurement data, that the soil moisture range of 0.2 to 0.5, roughness, slope (sig / cl) range from 0.04 to 1.0, select the corresponding simulated data set to analyze the data obtained RMS height and backscatter coefficient and the ratio of two-channel scatter meter decibel difference between the diagram(Fig.5).

According to the variation in Figure 6, the establishment of sig / cl and σ (dBHH)-σ (dBVV), exp (σ (dBHH)-σ (dBVV)) relation by regression analysis of the data to establish the following formula (3):

$$sig\big/_{cl} = 0.4231 \cdot \exp[\sigma(dB\text{VV}) - \sigma(dB\text{HH})] + 0.0034[\sigma(dB\text{VV}) - \sigma(dB\text{HH})] + 0.109 \quad (3)$$

Correlation coefficient is 0.9692, standard error is 0.0342, regression analysis of fitting errors shown in Figure 6.

Here we take   the 2008-07-02 global scattering plan the roughness of the estimate slope for example, Let's use of global HH, VV polarization scatter meter images to estimate the joint roughness slope, then Maqu region of the HH and VV scattering count data to estimate the roughness of the slope of the joint (Fig.7).

It can be seen from the figure, Maqu region in July 2 to estimate the roughness of the slope ranges from 0 to 0.5, consistent with the scope of information. Follow the same steps can be estimated from 2008-07-01 to 2008-08-29 Maqu region a total of 60 days roughness slope. It can be seen from the results, small changes over time, slope roughness, according to the average cycle of a certain treatment, such as ten days, in this time range roughness slope is very small fluctuations over time.
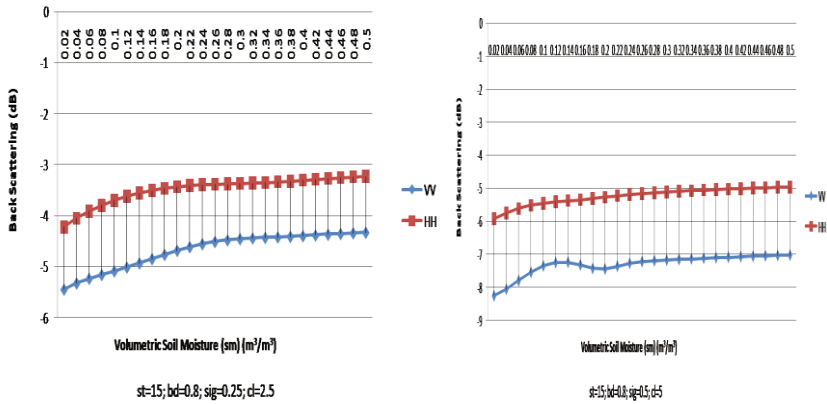
Fig. 4. Different roughness conditions the difference between the backscattering coefficient and soil moisture changes in the relationship between
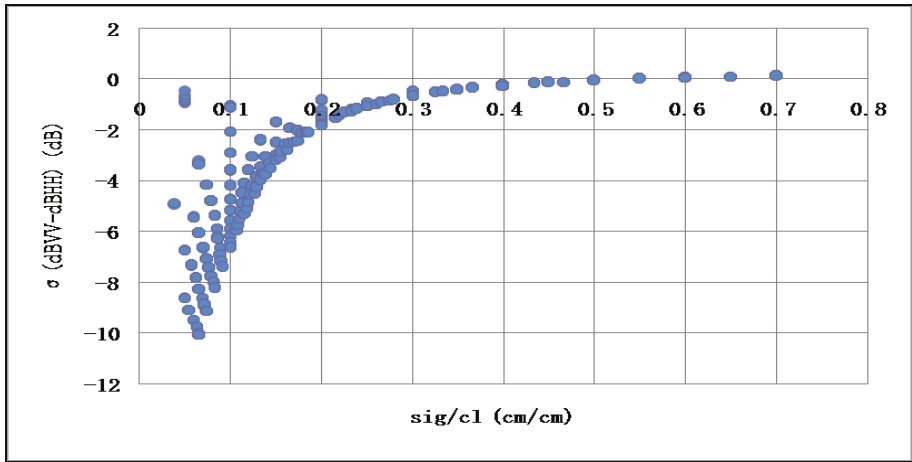
Fig. 5. RMS height and backscatter coefficient and the ratio of two-channel scatterometer decibel difference between
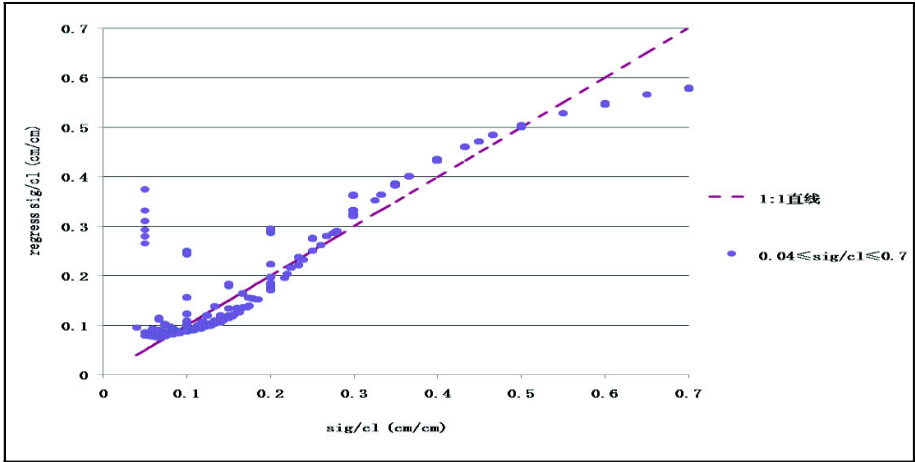
Fig. 6. Statistical regression analysis of fitting error
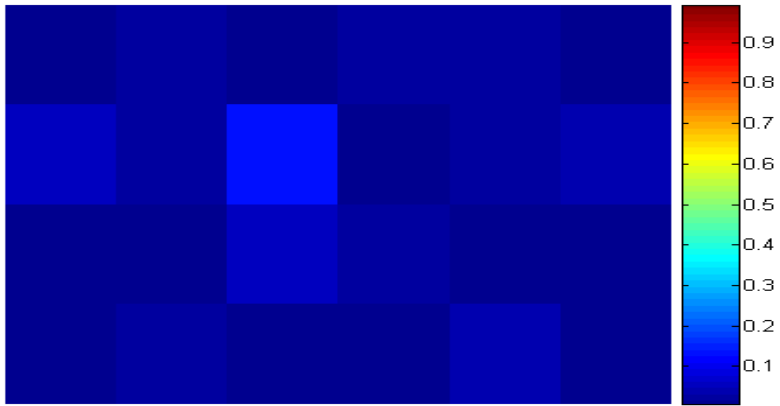
Fig. 7. With HH and VV Maqu region of scatter meter data to estimate the roughness of the slope image combined

4.3   Scatter Meter Soil Moisture Inversion

In the course of the study, due to the frequency when the incident wave near the C-band passive microwave atmospheric effects are negligible; when the frequency gradually increases, the atmospheric effects are also increased. And in order to eliminate measurement noise, where the use of 6 and 10 four-channel two-frequency inversion. Stay inversion of surface parameters to be soil moisture and surface temperature. Iterative inversion algorithm is the LM algorithm. Because of limited space, here only shows the average of all pixel inversion results (Fig.8), which results from the two diagrams: Scatter (left) and sequence diagram (right).
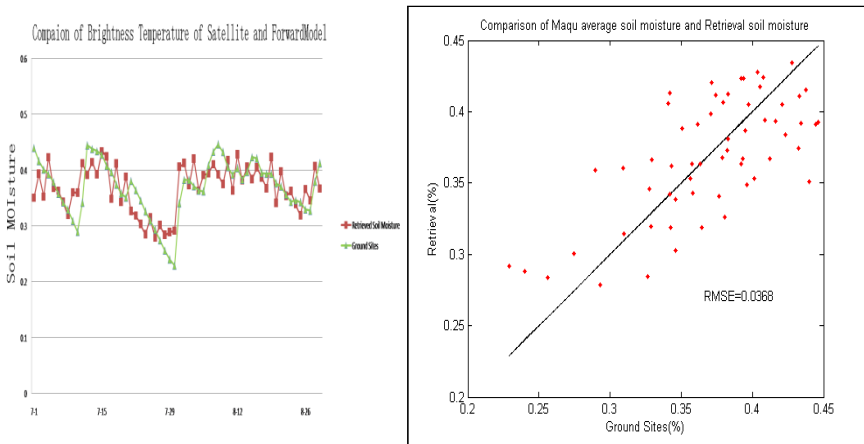
Fig. 8. The inversion of all pixels and averaging the results of ground truth comparison chart

5   Conclusions

Several conclusions can be drawn from our study.

(1) Inversion results has a high fitting accuracy with the ground measurements, indicating that the proposed method can take advantage of the complement of the active and passive microwave data to a certain extent. The scatter meter is more suitable for estimating the impact of surface roughness, and the low-frequency channel of the radiometer is the best choice for SM inversion, and the high-frequency of the radiometer is more relevant to the surface temperature and thus is fit for inversion of ST. Based on these above analysis, this research selects the best combination of observations to retrieve land surface parameters, to achieve a high retrieval accuracy.

(2) The other advantage of importing the Qp model is the active model and the passive model can be linked. This dealing is convenient to make full use of the active/passive's fittest channel to retrieve their most sensitivity land surface parameters.

(3) The proposed method in this research is for the bare soil surface or sparse vegetation cover surface, the surface of higher vegetation cover does not apply to.

References

[1]   Entekhabi, D., et al.: The Soil Moisture Active Passive (SMAP) Mission. Proceedings of the IEEE 98, 704–716 (2010)

[2]   http://hydros.gsfc.nasa.gov/

[3]   Njoku, E.G.: Observations of Soil Moisture Using a Passive and Active Low-Frequency Microwave Airborne Sensor During SGP99. IEEE Transactions on Geoscience and Remote Sensing 40, 2659–2673 (2002)

[4]   Bolten, J.D., Lakshmi, V., Njoku, E.G.: Soil moisture retrieval using the passive/active L-and S-band radar/radiometer. IEEE Transactions on Geo-science and Remote Sensing 41, 2792–2801 (2003)

[5]   Narayan, U., Lakshmi, V., Njoku, E.G.: Retrieval of soil moisture from passive and active L/S band sensor (PALS) observations during the Soil Moisture Experiment in 2002 (SMEX02). Remote Sensing of Environment, Papers 92(4), 483–496 (2004)

[6]   Shi, J., Njoku, E.G., Jackson, T., O'Neill, P., Chen, K.S.: Estimation of Soil Moisture with Dual-Frequency-PALS. In: IEEE International Geo-science and Remote Sensing Symposium, vol. 2, pp. 7–11 (2008)

[7]   Narayan, U., Lakshmi, V., Jackson, T.J.: High-resolution change estimation of soil moisture using L-band radiometer and radar observations made during the SMEX02 experiments. IEEE Transactions on Geo-science and Remote Sensing 44, 1545–1554 (2006)

[8]   Bindlish, R., Jackson, T., Cosh, M., Sun, R., Yueh, S., Dinardo, S.: Combined Passive and Active Soil Moisture Observations During Clasic. In: IEEE International Geo-science and Remote Sensing Symposium, vol. 6, pp. 644–648 (2008)

[9]   Piles, M., Entekhabi, D., Camps, A.: A Change Detection Algorithm for Retrieving High-Resolution Soil Moisture From SMAP Radar and Radiometer Observations. IEEE Transactions on Geo-science and Remote Sensing 47, 4125–4131 (2009)

[10] Zhong, R., Guo, H., Wang, W.: Combined passive and active microwave remote sensing of land surface parameters using AMSR-E and Seawinds. In: Proceedings of SPIE (2006)

[11] Scipal, K., Wagner, W., Kidd, R., Ringelmann, N.: Comparison of Ku- and C-band backscatter time series over land. In: IEEE International Geo-science and Remote Sensing Symposium, pp. 24–28 (2002)

[12] Scipal, K., Holmes, T., de Jeu, R., Naeimi, V., Wagner, W.: Error estimation of soil moisture derived from active and passive microwave satellite observations and model data. In: IEEE International Geo-science and Remote Sensing Symposium, vol. 2 (2008)

[13] Mladenova, I., et al.: An Assessment of Quikscat Ku-Band Scatterometer Data for Soil Moisture Sensitivity. IEEE Geo-science and Remote sensing letters 6, 640–643 (2009)

[14] Njoku, E.G., Li, L.: Retrieval of land surface parameters using passive microwave measurements at 6~18 GHz. IEEE Transactions on Geo-science and Remote Sensing, Papers 37(1), 79–93 (1999)

[15] Njoku, E.G., Jackson, T.J., Lakshmi, V., Chan, T.K., Nghiem, S.V.: Soil moisture retrieval from AMSR-E. IEEE Transactions on Geo-science and Remote Sensing 41, 215–222 (2003)

[16] http://nsidc.org/data/amsr_validation/soil_moisture

[17] Chen, K.S., Wu, T.D., Tsang, L., Li, Q., Shi, J.C., Fung, A.K.: The Emission of Rough Surfaces Calculated by the Integral Equation Method with a Comparison to a Three-Dimensional Moment Method Simulations. IEEE Transactions on Geo-science and Remote Sensing 41, 1–12 (2003)

[18] Shi, J.C., et al.: A Parameterized Multifrequency-Polarization Surface Emission Model. IEEE Transactions on Geo-science and Remote sensing 43, 2831–2841 (2005)

[19] Mao, K., Shi, J., Li, Z., Qin, Z., Li, M., Xu, B.: A physics-based statistical algorithm for retrieving land surface temperature from AMSR-E passive microwave data. Science in China (Series D), Papers 50(7), 1115–1120 (2007)

Grading Method of Crop Disease
Based on Image Processing

Youwen Tian, Lide Wang, and Qiuying Zhou

College of Information and Electric Engineering, Shenyang Agricultural University, China
youwen_tian10@163.com

Abstract. At present, in the crop disease harm degree is graded mainly by measure with the eye or paper cut primarily, which is greatly influenced by subjective factors, and results in obvious error. For improvement on identification precision of crop disease, this paper developed a new crop disease grading method based on computer image processing. Image preprocessing, segmentation and statistical calculation were applied effectively in this study. According to crop disease harm degree and classification standard, the crop disease harm degree was determined by computing the proportion of sickness spot area and the normal area on the leaf. The experiment results indicated that the identification accuracy was greatly improved, the grading time and costs is reduced by the manual evaluation, providing accurate data for the study of the crop's other aspects and has a broad prospect for application.

Keywords: Computer image processing, crop disease grading, grading system.

1   Introduction

Crop often subjected to various adverse environmental biotic and abiotic factors in growth and development process, leading to crop disease, resulting in a massive decline in crop production. Thus agricultural producers must obtain the crop disease hazard information quickly and accurately, adopt targeted prevention method, scientific and systematic management of disease in field, in the most economic way to restore the losses caused by diseases. But the current commonly used measurement method of crop disease harm degree is visual method or paper card classification method. Because this method is influenced by subjective factors grading error is relatively large, especially for small and many lesions of disease leaves time-consuming and great error were gotten to grade. Thus this no longer adapts to the development of modern agriculture needs. Therefore development of a rapid and accurate measurement methods of crop leaf disease grading is of great practical significance.

The computer image processing technique has many virtues such as high processing precision, good repeatability, strong quantitative and adaptability, capable of handling a large volume of data etc, so it is widely used in many fields, especially in agricultural engineering field, which shows unparalleled advantage.

According to the features of crop leaf spots when infected, the study adopted advanced digital image processing technology to make accurate judgment rapidly on the damage degree of crop disease, thus the agricultural producers can save loss caused by the crop disease at the most economical cost, can meet the demand of agricultural producers, reduce pesticide on agricultural products and environment pollution.

2   Materials and Method

2.1   Materials

The main purpose of this study is to make disease harm degree to grade on the crops leaves infected. In this study the most extensive crop diseases, namely in crops, fruit trees and vegetables are easily infected, and harm degree is the most serious symptoms on leaves, the most obvious manifestation of the disease is as study object. In the Cucumber Greenhouse, Field Corn and Fruit Grape research base of Shenyang Agricultural University, several disease leaves of downy mildew, powdery mildew and scab in cucumber, big spot, leaf blight, gray leaf spot in corn and downy mildew, powdery mildew, anthracnose in grape were collected on the field.

2.2   Testing Device

According to the requirements of research, the image acquisition system was constructed , as shown in Fig.1.The light box of the system is a homemade light box, which size is 80cm×80cm×80cm, inner 5 ring lamps act as the light source. The image sensor is used Panasonic WV-CP234 color CCD camera made in Japan Mataushita Electric Industrial Company or SONY DSC-F717 color CCD digital camera AverMedia, which made in SONY of Japan. PCI standard image acquisition card with the display module ensure the image real-time display in high quality, and with a communication interface. P4 compatible computer was applied in the system of computer image processing, with the frequency of 2GHz, memory of 1G, and hard disk of 80G.

2.3   Methods

Computer image processing technology was applied in this experiment. The crop disease severity degree was graded in fundamental steps as follows: image preprocessing, image segmentation and statistical calculation. First of all, acquire the images of the crop disease leaves. Subsequently, the color image vector median filtering method is applied to preprocess the crop leaf disease image. Then on the pretreated crop disease image selection of suitable color space based on lesion and normal color difference, segment the lesion and normal part using statistical pattern recognition method; then calculate the ratio percentage of the lesion area and the leaf area. According to crop disease harm degree and classification standard, the crop

disease harm degree were determined. Finally, the crop disease intelligent damage degree grading of hardware system and software system is developed for crop disease intelligent hazards degree grading system.
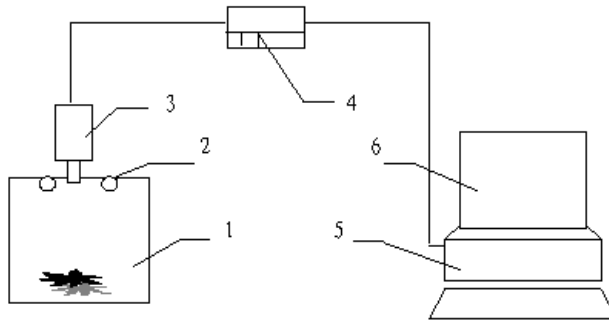
1-Light box;   2- light source;   3- video camera or a digital camera;   4- Acquisition card
5- host machine; 6- monitor

Fig. 1. Computer image acquisition system

3   Results and Analysis

3.1   Disease Image Preprocess

For color disease image, traditional image enhancement methods cannot properly remove noise and keep the edge. Therefore, this research make full use of color information of crop diseases image, the R, G, B value of each pixel in the image as a feature vector, applying a improved vector median filtering method enhanced crop disease image.

The improved vector median filtering method is to take the average of all vector Xi in the window to obtain mean vector $\overline{X}$, calculating the distance from the vector Xi to the mean vector $\overline{X}$, the vector whose distance is the nearest from the vector $\overline{X}$ as the element output value in window center. Therefore the algorithm is as follows:

(1) Calculating the mean vector in window, and calculating the distances from each vector to the mean vector Si′

$$\overline{r} = \sum_{i=1}^{s} r_i / s \qquad \overline{g} = \sum_{i=1}^{s} g_i / s \qquad \overline{b} = \sum_{i=1}^{s} b_i / s \qquad (1)$$

$$S_i^{'} = \left\| x_i - \overline{x} \right\| \qquad (2)$$

(2) After comparing all values of $s_i^{'}$, the smallest value of them is selected as $s_{min}^{'}$.

(3) The pixel X_{min} , which corresponding to S_{min} ' is the vector median in the window , replacing the pixel vectors in window center.

If the minimum value of S_{min} in not unique according to the above algorithm, how to determine the mean vector is not given in this method. In this research, the Xi is treated as median vector, which is corresponding to S_{min}.

With the improvement of color image vector median filtering method as above stated, the cucumbers, grapes, corn disease leaf images were enhanced. Experimental results show that the improved algorithm has the advantages of better edge and details and does not increase new colors. In addition it has a small amount of calculation, its running time is short, it is simple and easy to realize. So the improved vector median filtering method was applied in this study.

3.2   Image Segmentation

Crop disease image has complex components, the leaf spots arranged on no rules, and a different color depth. Since diseases are different, the spot color is not the same. It can be seen from the image of crop leaf disease, the leaf infected is composed of two parts: lesion and normal portion, essentially it is a two class of problems. Therefore this research applied the statistical pattern recognition classification approach to segmentation image.

Suppose discriminant function is linear, the general expression is:

$$g(x) = w^T x + w_0 \tag{3}$$

Where x is a three-dimensional feature vector, w is called the weight vector. They are respectively expressed as follow:

$$x = [R, G, B]^T, \quad w = [w_1, w_2, w_3]^T \tag{4}$$

Where w_0 is a constant, known as the threshold.

The classifier was designed according to the following steps:

 (1) Take representative disease spot part and normal part of training data set.

 (2) In the three dimensional RGB color space, to each type of sample of statistical analysis, calculating each type of data sets m1, m2 ,s1, s2, and concluded that w *.

 (3) According to $y = w^{*T} x$, draw a one-dimensional data set, and carries on the statistical analysis to determine the classification threshold y.

 (4) Scanning image, according to $y = w^{*T} x$, determine the image of each pixel point of projection y.

 (5) According to the decision rules, making decision classification.

$$y \geq y_0 \rightarrow \begin{cases} x \in L_1 \\ x \in L_2 \end{cases}$$

Applying above classifier, the disease leaf images of cucumber, corn and grape were isolated from background respectively. The test results as shown in Fig.2. As can be seen from the graph, good segmentation results were obtained, and spot part can be divided from the normal part basically.

(a) Original image   (b) Filtering image   (c) Segmentation of leaf   (d) Segmentation of spot
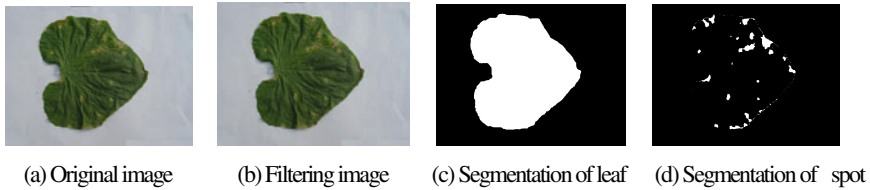
Fig. 2. Segmentation results of disease leaf image

After crop disease image segmentation, there are often isolated points, burr and small cavities in the image. In order to reduce the noise effect on the subsequent identification, open and close operation in mathematical morphology was applied in this study to eliminate noise, so as to achieve the ideal image.

3.3   Grading Method of Crop Disease Based on Image Processing

The crop disease harm degree classification level are expressed as the percentage of disease spot area to total leaf. Because the lesion area is proportional to the number of the pixel in the lesion region, the normal leaf area is proportional to the number of the pixel in the normal region, so the leaf damage degree classification grades can be expressed as the ratio k of the number of the pixel in the lesion region to that in the normal region.

$$k = \frac{A_1}{A} = \frac{N_1}{N} \tag{5}$$

Where A represents the disease leaf's total area, A_1 represents lesion area, N represents the pixel number in the leaf, N_1 represents the pixel number in lesion area.

Based on crop disease original image preprocessing and image segmentation, this study completed damage degree automatic grading by effective extraction of the crop disease leaf area and the disease spot area. Therefore, according to the ratio k of the number of the pixel in the lesion region to that in the normal region in pros and cons of diseased leaf segmented, disease severity can be identified and the leaf infected level can be obtained. According to documents standard and modification protocol of plant protection experts, crop disease is divided into 6 levels, the specific grading standards as shown in Table 1.

This study respectively applied image processing technology, the leaf area meter and paper cut method to determine the cucumber disease damage degree and classification.

The crop disease harm degree values determined by leaf area meter method were treated as the standard values of disease harm degree. The experimental results showed that the harm degree of crop diseases, either by the image processing system measurement or by paper cut measurement are compared with the standard value, there is an error of certain degree. The accuracy of the system by image processing measurement is higher than that of paper cut measurement. So crop disease damage degree automatic grading system is more effective based on the image processing.

Table 1. Crop disease classification standard

level	0	1	2	3	4	5
cucumbers	0%	0.1%~5%	5.1%~10%	10.1%~25%	25.1%~50%	50.1%以上
corn	0%	0.1%~5%	5.1~10%	10.1%~30%	30.1%~70%	70.1%以上
grape	0%	0.1%~5%	5.1~30%	30.1%~50%	50.1%	

3.4   System Realization

In this study the crop disease automatic recognition system is developed in the WINDOWS environment with MATLAB 7.01. Software system mainly includes the system interface, image preprocessing, image segmentation and harm degree of crop disease classification module.

Finally, according to the harm degree of plant disease and classification results, disease grade was shown, especially for cucumbers, grapes and corn disease harm degree of determination, specific for disease automatic grading of cucumber downy mildew, powdery mildew and scab, grape downy mildew, powdery mildew and anthracnose and corn big spot, leaf blight and grey spot disease. The results interface of the crop disease harm degree is shown in Fig.3.

Fig. 3. Results of disease harm degree

4   Conclusion

The grading accuracy was greatly improved by using image processing technology on crop leaf disease level classification, reducing the time and costs by the manual evaluation, providing accurate data for the study of the crop's other aspects. The hardware requirement of computer by this method is very low. The crop disease grading method can be implemented effectively; It can made up of defects and deficiencies by artificial classification, improving the classification grading and efficiency. This technology has broad prospects for application.

References

1. Mao, H., Xu, G.: Identification the deficit of tomato nutrient elements based on computer vision. Transactions of the Chinese Society for Agriculture Machinery 34, 73–75 (2003)
2. Tian, Y., Li, T., Li, C., et al.: Method for recognition of grape disease based on support vector machine. Transactions of the CSAE 23, 175–179 (2007)
3. Camargo, A., Smith, J.S.: Image pattern classification for the identification of disease causing agents in plants. Computers and Electronics in Agriculture 66, 121–125 (2009)
4. Cen, Z., Li, B., Shi, Y., et al.: Discrimination of cucumber anthracnose and cucumber brown speck based on color image statistical characteristics. Acta Horticulturae Sinica 34, 1425–1430 (2007)
5. Zhang, Z., Wang, Y., Xue, G.: Digital Image Processing and Machine Vision—Visual C++ and Matlab. People's posts and Telecommunications Press, Beijing (2010)

Evolution Characteristics for Water Eco-Environment of Baiyangdian Lake with 3S Technologies in the Past 60 Years

Yunkai Li[1,2,*,**], Lingyan Wang[1,*], Hua Zheng[2,*], Hai Jin[3,*], Tingwu Xu[4,*],
Peiling Yang[1], Xiaokai Tijiang[1], Zengcai Yan[5], Zhiheng Ji[6], Jianli Lu[7],
Zhanfeng Wang[6], and Zhiyun Ouyang[2,**]

[1] College of Water Conservancy and Civil Engineering, China Agricultural University,
Beijing 100083, P.R. China
[2] State Key Laboratory of Urban and Regional Ecology, Research Centre for Eco-Environmental
Sciences, Chinese Academy of Sciences, Beijing 100085, P.R. China
[3] College of Sciences, Beijing Institute of Technology, Beijing 100081, P.R. China
[4] International College at Beijing, China Agricultural University, Beijing 100083, P.R. China
[5] Hebei Daqing River Administrative Office, Hebei 071051, P.R. China
[6] Baoding Bureau of Hydrology and Water Resources Survey, Hebei 071000, P.R. China
[7] Baoding Meteorological Office, Hebei 071051, P.R. China

Abstract. Keeping water eco-environment health and virtuous circle of lake-wetland, and letting lakes recuperate and reproduce has become a common view all over the world. Although a great deal of research results were obtained, such as lake-wetland ecosystem management and eco-hydrological mechanism of eco-environment evolution, there are few researches about comprehensive analysis on evolution trend and driven mechanism of lake-wetland water eco-environment. Baiyangdian, the biggest grass-type lake-wetland in North China, was chosen as the research object in this paper. Based on the long-term and field observations of water environmental quality and analysis for remote sensing images of water ecological landscape pattern, the evolution trend and driving mechanism for water eco-environment of Baiyangdian in the past 60 years have been studied systematically. The results showed that the average annual water level fluctuated greatly with a decreasing tendency as a whole. During 1974 and 2007, there was a general pattern of land use types, which the emergent plants, water body and cultivated land were the main, and the residential area, forest land, grass land and unused land were distributed as flower arrangement state. During the periods, the water body and emergent plants areas were decreased respectively by 28.7% and 20.8%, while the cultivated land areas were increased by 69.7%. The main pollutants (NH4+-N, CODMn, BOD5, TP, and TN) concentrations in Baiyangdian Lake were changed greatly with spatiotemporal variation and the pollutant concentrations were increased obviously. In the three polluted regions (north zone, south zone, and

---

[*] Yunkai Li, Linyan Wang, Hua Zheng, Hai Jin and Tingwu Xu are equally contributed to this paper.

[**] Corresponding author.

mixed zone), the north zone was polluted most seriously. The main pollutants concentrations were fluctuated obviously over a year, and the pollution was more serious in summer of normal and dry years, while the pollution was serious at the beginning of summer flood in wet years, and then became less. At last, the evolution trend and driven mechanism of water eco-environment of Baiyangdian was analyzed from three aspects, including typical climate indexes (such as precipitation, evaporation, air temperature, and solar radiation), ecological/environmental hydrology regime in upstream watershed of Baiyangdian, and human activities. It would provide some references to the comprehensive treatment of water eco-environment in Baiyangdian Lakes.

0   Introductions

Lakes, natural or artificial, are important for human development and for the preservation of sound ecosystems and biodiversity on our planet. They contain 90 percent of the liquid freshwater on the earth's surface and are critical elements of the water cycle; they sustain aquatic biodiversity and provide livelihoods and social, economic, and aesthetic benefits that are essential for the quality of life in Lake Basin communities. However, under the circumstances of global climatic changes, industrialization and urbanization, the lake-wetland presents a good many ecological and environmental problems, such as reduced water surface, deteriorated water quality and destroyed water ecosystem, etc.[1] The management and protection of lake water eco-environment are facing severe challenges. To let lakes recuperate and multiply by rationally developing and utilizing lake-wetland resource, strengthening management and protecting of lake-wetland, maintaining lake-wetland water eco-environmental health, has become a common view all over the world[2] .

During the past century, a lot of research results about the forming process of lake-wetland ecosystem and its hydrological mechanism (including ecological/ environmental water requirement and water balance, dynamic changes of water level, water transfer, etc,) were obtained by all of the countries in the world [2], ant it has been become the main basis of managing lake-wetland resources. But on the whole, there are few researches about comprehensive analysis on evolution trend and driven mechanism of lake-wetland water eco-environment, which was centralized in the investigating research less than ten years, while the analytical results of evolution trend and driven mechanism of lake-wetland water eco-environment based on longtime locating monitoring are much less. Meanwhile, the complex relationship between lake-wetland ecological process and hydrological process, which not only relates to the development of wetland ecosystem itself, but also relates to ecological/environmental hydrological process of lakes upstream watershed, makes the evolving rule of each lake-wetland water eco-environment possess a certain common features as well as unique features. Therefore, it is necessary to comprehensively analyze evolution trend and driven mechanism for the water eco-environment of every typical lake-wetland.

In this paper, Baiyangdian Lake was chose as the research object, which was the biggest grass-type lake-wetland in North China, to study systematically the evolution trend of lakes water eco-environment with long-term and field observations of water environmental quality and analysis for remote sensing images of water ecological landscape pattern. And the driven mechanism for water eco-environment from climatic changes and human activities was analyzed. It would be hoping for providing the references about the protection and ecological restoration of Baiyangdian Lake and even lake-wetland all over the world.

1   General Situation of Baiyangdian Lake-Basin

Baiyangdian Lake, which is located between 38°43′-39°02′N and 115°38′-116°07′E and the heartland of three cities Beijing, Tianjin, and Shijiazhuang, is the most representative grass-type lake-wetland in North China[3]. In addition, Baiyangdian Lake is likely to play an important role in the South-to-North Water Division Project under construction[4] .Baiyangdian Lake is located at the middle reaches of Daqing River, Haihe River Basin. And it controls a total area of 31, 199 km^2, which accounts for 69.1% of the total area of the watershed. Baiyangdian Lake can be divided into four basin units according to river systems, including South Branch mountain area, North Branch mountain area, South Branch plain, and North Branch plain. Fig.1 showed that the distribution of Baiyangdian Lake and drainage basin water systems. Since ancient times, Baiyangdian Lake undertakes water from the south branch of Daqing-Zhaowang River (including Zhulong River, Xiaoyi River, Tang River, Fu River, Cao River, Baigouyin River, Pu River, and Ping River). The discharge is controlled by Zaolinzhuang regulation sluice and Shifangyuan Overflow Weir. It pours into sea via Zhaowangxin Canal, Duliujian River, and Beida Harbor[5].

Baiyangdian Lake watershed is located the monsoon climate of warm temperate area, which is hot and rainy in summer, cold and dry in winter. The lowest temperature is -30.6℃ and the highest is 43.5℃, and average annual temperature is 7.3-12.7℃, the average yearly accumulated temperature is 29.93-44.09℃. The average annual precipitation in the whole watershed is 546.2mm, and the average annual evaporation is 1000-1200mm. Baiyangdian Lake borders the dykes, which is 39.5km from east to west, 28.5km from north to south, and the total area is about 365km^2. The main landscape of Baiyangdian Lake is water, the bottom of which is high in the west and low in the east, the altitude of it is 5.5-6.5m. The whole lake was divided to unequal and interrelated 143 sub-lakes with 3700 staggered channels. The tunable water storage capacity is 2.9×10^8 m^3, water area accounts for about 50%[6]. Because of its special geographical position, Baiyangdian lake-wetland provides vital ecosystem service functions such as water conservation, water purification, regional climate regulation, species diversity maintenance, and so on. It plays an irreplaceable role in maintaining the ecological safety in North China, and is functioned as the "kidneys" of North China[5](Zhao,Cui,Yang,2005).

With the affection of both climatic changes and human activities in the recent years, Baiyangdian Lake has been threatened by drought and pollution. The vegetation in upstream watershed becomes scarce and the function of water conservation is weak. Eco-environment problems appear, including frequent drying up, water pollution, the extinction of rare biological populations, etc.[7] . With the rapid degeneration of ecosystems, Baiyangdian Lake has changed from an exuberant lake to a lake facing shrink or even disappearance. The existence or disappearance of Baiyangdian Lake not only affects people's life and production, but also affects economic development and environmental balance of the North China Plain. The environment problems of Baiyangdian Lake have led to concerns of all circles in society, and the researchers have carried on a large amount of researches on eco-environment problems of Baiyangdian Lake[8], but few researches about comprehensive analysis on evolution trend and driving mechanism for water eco-environment of Baiyangdian Lake are reported.

2   Materials and Methods

2.1   Monitoring on Water Quality and Water Level in Baiyangdian Lake

The monitoring data about the water level of Baiyangdian Lake in the recent 60 years (1950-2008) were collected. The locations of the monitored spots were shown in Fig.1.
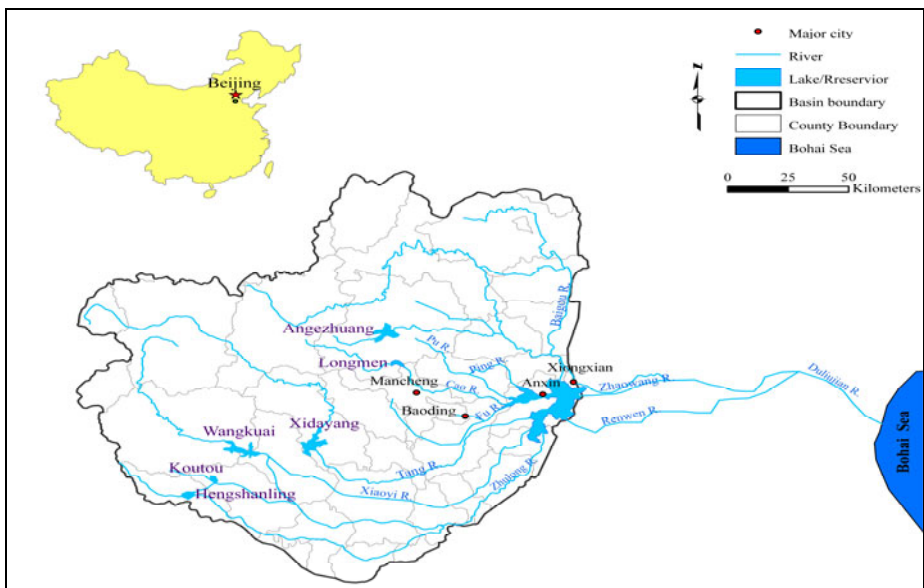
Fig. 1. Water System of Baiyangdian Lake and Its Upstream Watershed
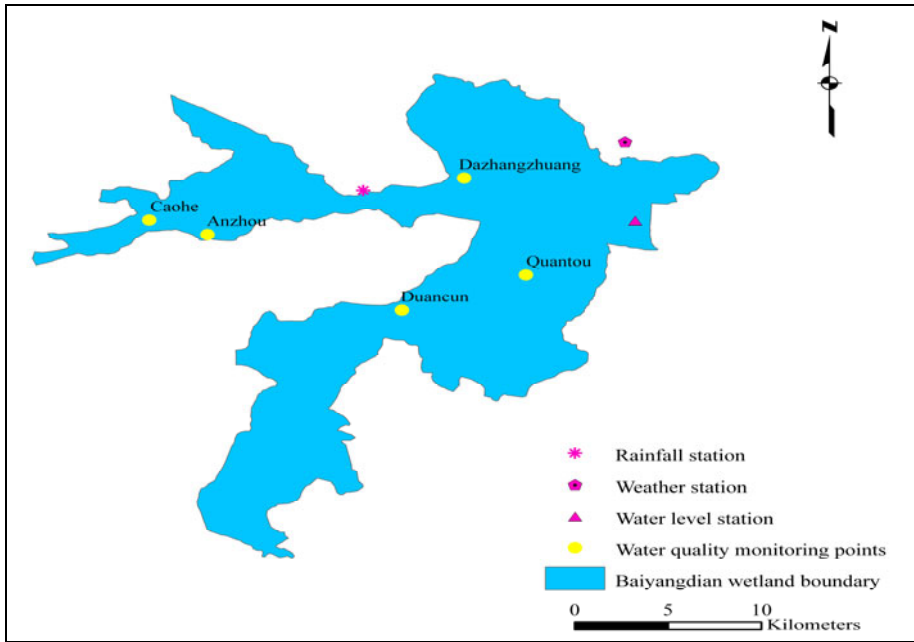
Fig. 2. Different Kinds of Monitoring Points

The pollutants in Baiyangdian Lake are from many sources, widely distributed, and constituted complexly. The main pollution indicators of water quality in Baiyangdian Lake are NH_4^+-N, COD_{Mn} (Chemical Oxygen Demand, Permanganate Index), BOD_5 (Biochemical Oxygen Demand after 5 days), TP (Total Phosphorus) and TN (Total Nitrogen). According to water sources and water flow characteristics in Baiyangdian Lake, it can be divided into three pollution zones, which is south pollution zone that undertakes water from south branch water system (mainly including Zhulong River, Xiaoyi River, and Tang River), north pollution zone that undertakes water from north branch water system (such as Fu River, Cao River, Baigouyin River, Bao River, Ping River, and so on), and mixed zone of the two pollution zone. The observation data (NH_4^+-N, COD_{Mn}, BOD_5, TP) of long time observation spots in the three zones from 1988 to 2008 were collected, including 1988 (restoration of water after the first drying up), 1991, 1994, 1996, 1999, and 2002-2008, the observation frequency was different from once per month to once per season; and each month's observation material of TN from 2002 to 2008. And the water quality monitoring data of the estuaries of Cao River and Fu River were collected, which was the most serious polluted location in North branch water system of Baiyangdian Lake watershed. The locations of water quality monitoring spots were shown in Fig.2.

2.2   Monitoring on Landscape Pattern of Water Ecosystem with Remote Sensing and Its Analysis Methods

2.2.1   Data Sources of Remote Sensing Images and Its Disposal Methods

The Landsat MSS/TM remote sensing images of studied region were chosen as the data source, which was Sep. 3, 1974, Aug. 28, 1987 (Dried-up Year), Jul. 13, 1996 (Wet Year), and Aug. 24, 2007. The software ENVI 4.5 was used to modify atmosphere of the images, and the map of roads in Baiyangdian Lake basin with a scale of 1:100, 000 were used to carry on Geometric Calibration and Registration. Actuality Map of Land Use in 1983, Administrative Map, and DEM images were used to help interpret Baiyangdian Lake wetland landscape pattern, as well as to enhance the accuracy of interpreting land type. Referring to *Technical Specification for Investigating the National Wetland Resources*[9], and combining characteristics of Baiyangdian Lake landscape pattern and wetland evolving, Baiyangdian Lake landscape patterns can be divided into six kinds, which are: (1) Emergent plant, mainly including the distributing region of emergent plants (Reeds and shrub) and hygrophyte plants. (2) Water area, mainly including the submergent plants region, clean water region, and trenches. (3) Forest/grass land, mainly including the forest land, grass land and forest-grass mixed land. (4)Cultivated land, mainly including the regions of planting grain and cash crops. (5)Residential area. (6)Unused land.

2.2.2   Image Disposal Methods Based on Support Vector Machine

Support Vector Machine (SVM) theory and method was used to classify the images. SVM is a new study method, which was developed based on Statistical Learning Theory put forward by Vapnik at AT&T Bell laboratory[10] . SVM, which can solve practical problems including small sample, nonlinearity, high dimension, local minimum point, has become a hot spot in machine learning area, and is successfully used in aspects such as classification and regression. According to the differences of the problems solved, SVM is called Support Vector Classification (SVC) and Support Vector Regression (SVR), respectively. The mechanism of SVC is to find an Optimal Hyperplane (OHP) which meets the demand of classification, in order to guarantee the accuracy of classification while not only classify the two kinds of data spots correctly as many as possible, but also make the blank area at the two sides of the Optimal Hyperplane the biggest - to make the classified two kinds of data spots farthest from classified plane[11]. SVC method was used to analyze the landscape classification of remote sensing images in this paper.

Given a training dataset $X = \left\{ (x_i, y_i) : x_i \in R^n, y_i \in \{+1, -1\} \right\}_{i=1}^{m}$, where each x_i is labeled by y_i. Linear SVM finds a boundary that separates two different classes of feature vectors with a maximum margin. It leads to the following convex quadratic programming problem:

$$\min_{w,b} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \xi_i \qquad (1)$$

$$\text{s.t.} \quad y_i(w^T x_i + b) \geq 1 - \xi_i \tag{2}$$

$$\xi_i \geq 0, i = 1, 2, \ldots, m \tag{3}$$

A nonlinear SVM projects feature vectors into a high dimensional feature space by using a kernel function such as a Gaussian kernel function $K(x_i, x_j)$. The linear SVM procedure is then applied to the feature vectors in this feature space. The corresponding dual quadratic programming to (1)–(3) for nonlinear SVM is as follows:

$$\min_{\alpha} \quad \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{m} \alpha_i \tag{4}$$

$$\text{s.t.} \quad \sum_{i=1}^{m} y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq C, i = 1, 2, \ldots, m \tag{5}$$

where C is a constant controlling the trade-off between maximizing the margin and minimizing the errors. After the determination of the solution α^* of (4)-(5) and b^*, we can get the decision function

$$f(x) = \text{sgn}\left\{ \sum_{i=1}^{m} \alpha_i^* y_i K x_i, x) + b^* \right\} \tag{6}$$

where the function sgn is a sign function. If $x>0$, sgn(x)=1; otherwise sgn(x)=-1. In the present work, the Gaussian kernel function $K(x_i, x_j) = \exp\left\{ -\|x_i - x_j\|^2 / 2\sigma^2 \right\}$ is used. Because of the need to classify the lake-wetland to six landscape patterns, we extended this bi-class classification problem studied by SVM to the field of multi-class classification, and used bi-class classification machine which constructed $n(n-1)/2$ to combine them together and realize n-class classification. In order to enhance classified accuracy of the research region, the optimal value of punishing factor C and slack variable ξ in SVM method were obtained by using LIBSVM process. The land use types of 516 points were investigated with GPS (The number of each landscape samples were almost the same), and it was divided two parts, which one half was used to form the training dataset, another one was used to calibrated the classification results combining the land use data in 1983. The results showed that the image classification was good, which the precision of image classification and Kappa index were over 0.92 and 0.88, respectively.

2.2.3 Analysis on Wetland Landscape Pattern and Its Dynamic Changes

According to the characteristics of Baiyangdian Lake and research need ,a set of landscape pattern indexes were chose to reflect the changes of landscape structure and landscape pattern. The calculating formulae were as follows:

(1) Landscape area index. Patch type area (*PA*): the total area of a certain landscape pattern.

Patch and landscape area ratio (*PR*): the percentage of a certain patch type accounts for the whole landscape area.

Patch number (*PN*): the total number of a certain patch type appears in the landscape.

(2) Landscape fragmentation index. Patch density (*PD*): the ratio of patch number and the total patch type area.

The average patch fractal dimension (*FRAC_MN*): it reveals the relationship between shape and area of landscape constituted by patches, measures the complexity of the landscape shape and eco-process of the effect of landscape shape on the internal patches, and reflects the effect of human activity on landscape. The calculating formula is:

$$FRAC\_MN = \frac{\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n}\left[\dfrac{2\ln(0.25P_{ij})}{\ln(a_{ij})}\right]}{N} \tag{7}$$

where a_{ij} is patch's area, P_{ij} is the perimeter of landscape type, N is the total number of patches.

(3) Landscape diversity index Shannon diversity index (SHDI): It reflects landscape heterogeneity and is sensitive to the Disequilibrium distribution condition of each patch type, which emphasizes on the contribution of rare patch type to information. The formula is:

$$SHDI = -\sum\limits_{i=1}^{m}\left[P_{ij}\ln(P_{ij})\right] \tag{8}$$

where, P_i is the percentage of a certain patch type accounts for the whole landscape area.

(4) Landscape evenness index   Shannon evenness index (SHEI): It reflects non-uniformity degree of each patch area distribution in landscape. It is represented by the ratio of diversity index and the maximum value. The formula is:

$$SHEI = SHDI/lnN \tag{9}$$

*(5) Dynamic change index of landscape pattern   Landscape type dynamic change index is the changing degree of a certain landscape type area per unit time. It can quantitatively characterize the changing speed of landscapes, which has a positive effect on comparing region differences of landscape type changing and predicting the changing tendency of landscape type in the future. The calculating formula is:

$$K = (U_b - U_a)/U_a \times 1/T \times 100\% \tag{10}$$

where K is dynamic degree of a certain landscape type in the research period of time, U_a and U_b are the number of a certain landscape type at the beginning and in the end, respectively; T is the length of the research time.

Transfer matrix of landscape is the area matrix and probability matrix of the transfer between different landscape types in a sequence unit, which not only reflects landscape type structures of different periods, but also reflects the condition of transfer between different land types. Transfer probability matrix can be calculated as follows:

$$P_{ij}=A_{ij}/A \tag{11}$$

where $0 \le P_{ij} \le 1$ and $\sum_{i=1}^{N} P_{ij} = 1$ $(i,j = 1, 2, \cdots, n)$, P_{ij} is the probability of transferring from type i to type j (from the beginning of the research to the end), A_{ij} is the area transferring from type i to type j (from the beginning of the research to the end), A is the area of type i at the beginning of the research; n is the number of landscape types.

2.3   Meteorological Data

The yearly and seasonal sequence observation data (including air temperature, solar radiation, evaporation) from 1956 to 2008 in Anxin county of Baoding City were collected, which is the location of Baiyangdian Lake. Yearly precipitation materials in Baiyangdian lake-wetland from 1956 to 2008 were also collected. The distribution of precipitation and weather stations were shown in Fig.2.

3   Evolution Trend for Water Eco-Environment in the Past 60 Years

3.1   Dynamic Water Level in Lakes

The ecological service functions of lake-wetland were closely correlated with the water level and water surface area, and the biodiversity and even the whole ecological function had great difference under different water lever conditions[12]. In the recent decades, the water level of Baiyangdian wetland has fluctuated greatly, and Fig.3 showed the annual average water level of Baiyangdian Lake observed by Shifangyuan in the recent 59 years (1950-2008). According to the result of Mann-Kendall significance testing, U=-4.96<0 , $|U|$=4.96>$U_{\alpha/2} = 2.32$ (α =0.01), which meant that water level showed obvious decreasing tendency. For the recent 60 years, the water level in Baiyangdian fluctuated between drying up level (Lake drying up was defined as the observed water level by Shifangyuan Monitoring Site was lower than 6.5m) and 10.03m. From 1950s to the middle 1960s, the water level was high, and the lowest water level was 8.12m. Between 1965 and 1983, water level was between 6.37m and 8.69m, and Baiyangdian Lake was dried up from 1984 to 1987(the gap in Fig.3). From 1988 to 2000, water level decreased significantly except the year 1991 and 1996.

Baiyangdian Lake was dried up again in 2001 and 2002. The average water level from 2003 to 2008 was 7.04m, and 1.55m lower than the average water level from 1950 to 1964 which was 8.99m.
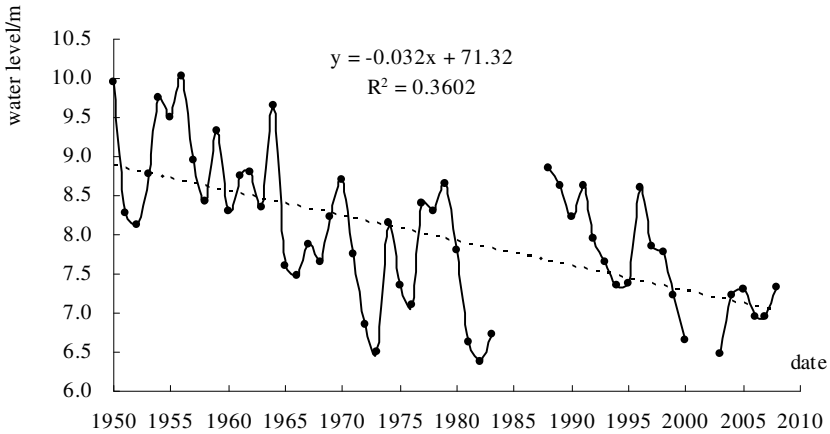
Fig. 3. Annual Average Water Level in Baiyangdian Lake between 1950 and 2008

3.2   Space-Time Evolution Trend of Water Ecological Landscape Pattern

Four remote sensing images were interpreted with ERDAS IMAGINE 8.7 and ENVI 4.5 software, and Fig.4 showed the results. The wetland landscape pattern evaluating indexes at different time were calculated by using software ArcGIS 9.2 and Fragstat 3.3, and Tab.1 and Tab.2 showed the results.

From 1974 to 2007, there were large proportion of emergent plants, water body and cultivated land in Baiyangdian Lake, small proportion of residential area, forest land, grass land and unused land. The proportions of emergent plants in different years were 35.4%, 43.6%, 39.0%, and 28.1% respectively. The proportions of water body were 31.1%, 7.3%, 31.4 %, and 22.2%. The proportions of cultivated land were 24.7%, 37.1%, 23.0%, and 41.8% respectively. Generally speaking, the four landscape types, including emergent plants, water body, forest/grass land, and used land, showed a decreasing tendency in different degrees. Cultivated land and residential land increased significantly.

Emergent plants mainly changed to cultivated land and water body, change rate were 31.07% and 13.35%, respectively. Water body mainly changed to emergent plants and cultivated land, change rate were 27.29% and 17.94%, respectively. And changing phenomenon between water body and emergent plants existed because of the weather condition. Due to human activity, economical and social modernization, and the idea of the conversion of cropland to forest and grassland, cultivated lands mainly changed to forest, grassland, and residential land. The most violent change of landscape pattern happened from 1996 to 2007, the areas of emergent plant and water body decreased by 3984.8 hm^2 and 3371.4hm^2 with the speed of 2.15%/a and 2.26%/a, cultivated land

increased by 6865.6hm^2 with the speed of 6.28%/a; the most violent change of residential land happened from 1987 to 1996, which increased by 437.3hm^2 with the speed of 4.1%/a in a decade.
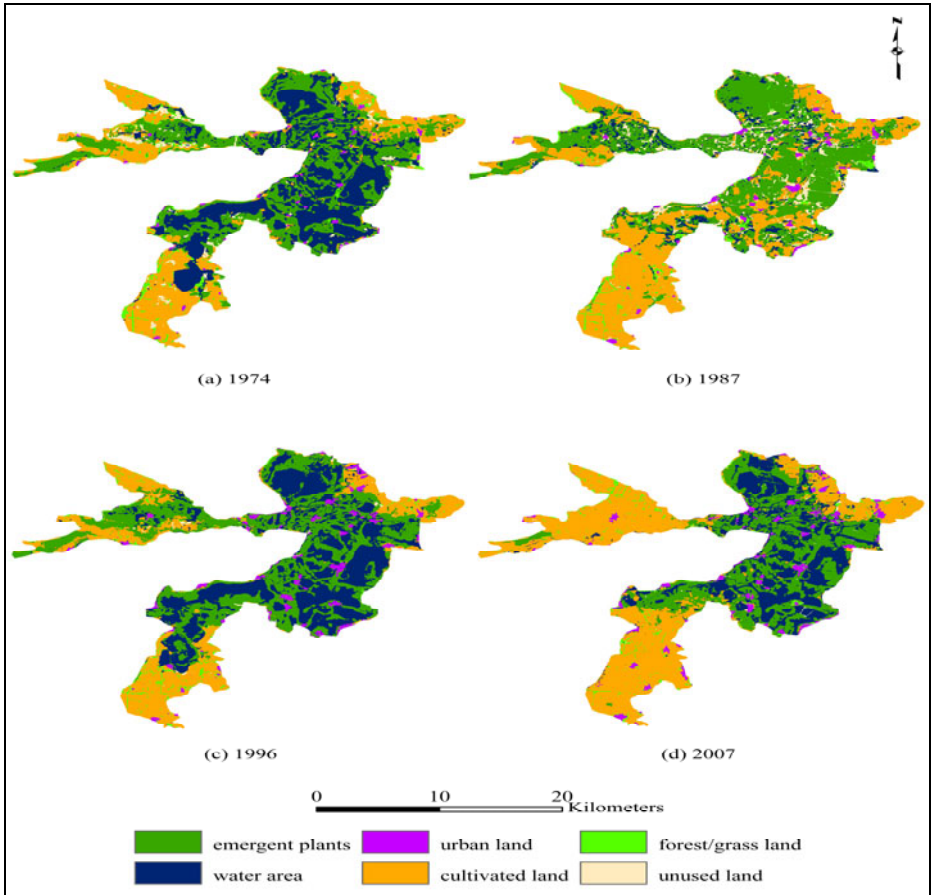
Fig. 4. Classification of Water Ecological Landscape Pattern in Baiyangdian Lake with Remote Sensing Monitoring

From landscape pattern evolving indexes listed in Tab.1, PN and PD had decreasing tendencies as a whole, which meant that heterogeneity of different landscape types decreased[13]. FRAC_MN showed a decreasing tendency, which indicated that landscape fragmentation degree increased and was more complex, suffered a tremendous influence from human factor, but had no significant difference[14]. In 1974, SHDI and SHEI in the research region were the biggest, which showed that landscape diversity degree was high in this period, heterogeneity was high, and dominance degree of patches in landscape was small; in 1987, SHDI and SHEI were the lowest, which meant that heterogeneity of different landscape types decreased,

dominance degree increased, being in conformity with the fact of the sharp decrease of water area and the significant increase of emergent plants and cultivated land; from 1996 to 2007, SHDI and SHEI showed an increasing tendency, which indicated that landscape developed to diversity and uniformity, and embodied the tremendous effect of human activity in this period.

Table 1. Water Ecosystem Landscape Pattern Evaluating Indexes of Baiyangdian Lake in 1974-2007

Items	Patch level PA/hm^2 (PR/%)						Landscape level				
	Emergent plants	Water area	Cultivated land	Forest/grass land	Urban land	Unused land	PN /ind	PD /ind/hm^2	FRAC_MN	SHDI	SHEI
1974	12928.2 (35.4)	11355.4 (31.1)	8999.6 (24.7)	1643.2 (4.5)	582.3 (1.6)	998.8 (2.7)	20109	0.55	1.10	1.37	0.77
1987	15923.4 (43.6)	2662.7 (7.3)	13531.8 (37.1)	1527.8 (4.2)	820.4 (2.3)	2041.4 (5.6)	23560	0.65	1.08	1.29	0.72
1996	14228.6 (39.0)	11462.8 (31.4)	8406.9 (23.0)	1005.1 (2.8)	1257.7 (3.5)	146.2 (0.4)	19894	0.54	1.09	1.31	0.73
2007	10243.8 (28.1)	8091.4 (22.2)	15272.5 (41.8)	1283.2 (3.5)	1490.1 (4.1)	126.5 (0.4)	19788	0.54	1.08	1.32	0.74

Table 2. Transfer Probability Matrix of Water Ecosystem Landscape Pattern in Baiyangdian Lake (1974-2007) (%)

Landscape types	Dynamic dgree K	Emergent plants	Water area	Cultivated land	Forest/grass land	Urban land	Unused land
Emergent plants	-1.60	50.4	27.29	4.75	6.55	2.61	9.9
Water area	-2.21	13.35	51.36	3.49	7.14	5.69	4.42
Cultivated land	5.36	31.07	17.94	81.16	69.18	12.74	76.21
Forest/grass land	-1.69	2.23	1.47	5.46	8.84	9.16	6.22
Urban land	11.99	2.66	1.57	4.89	7.87	69.44	2.93
Unused land	-6.72	0.28	0.37	0.25	0.42	0.37	0.32

3.3 Spatiotemporal Variation of Water Environmental Quality

From 1960s, because of self-purification capability decreased for water source deficiency and the effect of human activity, large amount of wastewater had flown into Baiyangdian Lake, which made the water environmental quality worse day by day[15] Fig.5 showed the dynamic changes of NH_4^+-N, COD_{Mn}, BOD_5, TP, TN in

south pollution zone, north pollution zone, and mixed pollution zone from 1988 to 2008. From it, we can find that North pollution zone was polluted most seriously in the three pollution zones, where the concentrations of NH_4^+-N and COD_{Mn} in 1994 and 1999-2008 were worse than water of Grade V. The concentrations of them were the highest in 2004, where the concentration of NH_4^+-N was 16 times the water quality standard of Grade V (2mg/L), and the concentration of COD_{Mn} was 2.2 times the water quality standard of Grade V (15mg/L). The concentrations of BOD_5 increased obviously from 1999 to 2005, worse than water quality standard of Grade V (10mg/L). The concentrations of TP were mostly between 0.2 mg/L and 1.8 mg/L, which were higher than surface water quality standard of Grade V (0.2mg/L). After the largest concentration in 1999, it showed a decreasing tendency. The concentrations of TN fluctuated obviously, which were 2 to 18 times the water quality standard of Grade V (2.0mg/L) each year, but generally showed a decreasing tendency. On the whole, the water quality of north pollution zone was at a worse than Grade V.

The pollution degree of south pollution zone was lower than that of north pollution zone, except that the TP and TN concentrations were comparatively higher, and the TP concentrations met the water quality standard of Grade V from 1999 to 2008, and the TN concentrations were worse than water quality standard of Grade V. The NH_4^+-N concentrations were mostly between 0.15 mg/L and 1.0 mg/L, which met the water quality standard of Grade Ⅲ. The COD_{Mn} concentrations were mostly between 2 mg/L and 10 mg/L, which were better than water quality standard of Grade Ⅳ. The BOD_5 concentration were within water quality standard of Grade Ⅳ(6mg/L) before 1999, and increased to the concentrations between the water quality standard of Grade Ⅳ-Ⅴ. The NH_4^+-N concentration in mixed pollution zone was affected by the pollution of both north and south pollution zone. But due to the affect of water body dilution in Baiyangdian Lake and the removal function of plants, which was decreased along the water movement path, the NH_4^+-N concentration was lower than both of the other two pollution zones.
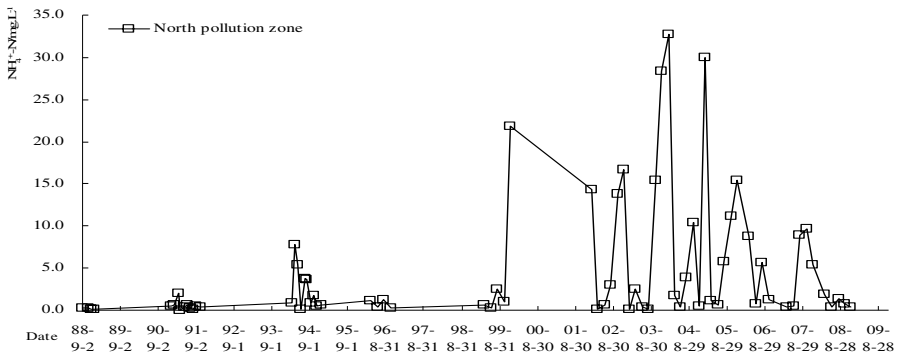
Fig. 5. Annual Pollutants Concentrations Dynamics
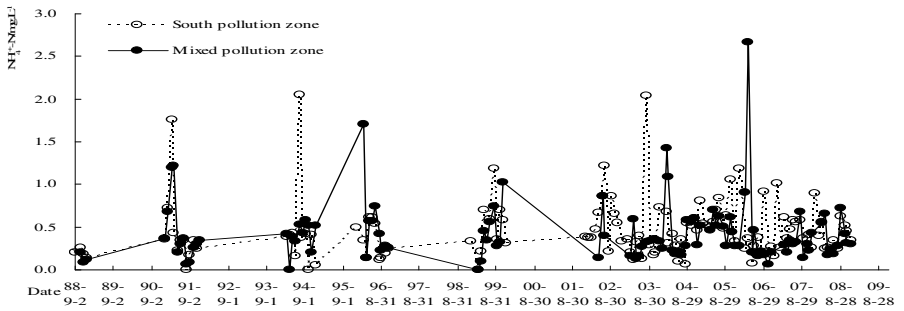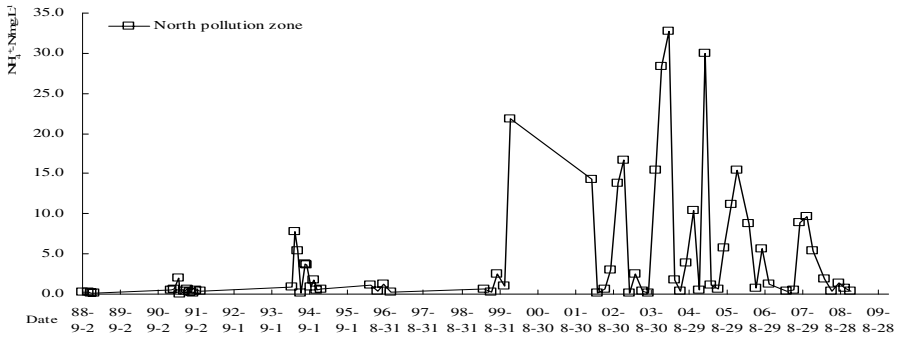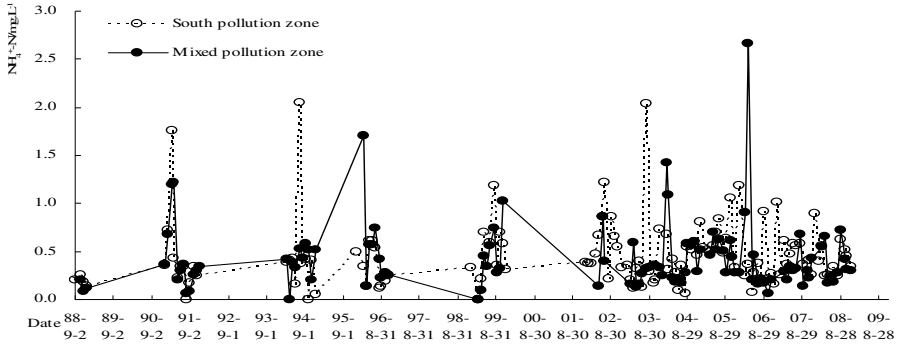
(a) NH$_4^+$-N

Fig. 5. (*Continued*)
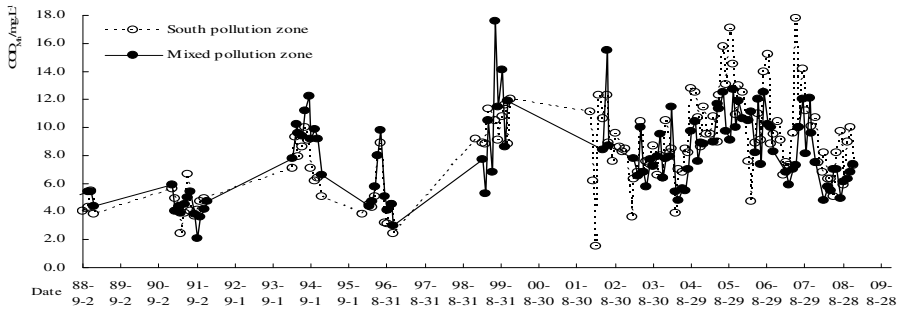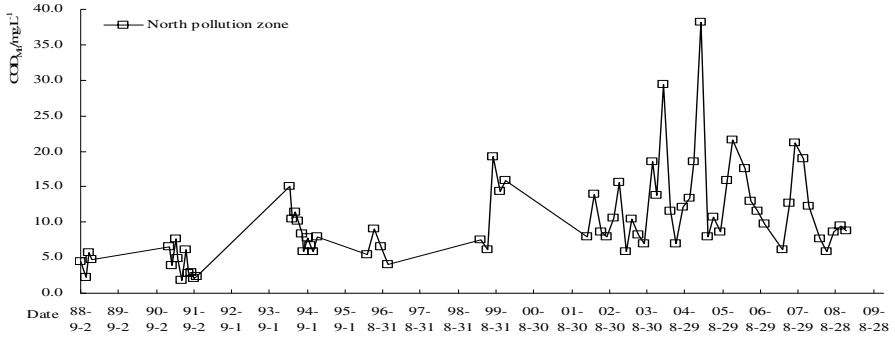
(b) COD$_{Mn}$
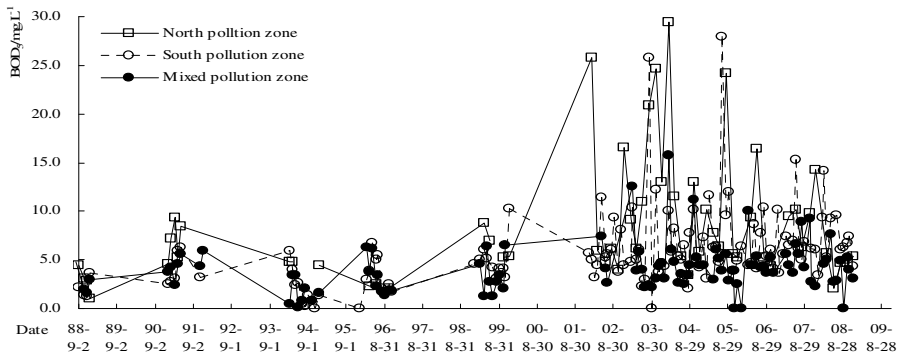
(c) BOD$_5$

Fig. 5. (*Continued*)
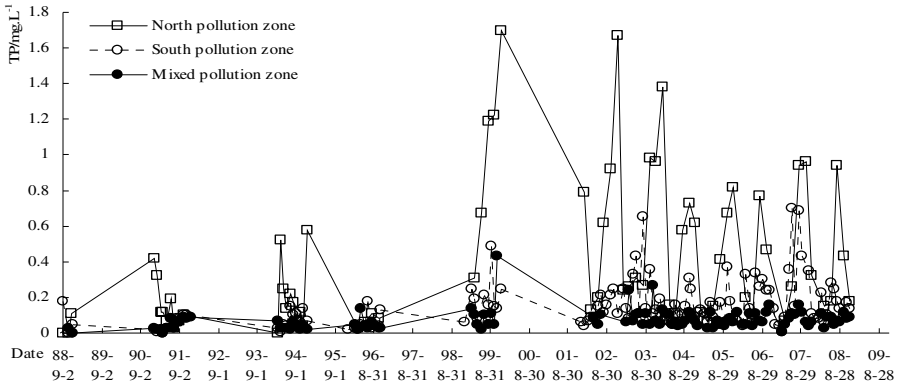
(d) TP
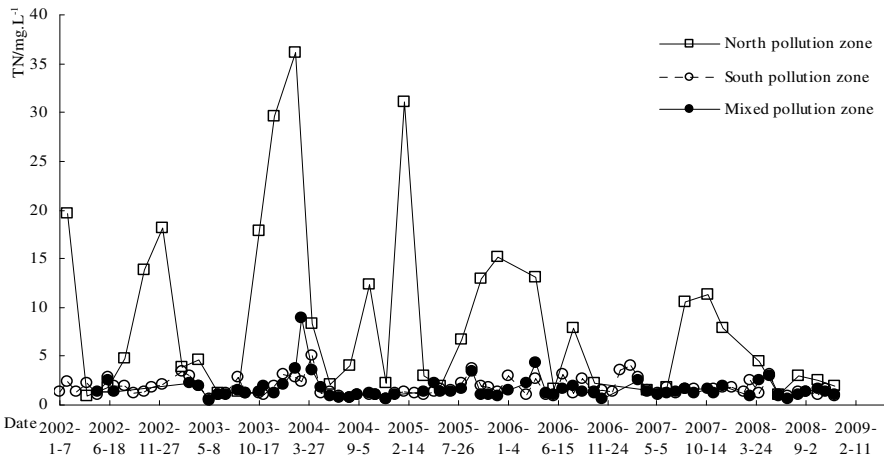
(e) TN

Fig. 5. (*Continued*)

Generally speaking, the concentrations of pollutants in different years fluctuated greatly from 1988, in which Baiyangdian Lake was refilled. The increasing of the concentrations of NH_4^+-N, COD_{Mn}, BOD_5, and TP were obvious, the TN concentration showed a decreasing tendency as a whole. Meanwhile, the main pollutants concentrations fluctuated obviously in a year. In normal and dry years, the pollution was more serious in summer, while the pollution was serious at the beginning of summer flood in wet years and then became less. It was the reason that upper regions of rivers were almost no flow and large amounts of pollutants stored in the rivers flowed into Baiyangdian Lake with the initial flood. For instance, a great flood occurred in 1996, the flood brought a large amount of pollutants in rivers at the beginning which caused the pollutants (such as NH_4^+-N, COD_{Mn}, and BOD_5) concentrations increased,

later with the increasing of water flowing into Baiyangdian Lake, the dilution function of river water emerged, which caused the pollutants concentrations decreased.

4   Analysis on Driving Mechanism of Water Eco-Environment Evolution

4.1   Climatic Changes in Baiyangdian Lake

4.1.1   Precipitation

Precipitation, as one of the main ways to income water in Baiyangdian Lake, had a vital effect on the evolution of water eco-environment. Average annual precipitation was 514.1mm in lake, and Fig.6 showed the results of precipitation in lake from 1974 to 2008. From it, we can find that the annual precipitation in Baiyangdian Lake changed greatly and the largest precipitation from 1955 to 2008 were 909mm (in 1996), 3.6 times as much as the smallest annual precipitation 252mm (in 1997). The precipitation showed a slow decreasing tendency before 1980s, while, the precipitation decreased rapidly after 1980s. It was present the phenomenon of obvious continuous wet and dry year. For instance, it was dry year occurred from 1983 to 1987 for five years, which directly caused Baiyangdian Lake dried up with a duration of 1650 days.
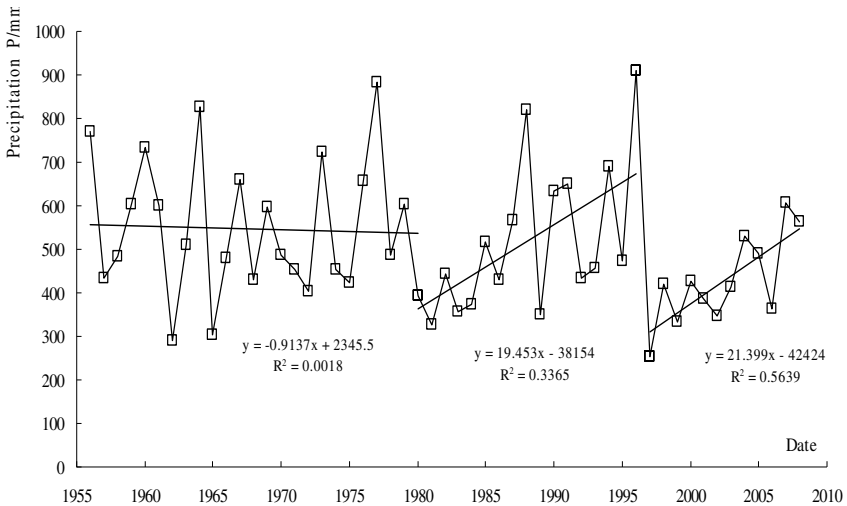
Fig. 6. Annual Precipitations of Baiyangdian Lake

4.1.2   Evaporation

Fig.7 and Fig.8 showed the results of evaporation in Baiyangdian Lake. The annual average evaporation in Baiyangdian was 1303.6 mm, this was 2.5 times the annual average precipitation. In the early 1970s, annual average evaporation with 10-year running means was 1289.0 mm, and decreased continuously. The lowest evaporation value was in the early 1990s, and then increased gradually to 1347.1mm in 2004.

We can see that, in the recent 30 years, water evaporation in Baiyangdian Lake has showed a increasing tendency which decreased at first and increased later. The significance test in Tab.6 also verified that the evaporation of Baiyangdian Lake showed a increasing tendency.

According to Fig.7, Fig.8, and the Mann-Kendall significance test in Tab.3, evaporation of different seasons except autumn in Baiyangdian showed unobvious increasing tendency, which passed the confidence level of 99%. The evaporation in summer increased the largest, which was 31.40 mm more than that of 1970s, and then was spring more 17.36mm than that of 1970s. The increase of air temperature would enhance evaporation capacity, so the increasing continuously temperature was the main impact factor caused the water surface shrinked[7].
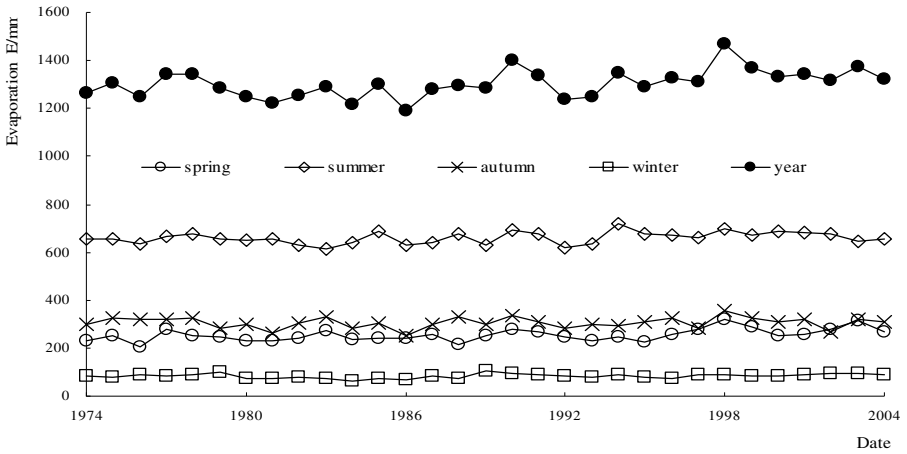
Fig. 7. Seasonal and Years Evaporations of Baiyangdian
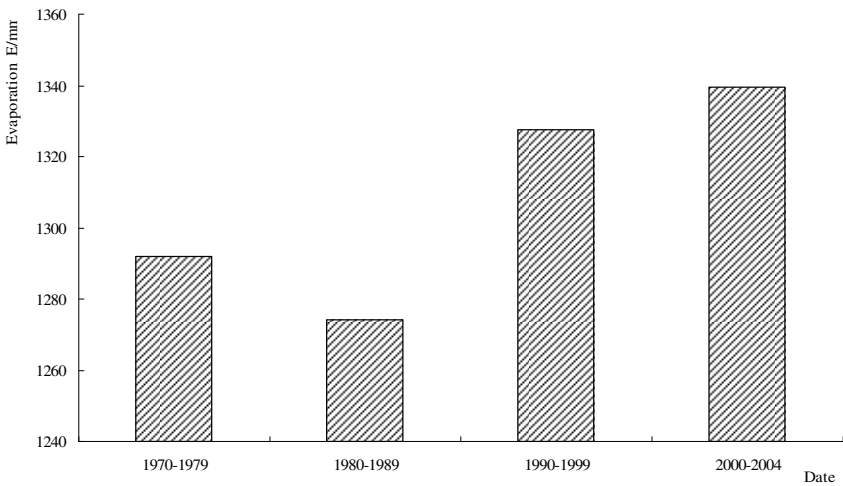
Fig. 8. Ten Yearly Evaporations of Baiyangdian

4.1.3   Air Temperature

Annual average air temperature in Baiyangdian was 12.42℃. Fig.9 and Fig.10 showed air temperature dynamic changes under three time scales of seasonally, yearly and ten-yearly, respectively, and Tab.3 showed the results of Mann-Kendall significance test. The result showed that annual average temperatures with 10-year running means in Baiyangdian Lake were increased and then decreased tendency in the past over 30 years. The annual average temperature in the 1980s was 0.13℃ higher than that of the 1970s. From the 1990s, the air temperature increased greatly, and the annual average temperature was 0.38℃ higher than that of 1970s. After 2000, the air temperature decreased, and the annual average temperature equal to the temperature of 1970s.
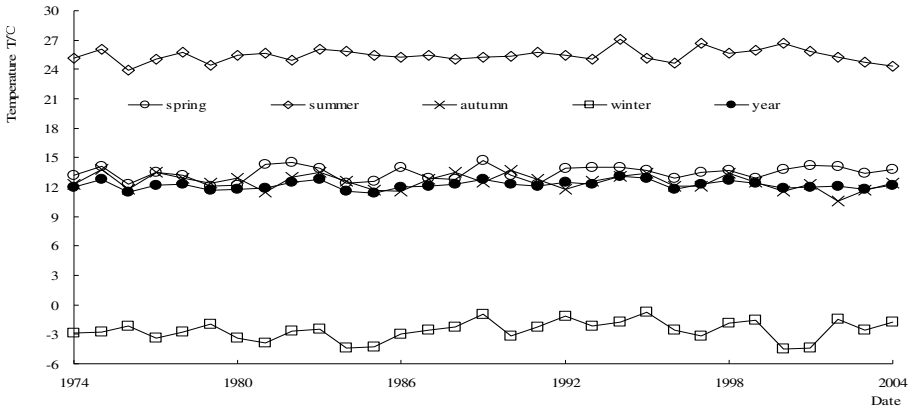
Fig. 9. Seasonal and Years Average Temperatures in Baiyangdian Lake
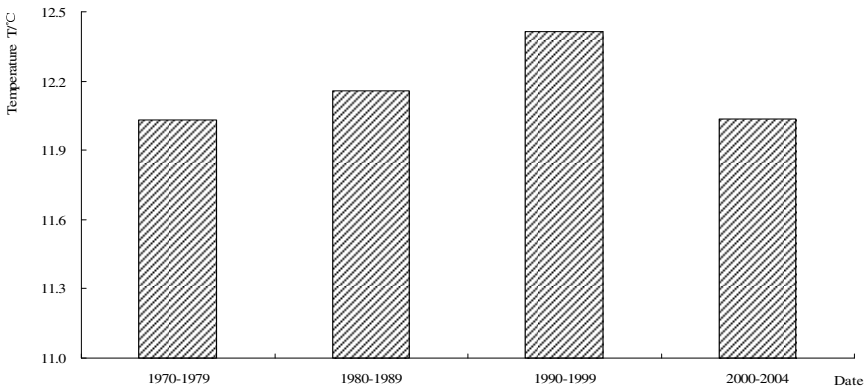
Fig. 10. Ten Yearly Average Temperatures in Baiyangdian

   Temperature changes of different seasons also showed obvious increasing tendency, which was more obvious in spring and winter. The average temperature in spring after the year 2000 was 0.94℃ higher than that of 1970s, and the average temperature in winter after the year 2000 was 0.21℃ higher than that of 1970s. The increase of temperature would cause all kinds of hydrological variables (such as the decreasing of precipitation, the increasing of soil temperature, the increasing of agricultural water consumption, and so on) changed, directly or indirectly caused the degeneration of Baiyangdian lake-wetland.

4.1.4   Solar Radiation

Fig.11 and Fig.12 showed solar radiation dynamic changes under three time scales of seasonal, yearly and ten-yearly, respectively. Tab.3 showed the results of Mann-Kendall significance test. The result showed that solar radiation always had existed a decreasing tendency in the past 30 years. The daily average solar radiation was 14.2MJ/(m^2·d) in the 1970s, and decreased to 14.0 MJ/(m^2·d) in 1980s, 12.8MJ/(m^2·d) in 1990s. The daily average solar radiation decreased a bit from the year 2000 to 2008, to 12.7MJ/ (m^2·d). According to the Mann-Kendall test, solar radiation of different seasons also showed obvious decreasing tendency, which was more obvious in summer and winter. The solar radiation in summer after the year 2000 was 2.09 MJ/ (m^2·d) lower than that of 1970s, and the solar radiation in winter after the year 2000 was 1.64MJ/ (m^2·d) lower than that of 1970s.
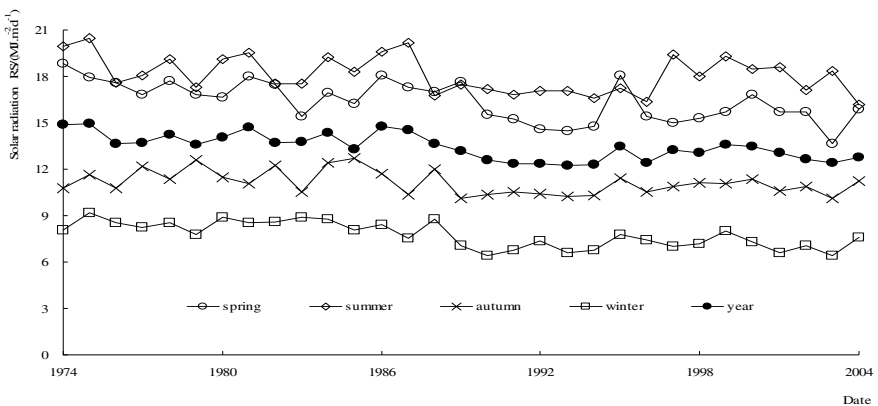
Fig. 11. Seasonal and Years Average Daily Solar Radiation in Baiyangdian Lake
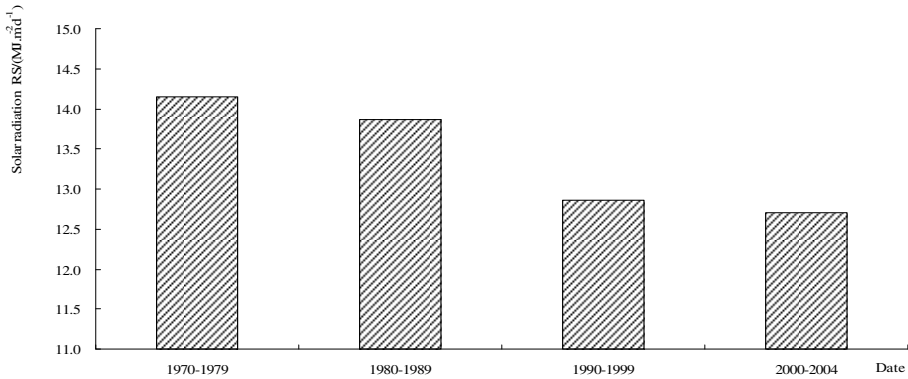
Fig. 12. Ten Years Average Daily Solar Radiation in Baiyangdian Lake

Table 3. Mann-Kendall Test Results of Yearly Evaporation, Air Temperature and Solar Radiation

Time \ Meteorological factors	Evaporation	Air temperature	Solar radiation
Year	1.30	4.08**	-4.63**
Spring	1.50	3.40**	-2.33*
Summer	1.19	0.38	-3.44**
Autumn	3.37**	0.62	-3.19**
Winter	1.07	3.95**	-4.20**

Note: * Correlation is significant at the 0.10 level(2-tailed)，

　　** Correlation is significant at the 0.01 level(2-tailed).

4.1.5   Relationship between the Local Climate and Water Level in Lake

Water level and meteor -ological factors of each year were carried on correlation analysis, the results of which were shown in Tab.4. The result showed that there was obvious positive correlation between water level and precipitation, the relativity of which was 0.483. And the water level showed negative correlation with temperature and evaporation, the relativity of which were 0.326 and 0.314, respectively. Water level and solar radiation were little related. Therefore, precipitation had a vital effect on the changes of Baiyangdian Lake eco-environment.

Table 4. Relativity between Water Level and Meteorological Factors

Pearson correlation	Precipitation	Air temperature	Evaporation	Solar radiation
Water level	0.483[**]	-0.326	-0.314	-0.008

Annotation:[**] Correlation is significant at the 0.01 level(2-tailed).

4.2   Human Activity

4.2.1   Eco-Hydrological Process in Upstream Watershed Changed Greatly, Directly Causing the Runoff into Lake Decreased Sharply

Watersheds are the source of lakes, and lakes are the converging of watersheds. Watersheds and lakes form a dynamic system which shows mutual feedback and close relationship between nature and society. At present, most of the researches on lakes and watersheds studied them respectively, which ignored the integrality of watershed hydrological process and ecosystem. The emphasis of some researches on lake water environment had already transferred to complicated lake-watershed macro-system, from watershed scale to carry through pollution prevention and cure, ecologic recovery, and ecosystem management. The changed eco-hydrological process in upstream watershed had significant effect on Baiyangdian lake-wetland.

The water storage of Baiyangdian Lake mainly relied on runoff formed by the precipitation in upstream watershed, which flowed into Baiyangdian through rivers. Runoff into Baiyangdian Lake in 1950s was abundant, but after the year 1958, 6 large reservoirs (Hengshanling, Koutou, Wangkuai, Xidayang, Longmen, and Angezhuang), 12 medium-sized reservoirs, and 116 small reservoirs were constructed in the upstream watershed in succession. The reservoirs controlled 54.93% of the watershed area in mountainous area, with a total storage of 3.619 billion m^3 （Zhang,Tian,Li,2007）, and had vital effect on intercepting the runoff from the upstream watershed. The building of water conservancy projects caused water flowing into Baiyangdian Lake decreased. In wet year, the runoff into Baiyangdian was abundant; in normal year, the control coefficient of reservoirs to runoff was 60%; while in dry year, the water in the reservoirs was mostly used for agricultural irrigation and little water was left, control coefficient to runoff was 96%.

The serious problems of Baiyangdian Lake, including resources water shortage, water quality-induced water shortage, unbalanced ecosystem, had attracted much attention from state council, ministry of water resources, Haihe River Commission, leaders and experts from departments in Hebei province. Deficiency of water was gradually acknowledged as the bottleneck of the Baiyangdian Lake eco-environmental management [16]. In order to ease this situation, water was refilled to Baiyangdian from upper region reservoirs for 28 times, with a total release of 2.35 billion m^3 and a total refill of 1.09 billion m^3 water to Baiyangdian Lake, and Table 5 showed the results. From it, we can find that Baiyangdian Lake had changed from a natural controlled lake-wetland to a totally artificial regulated lake.

Table 5. Situation of Different Reservoirs Supplying Water to Baiyangdian Lake

Year	Times	Water supplement source	Released water (million m³)	Water flowing into Baiyangdian (million m³)
1983	1	Angezhuang reservoir	25.07	9.79
	2	Xidayang reservoir	11.23	4.51
	3	Xidayang reservoi	51.75	15.10
	4	Xidayang reservoi	26.49	7.38
1984	5	Angezhuang reservoir	19.10	5.77
	6	Wangkuai reservoir	44.75	14.31
	7	Xidayang reservoi	31.16	12.19
	8	Wangkuai reservoir	45.14	27.09
1992	9	Xidayang reservoi	30.10	16.21
	10	Angezhuang reservoir	34.13	18.80
1997	11	Angezhuang reservoir	79.97	51.98
1998	12	Angezhuang reservoir	33.06	21.50
1999	13	Angezhuang reservoir	27.43	17.80
2000	14	Angezhuang reservoir	31.11	15.00
	15	Wangkuai reservoir	75.00	40.60
2001	16	Angezhuang reservoir	50.87	21.64
	17	Wangkuai reservoir	100.79	45.13
	18	Xidayang reservoi	50.15	35.01
2002	19	Xidayang reservoi	38.73	19.74
	20	Wangkuai reservoir	59.14	31.04
2003	21	Wangkuai reservoir	234.97	116.34
2004	22	Yuecheng reservoir	417.00	159.00
2005	23	Angezhuang reservoir	83.50	42.51
2006	24	Angezhuang reservoir	32.00	8.28
	25	Wangkuai reservoir	114.60	48.44
	26	Yellow river	200.00	100.00
2008	27	Yellow river	312.00	157.60
2009	28	Angezhuang reservoir	70.00	20.00
——	——	——	——	——
Total	——	——	2351.56	1090.36

Baiyangdian watershed is one of the primary grain-produce areas in north China, and the ratio of agricultural water is high. For instance, in 2006, the total water used in Baiyangdian watershed was 3.217 billion m³, and agricultural water was 2.696 billion m³, which accounted for 83.8% of the total water used. The construction of controlled reservoir in upstream watershed also greatly promotes the development of irrigation and drainage engineering. So far, 35 irrigation districts above 667 hectares have been

constructed, among which there are 4 large irrigation districts (Shahe, Tanghe, Yishui, Fanglaizhuo) above 2001 hectares and 31 medium irrigation districts (above 667 hectares and smaller than 2001 hectares). The whole effective irrigation area of cultivated land in watershed is 78706 hectares. The increasing of agricultural water in the upstream watershed would necessarily cause the runoff into Baiyangdian Lake decreased, especially in drought year, the conflict between agricultural water and water required in Baiyangdian Lake was more obvious. It would be aggravate rapidly the deterioration trend of water eco-environment in Baiyangdian Lake.

4.2.2   Environmental Hydrology Situation Changed a Lot in Upstream Watersheds, and Wastewater Discharge Quantity in the Watersheds Increased Rapidly

Except Baigouyin River of the 8 rivers whose water quality was good, water qualities of other rivers were worse than the water quality standard of Grade V. Fu River mainly undertook the domestic wastewater of Baoding City[17], Xiaoyi River mainly undertook the industrial wastewater from Li and Gaoyang counties, Cao River and Bao River mainly undertook the domestic and industrial wastewater from Mancheng and Xushui counties, respectively. Meanwhile, the increasing of agricultural land area in the upstream watersheds caused the application of chemical fertilizer and pesticide increased sharply, so the non-point pollution from agriculture was also one of the main pollution sources. The pollutants discharged from Fu River and Cao River were the most of the 8 rivers. Fu River flows through the center of Baoding City, Cao River flows through the northwestern part of Baoding City. In the past 40 years, the daily wastewater discharge in Baoding City increased continuously. The daily wastewater discharge of Baoding City in 1962 was 0.43×10^5 tons, which was 1.6×10^5 tons in 1974 and 2.1×10^5 tons in 1987. The daily wastewater discharge of Baoding City was 2.5×10^5 tons in 2005, and it was still increasing. However, at present there are only two wastewater treatment plants with a total daily treatment capacity of 1.6×10^5 tons, which results in 0.9×10^5 tons wastewater flowing into Fu River without being treated, and becomes the biggest pollution source in Baiyangdian Lake. Cao River was not polluted before the middle of 1980s. After the middle of 1980s, 156 paper mills were constructed in the upstream watershed of Cao River, and the daily wastewater increased to 0.8×10^5 tons, which was discharged into Cao River only after simple treatment except 0.3×10^5 tons wastewater was treated by Daceying Town wastewater treatment plant. In addition, with the domestic wastewater discharged from Mancheng County, Cao River had become a Sewage River. Fig.13 showed the monitoring results of the five main pollutants concentration (NH_4^+-N, COD_{Mn}, BOD_5, TP, TN) in the estuary of Fu River and Cao River to Baiyangdian Lake from 1988 to 2008. The result showed that all pollutants concentrations fluctuated greatly inter-annually and annually. The NH_4^+-N, COD_{Mn} and BOD_5 concentrations showed obvious decreasing tendency after the year 2002, however the TP and TN concentrations were more stable which showed a slight increasing tendency.
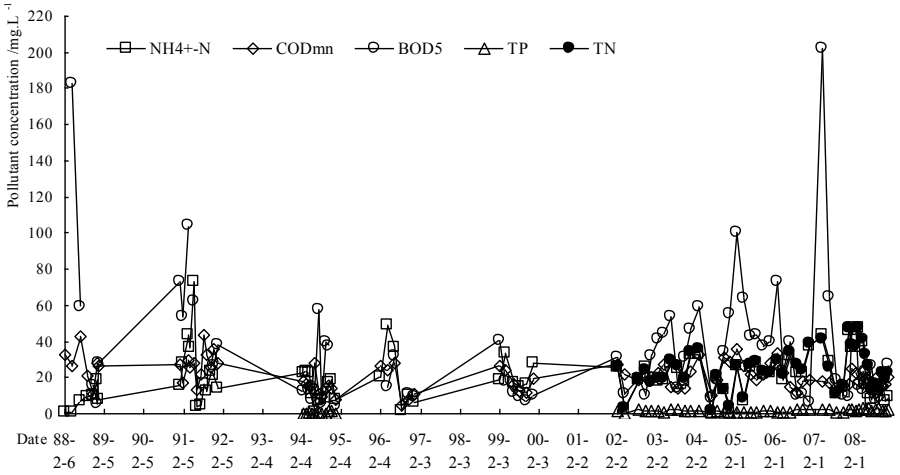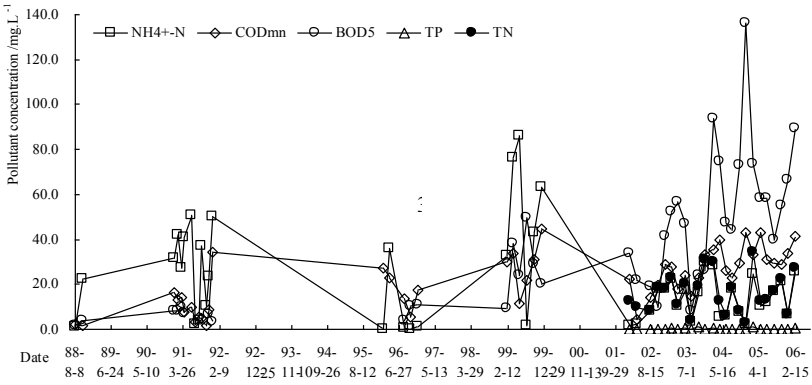
Fig. 13. The Dynamic Changes of the Concentrations of River Pollutants in Baiyangdian

From Fig.13 (a), it could be found that the water quality of Fu River was worse than the standard of Grade V, the five pollutants concentrations were far beyond the standards. The average concentrations of NH_4^+-N, COD_{mn} and BOD_5 were 19.88mg/L, 20.83mg/L and 34.29 mg/L, respectively, which were 10 times, 1.4 times and 3.5 times the water quality standard of Grade V. The annual average concentrations of TP and TN were 1.44mg/L and 24.34mg/L, which were 7 times and 12 times more than the standard of Grade V. From Fig.13 (b), it could be found that Cao River mainly discharged oxygen consumed pollutants, and the BOD_5 and COD_{Mn} concentrations were high, which were 3.6 times and 2.2 times the standard of Grade V. The average concentration of NH_4^+-N was 21.14mg/L, which was 10 times the standard of Grade V.

The concentration of TP was 0.27mg/L, which was similar to the standard of Grade V. The concentration of TN was 8 times the standard of Grade V.

4.2.3 The Sharp Increasing of Human Activity Directly Caused the Water Requirement in Baiyangdian Lake Increase Greatly

In the recent years, the population in Baiyangdian Lake increased greatly, which increased from 364,800 in 1989 to 387,800 by now, and the results of remote sensing interpretation also showed the residential land increased from 820.4hm^2 in 1987 to 1490.0 hm^2 in 2007. The increasing of population in Baiyangdian Lake directly caused the discharge of domestic wastewater increased.

After the water refilling with Baiyangdian Lake in 1988, the tourism developed quickly. As recorded, it received the tourists 3,704,000 from 1988 to 2000, and the peak of daily received tourists was 6833[18]. The water consumption and domestic pollutants of tourists, and the pollutants discharged by the motorboats accelerated the artificial pollution in Baiyangdian Lake.

From the 1950s, the counties and cities around Baiyangdian Lake continually developed agricultural irrigation. The agricultural water around lake increased from 2×10^7 m^3 per year in the early 1950s to2×10^8 m^3 now. The increasing of water diversion decreased the regulating storage volume of Baiyangdian Lake, which brought a series of environmental problems.

5   Results

Four categories of conclusions have been achieved based on the research in this paper:

(1) Monitoring with remote sensing on the water ecological landscape pattern in Baiyangdian Lake, it was found that the emergent plants, water body and cultivated land were the main land use types. During 1974 and 2007, the water body and emergent plants areas were decreased respectively by 28.7% and 20.8%, while the cultivated land areas were increased by 69.7%. And the heterogeneity and landscape fragmentation degree of different landscape types decreased constantly, but the latter wasn't to the significant difference level.

(2) During 1988 and 2008, the main pollutants (NH_4^+-N, COD_{Mn}, BOD_5, TP, TN) concentrations changed greatly with obvious spatial-temporal characteristics by analyzing the water quality data, and the pollutants concentrations showed increasing tendency. The pollution of North Zone was most severe in the whole lake, and the pollution was more serious in summer of normal and dry years, while the pollution was serious at the beginning of summer flood in wet years, and then became less.

(3) The local climate changed obviously in Baiyangdian Lake with Mann-Kendall test, which the precipitation decreased, while the air temperature and evaporation increased. The warming and drying climate tendency intensified water eco-environmental degradation, and the precipitation was the main impact factor on the changes water eco-environment in lake.

(4) Sharp increasing of human activity directly caused ecological/environmental hydrology regime in upstream watershed of Baiyangdian Lake changed obviously, which made it changed from a natural controlled lake-wetland to a totally artificial regulated lake.

References

1. Wang, H.: Study on creative ideas and key countermeasures of lake-wetland water ecological pollution treatment, pp. 22–25. Science Press, Beijing (2010)
2. Chinese Society of Environmental Sciences: The 13th World lake Conference, pp. 1–860. China Agricultural University Press, Beijing (2010)
3. Cui, B.S., Li, X., Zhang, K.J.: Classification of hydrological conditions to assess water allocation schemes for Lake Baiyangdian in North China. J. Hydrol. 385(1-4), 247–256 (2010)
4. Hu, G.C., Luo, X.J., Li, F.C., Dai, J.Y., Guo, J.Y., Chen, S.J., Hong, C., Mai, B.X., Xu, M.Q.: Organo chlorine compounds and polycyclic aromatic hydrocarbons in surface sediment from Baiyangdian Lake, North China: concentrations, sources profiles and potential risk. J. Environ. Sci. 22(2), 176–183 (2010)
5. Zhao, X., Cui, B.S., Yang, Z.F.: A study of the lowest ecological water level of Baiyangdian Lake. Acta Ecol. Sinica 25(5), 1033–1040 (2005)
6. Xu, M.Q., Zhu, J., Huang, Y., Gao, Y., Zhang, S., Tang, Y.: The ecological degradation and restoration of Baiyangdian Lake, China. J. Freshw Ecol. 13(4), 433–446 (1998)
7. Gao, Y.C., Wang, H., Long, D.: Changes in hydrological conditions and the eco-environmental problems in Baiyangdian watershed. Resour. Sci. 31(9), 1506–1513 (2009)
8. Li, Y.H., Cui, B.S., Yang, Z.F.: Influence of hydrological characteristic change of Baiyangdian on the eco-environment in wetland. J. Nat. Resour. 19(1), 62–68 (2004)
9. State Forestry Administration, P.R. China. Technical Specification for Investigating the National Wetland Resources (2009)
10. Vapink, V.N.: The nature of statistical learning theory, pp. 1–24. Springer, New York (2005)
11. Oded, M., Lior, R.: The data mining and knowledge discovery handbook, pp. 41–84. Springer Science and Business Media Inc. (2005)
12. Hofmann, H., Lorke, A., Peeters, F.: Temporal scales of water-level fluctuations in lakes and their ecological implications. Hydrobiologia 613, 85–96 (2008)
13. Chen, L.D., Fu, B.J.: Analysis of impact of human activity on landscape structure in Yellow River Delta - a case study of Dongying Region. Acta Ecol. Sinica 16(4), 337–344 (1996)
14. Li, C.Z., Yu, F.L., Liu, J.: Dynamic change of landscape and its driving forces in midstream of Heihe mainstream basin after water redistribution. Acta Ecol. Sinica 29(11), 5832–5842 (2009)
15. Zhong, P., Yang, Z.F., Cui, B.S., Liu, J.L.: Eco-environmental water demands for the Baiyangdian Wetland. Environ. Sci. Eng. China 2, 73–80 (2008)
16. Zhang, S.Z., Tian, J.W., Li, G.B.: Ecological problems and restoration measures of Baiyangdian Wetland. Bull. Soil and Water Conserv. 27(3), 146–150 (2007)
17. Hu, G.C., Dai, J.Y., Mai, B.X., Luo, X.J., Cao, H., Wang, J.S., Li, F.C., Xu, M.: Concentrations and accumulation features of organo chlorine pesticides in the Baiyangdian Lake freshwater food web of North China. Arch. Environ. Contam. Toxicol. 58(3), 700–710 (2009)
18. Zhang, M.Y.: Study on the change of eco-environment in Baiyangdian watershed based on RS, GIS and landscape pattern analysis, pp. 1–58. Nanjing Normal University, Nanjing (2004)

Application and Evaluation of Wavelet-Based Denoising Method in Hyperspectral Imagery Data

Hao Yang[1], Dongyan Zhang[1,2], Wenjiang Huang[1], Zhongling Gao[1], Xiaodong Yang[1], Cunjun Li[1], and Jihua Wang[1,*]

[1] Beijing Research Center for Information Technology in Agriculture, 100097 Beijing, China
wangjh@nercita.org.cn
[2] Institute of Agricultural Remote Sensing & Information Technology Application, Zhejiang University, 310029 Hangzhou, China

Abstract. The imaging hyper-spectrometer is highly susceptible to the presence of noise and its noise removal is regularly necessary before any derivative analysis. A wavelet-based(WT) method is developed to remove noise of hyperspectral imagery data, and commonly used denoising methods such as Savitzky-Golay method(SG), moving average method(MA), and median filter method(MF) are compared with it. Smoothing index(SI) and comprehensive evaluation indicator(η) are designed to evaluate the performance of the four denoising methods quantitatively. The study is based on hyperspectral data of wheat leaves, collected by Pushbroom Imaging Spectrometer (PIS) and ASD Fieldspec-FR2500 (ASD) in the key growth periods. According to SI and η, the denoising performance of the four methods shows that WT>SG=MA>MF and WT>MA>MF>SG, respectively. The comparison results reveal that WT works much better than the others with the SI value 0.28 and η value 5.74E-05. So the wavelet-based method proposed in this paper is an optimal choice to filter the noise, in terms of balancing the contradiction between the smoothing and feature reservation ability.

Keywords: imaging hyper-spectrometer, noise, filtering, wavelet analysis, quantitative evaluation.

1 Introduction

Hyperspectral imaging system is characterized by integrating both spectroscopy and imaging techniques in one system to provide detailed information of target which otherwise cannot be achieved with either conventional imaging or spectroscopy alone[1]. The simultaneous spatial and spectral information provided by this system along with its non-destructive and chemical-free nature nominated this technology to be a deliberated tool for monitoring crop diseases[2], vegetation nutrition[3], food safety[4] etc.

---

* Corresponding author.

But the obtained data is highly susceptible to the presence of noise. The sensors because of their narrow bandwidth are not able to capture a lot of energy and cause self-generated noise inside the sensor. Variations and atmospheric absorption of sun illumination also reduce the amount of recorded spectral signal[5]. Therefore, noise removal is regularly necessary before any derivative investigation.

Many scholars have developed different methods for how to reduce noise of spectral signal[6-9] and 2-dimension image[10-13]. At the beginning, noise reduction applied spatial average method directly, such as moving-average or median filter method based on multi-points. But it was found that in the shortwave infrared spectrum, the mean and median filter cannot remove noise satisfactorily and a slight jaggedness remains[14]. With the development of fast Fourier transform and discrete Fourier transform, noise reduction method in frequency domain[15] become acceptable. Moreover, a number of methods for hyperspectral data dimensionality reduction have been developed, such as principal component analysis (PCA) [11,16,17], linear discriminate analysis (LDA)[18] etc. Especially in recent years, booming wavelet technology aims to provide a new way of thinking, there are many wavelet-based and improved filter has been widely applied[7,10,19,20].

However, these denoising methods aforementioned were designed mainly for field spectrometer(ASD, e.g.), or spspaceborne and airborne hyperspectral sensors (CHRIS/PROBA, OMIS, e.g.) where noise was caused by atmosphere and water vapour etc. Rarely research about denoising of imaging hyper-spectmeter data was reported. Recently, a new field-based Pushbroom Imaging Spectrometer(PIS) is developed independently by own. So more original data can be acquired, nevertheless coupled with more noise signal, which due to physical instruments, sun's illumination, field environment condition and others. An ideal method which is simple and effective to reduce noise of the collected data is an actual need. Furthermore, as to the evaluation of denoising effect, most research is based on visual judgment or simple signal noise ratio(SNR), which always can not satisfy requirements for their complex nature.

This paper attempts to find an optimal denoising method aiming at homemade Imaging Spectrometer (PIS). The rest of this paper is structured as follows. Section 2 describes the hyperspectral imagery data collection and introduces four denoising methods along with the quantitative evaluation. Section 3 presents the denoising result and evaluating result. Conclusions are drawn in Section 4.

2   Materials and Methods

2.1   Hyperspectral Data Collection

The study site located in the National Experimental Station for Precision Agriculture, the Changping district of Beijing, China (40°11′N, 116°27′E). The cultivar of winter wheat, Jing 411 was chosen as research material and sown in October, 2009. When the winter wheat grown at jointing stage, 30 leaves samples were picked and corresponding hyperspectra was collected by PIS (Pushbroom Imaging Spectrometer) and ASD (FieldSpec®3portable spectrometer) simultaneously. The detailed process was as follows: Put leaves flatly on the black cloth and corresponding spectrum was measured

with 120cm distance from the lens of PIS, the experimental time was from 10:30am to 15:30pm in a sunny and windless day.

PIS was developed by National Engineering Research Center for Information Technology in Agriculture and University of Science and Technology of China. Before the instrument was applied, the laboratory calibration had been carried out to determine location of the wavelengths, spectral response function, radiation accuracy and spatial displacement etc. The specific technical parameters of PIS and ASD were presented in Table 1.

Table 1. Specific technical parameters of PIS and ASD

Parameters	PIS	ASD
Spectral Range	400-1000nm	350-2500nm
Spectrum Resolution	To 2nm	3nm@700nm; 10nm@1400nm&2100nm
Spectrum Interval	0.7nm	1.4nm,350-1000nm; 2nm, 1000-2500nm
Spatial Resolution	To 5mm	/
Field of View	16°	25°

2.2 Data Pre-processing

Hyperspectral imagery data can be viewed as a 3-dimensions imagery sequence, which is constituted by 2-dimensions image and 1-dimension spectrum. The 2-dimensions image describes target's spatial information feature while 1-dimension spectrum information reveals spectral profile feature of every pixel in image. For every pixel in image acquired by PIS, a sequence of digital number(DN) was obtained with spectrum interval of about 0.7nm.

For the purpose of follow-up quantitive research, DN value of wheat leaves must be transformed to relative reflectance which involved with instrument's calibration. In calibration, linear empirical method was adopted.

$$DN = a\rho + b \tag{1}$$

In Eq.(1), the coefficient a and b was determined by experiment data and least square method. And then the reflectance$\rho$was calculated.

ASD data is also be transformed to reflectance through standard reference white panel.

$$\rho_{t\arg et} = (DN_{t\arg et} / DN_{panel}) \times \rho_{panel} \times 100\% \tag{2}$$

In Eq.(2), $\rho_{t\arg et}$ and ρ_{panel} mean reflectance of target and standard reference white panel, $DN_{t\arg et}$ and DN_{panel} indicate DN value of target and standard reference white panel, respectively.

2.3   Denoising Methods

In this paper, aiming to the quantitative analysis of crop physiological and biochemical parameters, a wavelet transform method was explored to remove noise of hyperspectral imagery data, and common denoising methods such as Savitzky-Golay method(SG), moving average method(MA) and median filter method(MF) were also chosen and compared with it. Some indicators were designed to evaluate abilities of noise removal.

2.3.1   Wavelet Transform

Denoising techniques based on the wavelet transform, also known as wavelet shrinkage has been proposed to eliminate noise in various types of signals, by applying the transform and zeroing the coefficients below a certain threshold value, on the basis that noise coefficients would have lower amplitude than the coefficients corresponding to the feature being studied. Additionally, the coefficients remaining after the threshold operation may be reduced by the threshold value as well; this is known as soft threshold. It has been shown by Donoho and Johnstone (1994) that by using the appropriate threshold value, soft threshold results in the optimum denoising with the least distortion in the signal.

In this paper, threshold-based wavelet denoising algorithm includes mainly three steps.

1) The discrete wavelet transform and selection of wavelet filter and number of decomposition layers:

$$Y(d) = Y(f) + Y(z) \tag{3}$$

In Eq.(3), $Y(d)$, $Y(f)$, $Y(z)$ is high-frequency wavelet transform coefficient of noisy data, truth data and noise data.

2) Choose a threshold for every level high-frequency coefficient and wavelet transform coefficient $\hat{Y}(f)$ was estimated by the threshold. Soft threshold and hard threshold can be used such as Eq.(4) and Eq.(5). T is threshold in them.

For soft threshold: $\hat{Y}(f) = \begin{cases} \text{sgn}(Y(d))(Y(d)-T) & |Y(d)| \geq T \\ 0 & |Y(d)| < T \end{cases} \tag{4}$

For hard threshold: $\hat{Y}(f) = \begin{cases} Y(d) & |Y(d)| \geq T \\ 0 & |Y(d)| < T \end{cases} \tag{5}$

3) Wavelet reconstruction $\hat{Y}(f)$ to get estimation of $\hat{f}(t)$:

$$\hat{f}(t) = \omega^{-1}\hat{Y}(f) \tag{6}$$

In Eq.(6), ω-1is the operator of wavelet inverse transformatin.

The wavelet transform filter method lies in the choice of the threshold scheme, commonly used scheme is Rigrsure, Sqtwolog, Heursure and Minimaxi. In them, Rigrsure selects an adaptive threshold based on Stein unbiased likelihood principle,

Sqtwolog adopts fixed threshold $\sqrt{2\times\log(length(x))}$, Heursure is a mixture of the first two,   Minimaxi selects the mean square extreme error under non-ideal circumstance as the threshold.

The denoising effect is influenced by threshold adjustment ways. If the adjustments, there are two ways: one is that threshold is adjusted according to the first level noise(global threshold), while the other one is according to each level noise(hierarchy threshold).

2.3.2   Savitzky-Golay Filter, Moving Average, Median Filter

The method of Savitzky-Golay filtering for hyperspectral analysis is discussed in detail by King et al. (1999). Savitzky-Golay filtering has its roots in least squares polynomial smoothing which is to find a smoothed value for each point in the spectrum of a subset of data within a window. The window contains the point to be smoothed in the center position of the window as well as several of its neighbors to either side in the spectrum. All the data within the window are used to perform the least squares fit, but only the central point is smoothed for each window position. The other points are smoothed by moving the window across the spectrum point by point, performing a least squares approximation to the windowed data at each location.

Moving average takes the mean spectral value of all points within a specified window as the new value of the middle point of the window(Tsai and Philpot, 1998). The method is solely based on linear calculations and has one key parameter, the filter size. The median filter is similar to moving average but it is median value not average value.

2.4   Evaluating Methods

Smoothing methods may cause some loss of information or change the original spectral so they will produce incorrect results in subsequent analyses such as red edge discrimination. Thus, smoothing techniques besides being powerful in the smoothing ability, should be used in the way that preserves the absorption and local spectral features like minima, maxima and inflection points. Fine filtering method evaluation criteria should be that the denoised spectral curve remains curve characteristics (curve shape and location of characteristics), and is as smooth as possible.

A smoothing index(SI) was designed to measure the smoothness of denoised spectral curve. In Eq.(7), n is number of wavebands, $\hat{\rho}$, ρ is the denoised spectrum and original spectrum respectively; the SI is smaller, it means the denoised spectrum is smoother.

$$\text{Smoothing index:} \qquad SI = \frac{\sum_{i=1}^{n} |\hat{\rho}_{i+1} - \hat{\rho}_i|}{\sum_{i=1}^{n} |\rho_{i+1} - \rho_i|} \qquad (7)$$

Signal-to- Noise Ratio(SNR), Mean Square Error(MSE), Spectral Angle(SA), were also be used to evaluate smoothing effect. As the denoising effect is proportional to SNR while MSE is inversely proportional to SA, we designed a comprehensive

evaluation indicator to ensue the consistency of the aforementioned three indicators. They are defined as follows:

$$\text{Mean Square Error:} \quad MSE = n^{-1} \sum_{i=1}^{n} (\hat{\rho}_i - \rho_i)^2 \tag{8}$$

$$\text{Signal-to-Noise Ratio:} \quad SNR = 10 \cdot \log \frac{\sum_{i=1}^{n} \rho_i^2}{\sum_{i=1}^{n} (\hat{\rho}_i - \rho_i)^2} \tag{9}$$

$$\text{Spectral Angle:} \quad SA = \arccos \frac{\sum_{i=1}^{n} \rho_i \hat{\rho}}{\sqrt{\sum_{i=1}^{n} \rho_i^2} \cdot \sqrt{\sum_{i=1}^{n} \hat{\rho}_i^2}} \tag{10}$$

$$\text{Comprehensive evaluation indicator:} \quad \eta = \frac{MSE \cdot SA}{SNR} \tag{11}$$

In above equations, $\hat{\rho}$, ρ is the denoised spectrum and original spectrum, n is number of wavebands respectively.

3   Results and Analysis

The result of applying wavelet transform denoising together with other methods was demonstrated in Figure 1. We chosen an area in the middle of leave and got its DNs average in the area in every band, so an average spectrum was acquired. In fact, this process meant the correlation between pixels in spatial domain is utilized. The average spectra were showed in blue curves in Figure 1. Red curves displayed the results after noise removal. All algorithm was realized in MATLAB platform. In wavelet transform filter, sqtwolog, soft threshold, hierarchy threshold scheme were selected, the number of decomposition layers was set to 5, wavelet radix chosen "symlets20". For SG filter, the number of envelope iterations was 2 and window size was 5. And in MA and MF filter, the filter step was 5.

For four methods, the effects of noise removal can be seen in the Fig.1. The obvious noises were removed and the whole shapes of spectral curves were preserved. When compared to the other, WT was smoother than others, especially in wavelength greater than 480nm. Contrasting the curves in 750-950nm, some small disturbances remained in Fig.1(b,c,d) while Fig.1(a) was smooth.

Evaluating methods in section 2.4 were applied to further assess the effect of denoising quantitatively. The result was showed in Table 2. In Table 2, two critical indicators should be concerned: SI and η. SI represents smoothness, the less means the smoother. WT had the minimum value of 0.28, MF had the maximum of 0.36. The denoising effect was WT>SG=MA>MF.η measured denoising ability from another

view. Similarly, WT had the minimum value of 5.74E-05, it meant WT had the best performance. The effect was WT>MA>MF>SG. As a result, both of evaluating indicators showed WT had the best denoising performance.

Above result was based on region of pixels. For single pixel, the result of WT filter was showed in Fig.2. The red curve represented denoising spectrum. It can be found WT was still very efficient though much noise existed in this case. So the WT filter was stable in denoisng, especially in >480nm wavelength range.
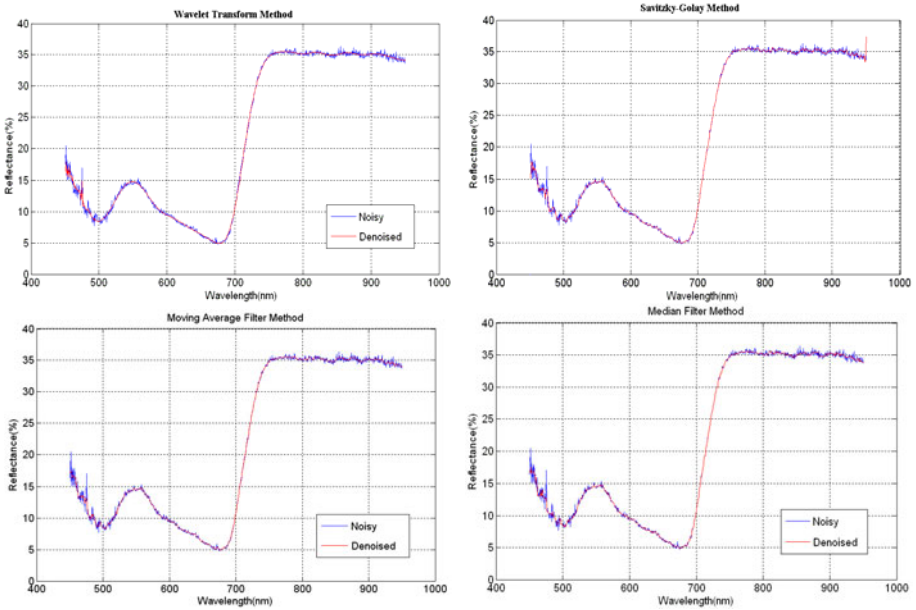
Fig. 1. The results of four denoising methods
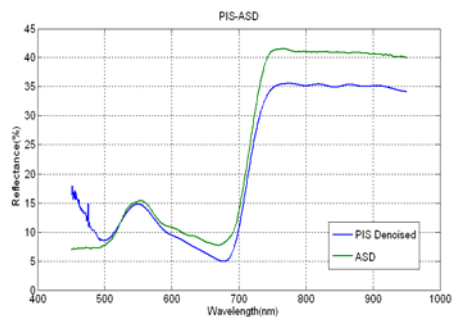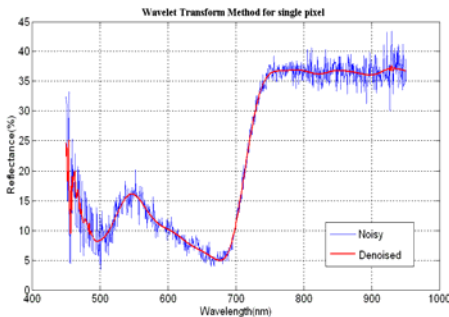
Fig. 2. Denoising result of single pixel by WT   **Fig. 3.** Denoising Spectra of ASD and PIS by WT

Table 2. The statistics of evaluation results for four methods

Indicator	SI	SNR	MSE	SA	η
WT	0.28	36.56	0.14	0.015	5.74E-05
SG	0.35	35.12	0.20	0.018	1.03E-04
MA	0.35	36.15	0.16	0.016	7.08E-05
MF	0.36	35.65	0.18	0.016	8.08E-05

When the selected region from PIS image was same to the view field of ASD, the same target was measured. Owing to filter algorithm was built-in ASD, the obtained ASD spectrum had little noise and can be viewed as reference spectrum. Denoised PIS spectrum by WT and the corresponding ASD spectrum was showed in Fig. 3. As shown, the former approach to the later. The trend was very consistent though absolute reflectance differences. The absolute reflectance difference can be explained by the two instrument sensitivity itself, illumination condition, calibration accuracy, and time etc.

4   Discussion and Conclusions

This study presented an approach to removing noise of hyperspectral imagery data which were collected by PIS. Aiming to the quantitative inversion of crop physiological and biochemical parameters, the wavelet-based method was developed, and commonly used denoising methods such as Savitzky-Golay method(SG), moving average method(MA), and median filter method(MF) were compared with it. And a smoothness index and a comprehensive evaluation indicator were designed to appraise denoising effects. The denoised effects of four methods were assessed by them. For index SI, the denoising effect was WT>SG=MA>MF; forη, the result was WT>MA>MF>SG. So WT had the best noise removal performance. The denosing result by WT was also compared with ASD data, the effect was close to each other. As a result, WT filter was a very efficient means to removing noise for hyperspectral imagery data.

Though WT method showed powerful denoised ability, there was space for further research. Firstly, this paper focused mainly on spectrum dimension, and put little consideration on image dimension noise reduction. Secondly, for evaluation of denoised effect, the characteristic preservation ability can be evaluated through detailed key feature such as location of small peak and valley which depends on specific application. Denoising data was applied to biochemical parameters inversion, and then best denoising method was evaluated. In addition, the factors, such as illumination, temperature, wind, and human are all affect the experimental results, so how to ensure optimal spectrum information that reflects truth need further to study.

References

[1] Wang, J.H., Zhao, C.J., Huang, W.J.: Basis and Application of Quantitative Remote Sensing in Agriculture, pp. 141–184. Science Press, Beijing (2008) (in Chinese)

[2] Singh, C.B., Jayas, D.S., Paliwal, J., White, N.D.G.: Detection of insect-damaged wheat kernels using near-infrared hyperspectral Imaging. Journal of Stored Products Research 45, 151–158 (2009)

[3] Nguyen Do Trong, N., Tsuta, M., Nicola, B.M., De Baerdemaeker, J., Saeys, W.: Prediction of optimal cooking time for boiled potatoes by hyperspectral imaging. Journal of Food Engineering 105, 617–624 (2011)

[4] Liu, L., Ngadi, M.O., Prasher, S.O., Gariépy, C.: Categorization of pork quality using Gabor filter-based hyperspectral imaging technology. Journal of Food Engineering 99, 284–293 (2010)

[5] Yusof, M.R.M.: Trends and Issues in Noise Reduction for Hyperspectral Vegetation Reflectance Spectra. European Journal of Scientific Research 29(3), 404–410 (2009)

[6] Wang, Y., Mo, J.: A New De-Noising Technique for Spectra Based on Mexican Hat Wavelet. Spectroscopy and Spectral Analysis 25(1), 124–127 (2005)

[7] Zhou, D., Wang, Q., Tian, Q., Lin, Q., Fu, W.: Wavelet Analysis and Its Application in Denoising the Spectrum of Hyperspectral Image. Spectroscopy and Spectral Analysis 29(7), 1941–1945 (2009)

[8] Hu, B., Li, Q., Smith, A.: Noise reduction of hyperspectral data using singular spectral analysis. International Journal of Remote Sensing 30(9), 1954–1957 (2009)

[9] Jin, L., Wan, W., Wu, Y., Cui, B., Yu, X.: A General Framework for High-Dimensional Data Reduction Using Unsupervised Bayesian Model. CCIS, vol. 98(2), pp. 96–101 (2010)

[10] Sun, L., Gu, D.-F., Luo, J.-S.: Hyperspectral Imagery Denoising Method Based on Wavelets. Spectroscopy and Spectral Analysis 29(7), 1954–1957 (2009)

[11] Chen, G., Qian, S.: Simultaneous dimensionality reduction and denoising of hyperspectral imagery using bivariate wavelet shrinking and principal component analysis. Can. J. Remote Sensing 34(5), 447–454 (2008)

[12] Gómez-Chova, L., Alonso, L., Guanter, L., Camps-Valls, G., Calpe, J., Moreno, J.: Correction of systematic spatial noise in push-broom hyperspectral sensors: application to CHRIS/PROBA images. Applied Optics 47(28), 46–60 (2008)

[13] Minh, N.Q.: Image smoothing of multispectral imagery based on the HNN and geo-statistics. Journal of Remote Sensing 15(3), 640–644 (2011)

[14] Schmidt, K.S., Skidmore, A.K.: Smoothing vegetation spectra with wavelets. International Journal of Remote Sensing 25(6), 1167–1184 (2004)

[15] Atkinson, I., Kamalabadi, F., Jones, D.L.: Wavelet-based hyperspectral image estimation. In: Proceedings of International Geoscience and Remote Sensing Symposium, vol. 2, pp. 743–745 (2003)

[16] Jolliffe, T.: Principal component analysis. Springer, New York (2002)

[17] Chang, W., Guo, L., Liu, K., Fu, C.: A noise removal method for hyperspectral data based on Contourlet transformation and PCA analysis. Journal of Electronics & Information Technology 31(12), 2892–2896 (2009)

[18] Fukunaga, K.: Introduction to statistical pattern recognition. Academic Press, San Diego (1990)

[19] Othman, H., Qian, S.-E.: Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage. IEEE Transactions on Geosciences and Remote Sensing 44, 397–408 (2006)

[20] Huang, M., Wang, K., Shi, Z., Gong, J., Li, H., Chen, J.: Quantitative Evaluation of Soil Hyperspectra Denoising with Different Filters. Spectroscopy and Spectral Analysis 29(3), 722–725 (2009)

Estimation of Maize Planting Area through the Fusion of Multi-source Images[*]

Xiaohe Gu, Yuchun Pan[**], Xin He, and Jihua Wang

Beijing Agriculture Information Technology Research Center, Beijing 100097, China
{guxh,panic}@nercita.org.cn

Abstract. The limitations on spatial resolution and on the availability and measurement accuracy of remote sensing images are the primary problems in the estimation of the large-scale planting area for maize. The integration of mid- and low-resolution images is the one of primary methods used for the estimation of large-scale crop planting areas using remote sensing. The use of a single-temporal thematic mapper (TM) image results in a low accuracy of maize recognition, so a mid-scale time-series normalized difference vegetation index (NDVI) dataset, which was derived from the fusion of the moderate-resolution imaging spectroradiometer (MODIS) and TM images based on the wavelet transform, was established. The planting area was estimated using the minimum distance model and the accuracy was evaluated using in-situ samples. The results show that the estimation of the maize-sown area based on the time-series NDVI information of the integrated images reached high levels of gross and position accuracy (89% and 90%), indicating that this method could fully utilize the time-series information from the MODIS images and the spatial resolution of a TM image. The use of the difference in phenophases among fall crops enables the effective classification of the spatial distribution of these crops.

Keywords: maize, planting area, MODIS, TM, remote sensing, wavelet transform, minimum distance.

1   Introduction

Maize is one of the primary crops in China, the planting area and yield of which comprise one-fifth of all crops [1]. Accurate information on the planting area and distribution of maize are important for effective yield estimation, agricultural management, and national food security [2]. The spatial characteristics of maize distribution in China include complex planting structures and fragmented planting parcels. The limitations on spatial resolution and on the availability and measurement of the accuracy of remote sensing images are the primary problems in the estimation of the large-scale planting area for maize [3].

Previous studies show that the integration of mid- and low-resolution images is the primary methods used for investigating large-scale crop planting areas [4 – 7]. A mid-resolution image has a high level of accuracy in crop recognition, but lacks availability and coverage [8 – 9]. On the other hand, a low-resolution image enables the acquisition of long time-series information that match crop growth [10 – 11]. Thus, utilizing mid- and low-resolution images to conduct crop recognition through the integration of the multi-source data is necessary. Currently available studies on crop planting area estimation primarily focus on winter wheat [12–13] and rarely emphasize maize. A number of crops are synchronized with maize, so this crop is of great significance for research on the estimation of planting areas through the integration of mid- and low-resolution images.

The Yuanyang County in Henan province was selected as the study area. The data used in the current paper included multi-temporal moderate resolution imaging spectroradiometer (MODIS) normalized difference vegetation index (NDVI) images and a single-temporal thematic mapper (TM) NDVI image. A mid-scale time-series NDVI dataset, which was derived from the fusion of the MODIS and TM images based on wavelet transform, was established because of the low accuracy of maize recognition when using only a single-temporal TM image. The planting area was then estimated using the minimum distance model and the accuracy was evaluated using in-situ samples.

2   Study Area and Data Preprocess

2.1   Study Area

The Yuanyang County in Henan province, a typical agricultural region where the planting structures are stable and the fall crops include maize and paddy, was chosen as the study domain. The phenological differences of the crops are often used to select the temporal remote sensing data for crop recognition. Based on the crop phenology in Yuanyang County, growing time is generally from mid-June to mid-October. The synchronized fall crops include paddy, peanut, and soybean, which tend to obstruct maize recognition.

2.2   Data Preprocess

The remote sensing data used in the current paper included MODIS and TM images covering the entirety of Yuanyang County. The Earth Observing System/MODIS data were composed of the NDVI images with a maximum synthesis of 16 d downloaded from the NASA website. The MODIS images were captured from May 24 to October 30, 2008, with a spatial resolution of 250 m. On the other hand, the TM image was captured on August 25, 2008, with a spatial resolution of 30 m. The data preprocess included geometric correction, atmospheric calibration, and projection transformation. The polynomial model of bilinear interpolation, which could ensure an error of less than 0.5 pixels, was used for the geometric correction. The 6S model was used for the atmospheric calibration. The projection type used for all spatial data used the Universal Transverse Mercatol system, whereas the geographic coordinate system was WGS 84.
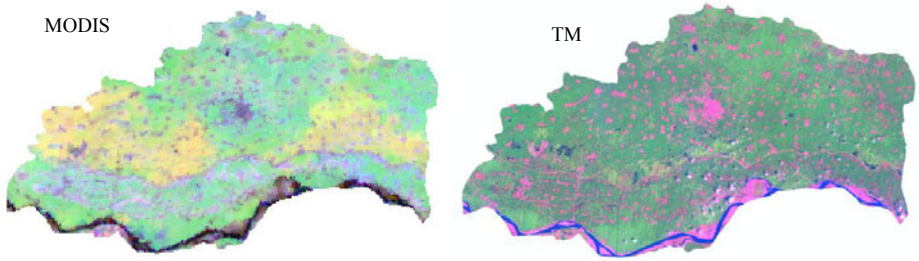
Fig. 1. MODIS and TM images covering Yuanyang County in 2008

Farmland parcel data, which were digitized from SPOT5 fusion images, provided the basic unit for the in-situ sample location and accuracy evaluation. The in-situ data included 49 sample parcels from August 2008, the attribute of which includes the parcel and planting area percentages for maize. The in-situ samples were primarily used to evaluate the accuracy of the method.

3   Method and Technical Process

The technical processes for the estimation of the maize planting area using integrated MODIS and TM images including wavelet transform fusion, reconstruction of time-series NDVI, classification by minimum distance model, and accuracy evaluation, are discussed in this section.

3.1   Fusion Model of Wavelet Transform

According to the theory and algorithm of wavelet transform, an image could be divided into one approximate component and three detail components. The approximate and detail components represent low- and high-frequency information, respectively. Using the inverse wavelet transform, high frequency information derived from the high-resolution image and low-frequency information derived from the multi-spectrum image could be transformed into the fusion image.

The algorithm of the wavelet transform model derived from the orthogonal basis of compact supported wavelets can be shown as follows:

$$\begin{cases} C_{j+1} = H_r H_c C_j \\ D_{j+1}^1 = H_r G_c C_j \\ D_{j+1}^2 = G_r H_c C_j \\ D_{j+1}^3 = G_r G_c C_j \end{cases} \tag{1}$$

where J is the decomposing order of the wavelet transform, C is the matrix of the original image, G is the high-pass filter, H is the low-pass filter, C_j is the low-frequency component, and D_j represents the three high-frequency components.

The key parameters of the wavelet transform fusion model include the wavelet basis, decomposition level, and fusion rule. The wavelet fusion is unable to provide a uniform theory and method for multi-source sensors because of the various remote sensor types and different spatial resolutions. Thus, performing numerous experiments was necessary to analyze the optimal wavelet basis, decomposition level, and fusion rule of the wavelet fusion between the MODIS and TM images.

After several experiments, the fusion of the wavelet transform between the MODIS and TM data was conducted using the BIOR6.8 wavelet basis and three decomposition layers and employing the fusion rule of high–low frequency conversion. Figures 2 and 3 show the final fusion image.
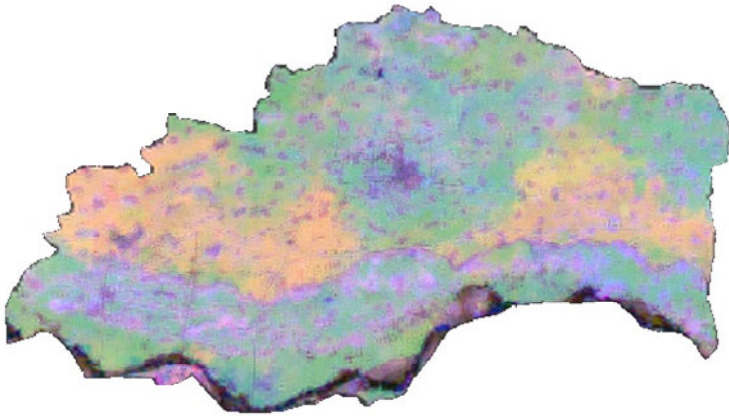
Fig. 2. Fusion image from the MODIS and TM images based on the wavelet transform
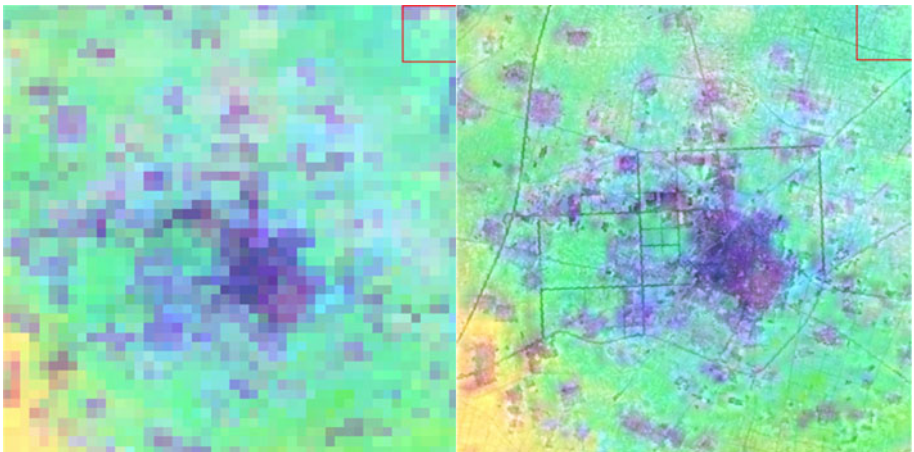
a) Original images of MODIS   b) Final fusion images

Fig. 3. Contrast between portions of the original MODIS images and the final fusion image

3.2   Reconstruction and Analysis of the Time-Series NDVI

The NDVI is the optimal indicator for the features, growth condition, and coverage of vegetation. In the current study, the standard growth curves of the primary fall crops were obtained using the time-series fusion image from the NDVI with a resolution of 30 m. First, using the located in-situ samples, pure parcel samples from the NDVI time-series fusion images, including maize, paddy, peanut, and soybean, were selected.
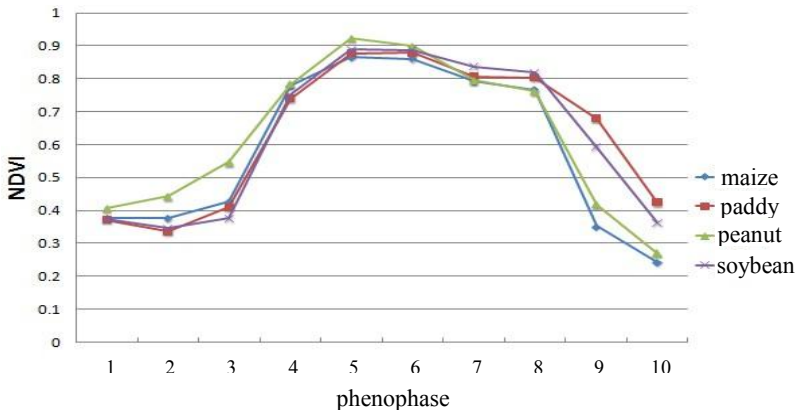
Fig. 4. NDVI time-series curves of the fall crops

As shown in Figure 4, the reflectance of the red band decreased when the coverage and leaf area index of the four crops increased, whereas the reflectance of the near infrared band increased along with the phenological changes. Thus, each of the four crops had peaks in the NDVI time-series curves.

The peak of maize in the NDVI curve appeared between the end of July and the first week of August, which corresponded to the grain-filling stage. The peak resulted from the end of the nutrition growth and the beginning of the complete reproductive growth of maize.

The features of the NDVI curves of the four crops were different to a certain extent. At the early-growth stages, the NDVI value of maize was difficult to distinguish because it was similar to that of paddy and soybean. However, at the late-growth stages, the NDVI values of maize and peanut rapidly decreased and showed significant differences with the NDVI values of paddy and soybean.

3.3   Minimum Distance Classifier

The selection of an effective distance measurement is important for the algorithm of the minimum distance classifier. The universal distance parameters include the Euclidean distance (ED) and the Mahalanobis distance, among others. In the current study, the typical ED was selected to determine the pixel classification using the minimum distance model. The function of Euclidean distance is as follows:

$$d_E = \sqrt{(NDVI_r - NDVI_t)^2}$$ (2)

3.4   Accuracy Evaluation

The in-situ sample parcels were assumed to have the real values, and the accuracy of the estimation of the maize planting area was evaluated based on the fusion of the MODIS and TM images with the indexes of position accuracy and gross accuracy.

The contributions of the different parcels varied because of the area differences of each farmland parcel. Taking the parcel area as weight, the function of the position accuracy is as follows:

$$K_i = \frac{\sum_{i=1}^{n}(1-|p_i - p_{io}|) * A_i}{\sum_{i=1}^{n} A_i} * 100\%$$ (3)

where P_i is the estimated percentage of maize in one parcel, P_{io} is the maize percentage of the in-situ sample parcel, A_i is the area of a certain parcel and n is the number of in-situ samples.

The evaluation of the gross accuracy was performed using the planting area account of all in-situ sample parcels as the true value. The function of the gross accuracy is as follows:

$$K_\omega = (1 - \frac{\left|\sum_{i=1}^{n} p_i * A_i - A_o\right|}{A_o}) * 100\%$$ (4)

where P_i is the estimated percentage of maize in one parcel, A_i is the area of a certain parcel, n is the number of in-situ samples of maize and A_o is the account of maize planting area of the in-situ sample parcels.

4   Results

The time-series curves of the four crops show that the NDVI features of several crops were similar to that of maize. At the seedling stage, the NDVI values of maize, paddy, and soybean were near that of one another, indicating that the growth conditions of the three crops were similar. At the jointing stage, the NDVI value of maize was close to that of the other crops, which were hardly distinguished. At the maturity stage, the NDVI value of maize was close to that of peanut when they were divided. However, the NDVI value of peanut was different with that of maize at the seedling stage, so the spatial distribution could be extracted using the stratified classification method.

Considering the differences in terms of phenomena and growing process among the studies crops, the spatial distribution of the NDVI values of peanut and maize at the seedling stage were extracted first. Then, the total spatial distributions of maize and peanut were distinguished based on their NDVI characteristics at the maturity stage. Finally, the joint of both spatial distributions was considered as that of maize, through which the account of the maize planting area in Yuanyang County could be estimated.

The maize planting area was also derived using a supervised classification of the in-situ samples. Considering the in-situ sample parcels as the true value, the accuracy of the supervised classification and time-series analyses were evaluated. The position and gross accuracies of maize derived using supervised classification were 62.89% and 49.01%, respectively. The estimated accuracies were derived from an improved time-series analysis from which the position and gross accuracies were 78.76% and 88.57%, respectively.

Table 1. Analysis of the accuracy of the maize planting area derived from the supervised classification and time-series analysis methods

method	position accuracy	gross accuracy
supervised classification	62.89%	49.01%
time-series analysis	78.76%	88.57%

5   Conclusion and Discussion

The fusion of MODIS and TM images based on wavelet transform enabled the plotting of the standard growth curves of the primary fall crops with a 30 m resolution, and the maize planting area was estimated using a minimum distance classifier. The results show that the estimation of the maize-sown area based on the time-series NDIV information of the fusion images could yield high values of gross and position accuracy (89% and 90%), indicating that this method could fully utilize the time-series information from the MODIS images and the spatial resolution of a TM image.

The present study requires further improvement. First, studying the estimation method for the maize planting area using daily time-series images is necessary because the MODIS NDVI data are synthesized from the maximum synthesis value of 16 d and because the phenomena observed from the fall crops are similar. Second, although the current study improved the spatial resolution of time-series data from 250 to 30 m, the spatial character of maize distribution, including the complex planting structure and fragmented planting parcels, result in mixed pixels. Therefore, studying soft classification or mixed pixel decomposition is necessary to reduce the effect of mixed pixels.

References

[1] Li, Y., Tan, K.: A preliminary study on the method for corn yield estimation by remote sensing in north China. Quarterly Journal of Applied Meteorology 6(supplement), 33–41 (1995)
[2] Chen, S., Liu, Q., Chen, L., et al.: Review of research advances in remote sensing monitoring of grain crop area. Transactions of the Chinese Society of Agricultural Engineering 21(6), 166–171 (2005)
[3] Gu, X., Pan, Y., Zhu, X., et al.: Consistency study between MODIS and TM on winter wheat plant area monitoring. Journal of Remote Sensing 11(3), 350–358 (2007)

[4] Van Niel, T.G., McVicar, T.R.: A simple method to improve field-level rice identification: toward operational monitoring with satellite remote sensing. Australian Journal of Experimental Agriculture 43, 379–387 (2003)

[5] Xiao, X., Stephen, B., et al.: Mapping paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images. Remote Sensing of Environment 100, 95–113 (2006)

[6] Gao, J.F., Pan, G.X., Jiang, X.S., et al.: Land-use induced changes in topsoil organic carbon stock of paddy fields using MODIS and TM/ETM analysis: A case study of Wujiang County, China. Journal of Environmental Sciences 20, 852–858 (2008)

[7] Wu, M., Wang, C., Niu, Z.: Mapping paddy fields in large areas, based on time series multi-sensors data. Transactions of the CSAE 26(7), 240–244 (2010)

[8] Zhang, J., Shen, K., Pan, Y., et al.: HJ-1 remotely sensed data and sampling method for wheat area estimation. Scientia Agricultura Sinica 43(16), 3306–3315 (2010)

[9] Ma, L., Gu, X., Xu, X., et al.: Remote sensing measurement of corn planting area based on field-data. Transactions of the CSAE 25(8), 147–151 (2009)

[10] Yang, X., Zhang, X., Jiang, D.: Extraction of multi-crop planting areas from MODIS data. Resources Science 26(6), 17–22 (2004)

[11] Xu, W., Zhang, G., Fan, J., et al.: Remote sensing monitoring of winter wheat areas using MODIS data. Transactions of the CSAE 23(12), 144–149 (2007)

[12] Feng, M., Yang, W., Zhang, D., et al.: Monitoring planting area and growth situation of irrigation- land and dry-land winter wheat based on TM and MODIS data. Transactions of the CSAE 25(3), 103–109 (2009)

[13] Gai, Y., Li, X., Li, X., et al.: Agricultural drought monitoring based on TM and MODIS remote sensing data—a case study of HebeiProvince. Journal of Nature Disasters 17(6), 91–95 (2008)

[14] Wang, H., Zhou, J., Wu, W.: Applications of wavelet analysis to remote sensing image processing. Remote Sensing Information 01, 93–99 (2009)

[15] Zhang, A., Li, N.: Research and analysis of multi-sources remotely sensed imagery fusion on the basis of wavelet transform technique. Modern Surveying and Mapping 32(5), 39–41 (2009)

[16] Han, L., Chen, Q., Han, T., et al.: Study on 3S-Based Estimation of Cotton-growing Area. Arid Zone Research 25(2), 207–211 (2008)

[17] Jing, R., Chun, L.: Improved minimum distance classifier — weighted minimum distance classifier. Computer Applications 25(5), 992–994 (2005)

[18] Gu, X., Pan, Y., He, X., et al.: Measurement of sown area of winter wheat based on per-field classification and remote sensing imagery. Journal of Remote Sensing 14(4), 789–805 (2010)

Discriminant Analysis of Red Wines from Different Aging Ways by Information Fusion of NIR and MIR Spectra

Sijia Tao[1,*], Jingming Li[2], Junhui Li[1], Jianbo Tang[1], Jinrui Mi[1], and Longlian Zhao[1]

[1] College of Information & Electrical Engineering, China Agricultural University,
Beijing, 10083, China
`taosijia530@163.com`
[2] College of Food Science & Nutrition Engineering, China Agricultural University,
Beijing, 10083, China

Abstract. Two different levels' information fusion of near infrared (NIR) spectra and mid infrared (MIR) spectra coming from red wines aged in different ways has been investigated. A total of 96 red wines, including 44 wines aged in oak barrel, 26 wines aged in stainless steel tank with oak chips and 26 wines aged in stainless steel tank, were analyzed. The NIR transmission spectra and MIR attenuated total reflectance (ATR) spectra of samples were collected. For data level fusion, the NIR and MIR spectra were max-min normalized respectively and then merged. For feature level fusion, five PCs and db3 coefficients of 3-layer wavelet decomposition were calculated. Discriminant models of the three different aging wines were established using DPLS and Fisher method before and after fusion. All the fusion methods mentioned above have achieved better results on classification than individual spectroscopy. These results suggest that the spectral fusion of NIR and MIR is a promising technology for discriminating different aging wines.

Keywords: Information fusion, NIR, MIR, Wine, Aging way.

1   Introduction

The red wine is welcomed by customers at home and abroad for its rich nutritional value. The wine aging process is the wine's mature process. It is a necessary procedure in the production of the aged wine. Different aging ways had great effects on the content of polysaccharides and wine chromaticity. So that it can affect the quality of the wine.

The research from home and abroad using spectroscopy studying wine are mainly focused on two aspects. On one hand, the application of mid infrared spectroscopy alone is widely used for classifying the varieties [1][2], vintages[3] and geographical region of red wine. On the other hand, near infrared spectroscopy is applied for quantitative analysis which is becoming a new rapid analysis technology that it is primarily applied for contents of samples as well as wine [1][4]. Although a large number of studies have been published, there are very few systematic investigations exploring classification of different aging ways of red wines to our knowledge.

---

[*] Corresponding author.

Information fusion is the merging of information from disparate sources, thus we can obtain more accurate and reliable description than single source achieved. The information of MIR comes from the frequency of the atomic vibration between the molecules while the information of NIR comes from the conjunction of vibration frequency and times frequency. The MIR has a lower limit of detection of $10^{-5} \sim 10^{-6}$ compared with NIR of $10^{-3} \sim 10^{-4}$. Based on the different characteristics of NIR and MIR, the information fusion of near infrared spectra and mid infrared spectra is supposed to get better classification results.

Here, two different levels' information fusion, data level fusion and feature level fusion, are studied, and the applicability of spectra fusion technique to discriminate different aging wines is investigated. The aim of this study is to realize the rapid and reliable discrimination of wines aged in three different ways.

2   Experimental

2.1   Wine Samples

A total of 96 red wines were analyzed in this study, including 44 wines aged in oak barrel, 26 wines aged in stainless steel tank with oak chips and 26 wines aged in stainless steel tank. The wine samples were from two kinds of grape varieties (Cabernet Sauvignon and Merlot) ranged from 2003 to 2007 vintages. The red wine samples were collected from wineries in Changli and Huailai of Hebei, China.

2.2   Instruments and Procedures

2.2.1   Spectra Collection
NIR Transmission Spectra: Wine samples were scanned in transmission mode using FT-NIR spectrometer of MPA (Bruker Optics), equipped with a tungsten lamp and a
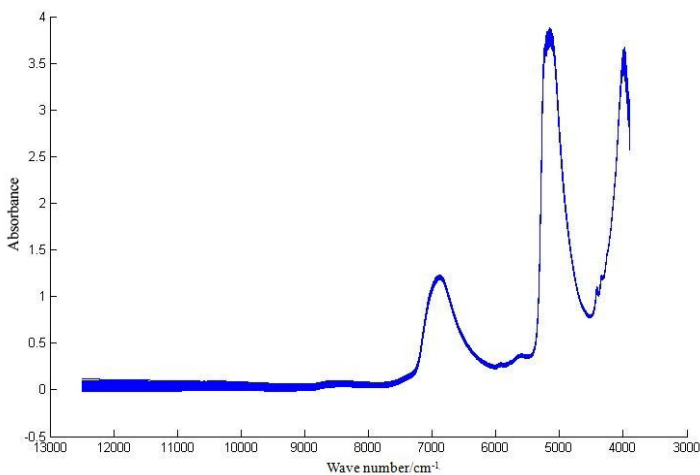
Fig. 1. Near infrared spectra of red wine samples

cooled PbS detector. Samples were scanned in a rectangular cuvette with a 1 mm path length. The spectra were measured by summing 32 scans acquired over the wavelength range 12500~3900cm^{-1} at a resolution of 4 cm^{-1}. The spectrum of each sample was obtained in triplicate and the mean of the three measurements was used for classification. NIR spectra are shown in Fig. 1.

MIR ATR Spectra: The MIR spectra of 96 wine samples were obtained from 550 to 4000 cm^{-1} on TENSOR 27 (Bruker Optics) FT-MIR instrument with an ATR attachment. The experiment used air as reference in room temperature. For both background and samples, the spectra were measured by summing 32 scans acquired at a resolution of 4 cm^{-1}. MIR spectra of the wine samples are shown in Fig. 2.
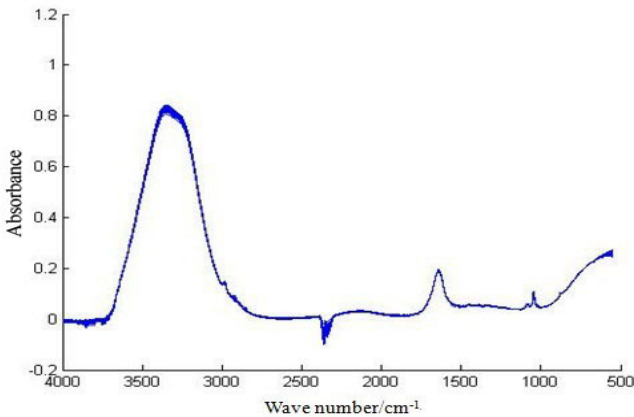
Fig. 2. Mid infrared spectra of red wine samples

2.2.2   Chemometric Techniques for Classification Analysis

The spectra before and after fusion were used to develop discrimination models for different wine aging ways using two pattern recognition methods, discriminant partial least squares (DPLS) and Fisher discriminant analysis (FDA). DPLS is an application of Partial Least Squares (PLS) [5] for discrimination analysis. Different from PLS, the output variable of DPLS was expressed as a binary variable, while the PLS was concentration value. Generally a critical threshold was pre-set, by comparing the predictive value with the threshold, the samples' categories were determined. FDA has been widely used in the domain of pattern recognition. The Fisher discriminant criterion was to find a linear projection that the classes of the samples are separated with maximal ratio of between-class and within-class variability [6].

3   Results and Discussion

3.1   Data Level Fusion

For data level fusion, the NIR and MIR spectra are preprocessed (min-max normalized) respectively. These two sources of information are merged then.

To develop the discriminant models, 96 wine samples were divided into two sets randomly, 2/3 training set and 1/3 for validation sets. Both sets were independent; the training set was used to build discriminant models and the validation set was only used for testing the models. Then DPLS and FDA modeling methods were used to develop classification models of different aging wines before and after fusion respectively. To test the stability of the models, wine samples were divided into training and validation set for 10 times randomly. Table 1 and table 2 show the mean recognition accuracy of cross-validation (training set) and prediction.

Table 1. Classification results for 3 different aged wine samples by DPLS

	Recognition accuracy（%）of training set	Recognition accuracy（%）of validation set
NIR	90.60	86.02
MIR	97.81	96.25
After fusion	99.06	96.59

Table 2. Classification results for 3 different aged wine samples by FDA

	Recognition accuracy（%）of training set	Recognition accuracy（%）of validation set
NIR	93.68	93.66
MIR	96.00	97.34
After fusion	98.01	97.98

From table 1 and table 2 we can see, both recognition accuracy of training set and recognition accuracy of validation set are improved. For DPLS model, the recognition accuracy of training set was improved to 99.06% while the recognition accuracy of validation set was advanced to 96.59%. For FDA model, the recognition accuracy of raining set and validation set were 98.01% and 97.98%, respectively.

3.2   Feature Level Fusion

Two kinds of feature extraction methods, PCA and wavelet transform, were used to achieve the goal of feature level fusion.

3.2.1   Principal Components Analysis (PCA)
PCA [7] was a widely used feature extraction method for reduction the number of variables showing co-linearity. In this study, five PCs were extracted from NIR spectra and MIR spectra respectively, then these PCs were merged as the input of the chemomtric methods in order to get the fused classification discrimination through modeling at the feature level. FDA method was applied to build the classification models of three different aging wines using the NIR spectra, MIR spectra and the merged spectra. Table 3 shows the classification result of the 96 red wine samples.

Table 3. Classification results for 3 different aged wine samples by FDA

	Recognition accuracy（%）of training set	Recognition accuracy（%）of validation set
NIR	90.63	87.50
MIR	94.38	91.87
After fusion	95.95	98.67

As it can be seen from table 3, the recognition accuracy of training set and validation set was 95.95% and 98.67 respectively. When compared with the results of NIR and MIR spectra, it can be easily found the accuracies were improved after spectra information fusion

3.2.2   Wavelet Transform

Wavelet transform [8] is now being adopted for a vast number of applications. Wavelets are mathematical functions that cut up data into different frequency components, and then study each component with a resolution matched to its scale. These coefficients can then be compressed more easily because the information is statistically concentrated in just a few coefficients.

The db3 wavelet was chose for 3-layer wavelet decomposition. The approximate coefficients of the 3rd layer were used to information fusion. Then DPLS and FDA modeling methods were used to develop classification models of different aging wines before and after fusion respectively.

we can obtain 99.52% as the recognition accuracy of training set and 95.96% as the recognition accuracy of validation set by taking an average of ten times as a result when use the DPLS model. While for the FDA method, the recognition accuracy of training set is 97.61% and the recognition accuracy of validation set is 98.65%. It can be found the accuracy was improved after information fusion.

4   Conclusions

In this paper, two different levels' information fusion of NIR and MIR spectra of three different aging wines were studied. For classifying three different aged wines, discriminant models are developed to analysis fused spectra with FDA and DPLS methods. The recognition accuracies were all increased to a certain extent after spectra fusion. For data level fusion, the recognition accuracies of validation set were improved to 96.59% and 97.98% by DPLS and FDA respectively. For feature level fusion, the recognition accuracies of validation set were improved to 98.67% using DPLS when PCA was used while the recognition accuracies of validation set were improved to 95.96% and 98.65% respectively by DPLS and FDA when we used wavelet transform. All the results above showed that information fusion of NIR and MIR spectra combined with the pattern recognition methods (DPLS and FDA) had the capability to extract feature of red wines and play a good role in discriminating wines from different aging systems.

However, some work still can be done for better results: some other pattern recognition methods could be tried, and the level of information fusion could be extended to the decision level [9]. It should be surely useful to extend the study in this direction.

References

1. Roussel, S., Bellon-Maurel, V., Roger, J.-M., Grenier, P.: Fusion of aroma, FT-IR and UV sensor data based on the Bayesian inference. Application to the discrimination of white grape varieties. Chemometrics and Intelligent Laboratory Systems 65(2), 209–219 (2003)
2. Picque, D., Cattenoz, T., Corrieu, G., et al.: Discrimination of red wines according to their geographical origin and vintage year by the use of mid-infrared spectroscopy. Sciences Des Aliments 25(3), 207 (2005)
3. Edelmann, A., Diewok, J., Schuster, K.C., Lendl, B.: Rapid method for the discrimination of red wine cultivars based on mid-infrared spectroscopy of phenolic wine extracts. Journal of Agricultural and Food Chemistry 49(3), 1138–1145 (2001)
4. Guillen, M.D., Cabo, N.: Infrared Spectroscopy in the study of edible oils and fats. Journal of the Science of Food and Agriculture 75, 1–11 (1997)
5. Rao, Xiang, Zhou, Wang, Xie, Xu (2009); Xie, Ying, Ying, Tian, Niu, Fu (2007)
6. Garcia-Allende, Conde, Mirapeix, Cobo, Lopez-Higuera (2008); Markiewicz, Matthews, Declerck, Herholz (2009); Zhu, Zang, Cao, Yan, He, Jiang, Sui, Wang (2008)
7. Vandeginsten, B.G.M., Massart, D.L., Buydens, S., De Jong, S., Lewi, P.J., Smeyers-Verbeke, J.: Handbook of Chemometrics and Quali- metrics: Part B. Elsevier, Amsterdam (1998)
8. Mallat, S., Hwang, W.L.: Singularity Detection and Processing with Wavelet. IEEE Trans. on Information Theory 38(2), 618–643 (1992)
9. Windeatt, T., Ghaderi, R.: Binary labeling and decision-level fusion. Information Fusion 2, 103–112 (2001)

Study on Automatic Composition of Semantic Geospatial Web Service

Dacheng Wang[1,2], Hongfeng Wu[3], XiaodongYang[1,*],
Wei Guo[1], and Weihong Cui[4]

[1] National Engineering Research Center for Information Technology in
Agriculture, Beijing, 100097, China
[2] Institute of Remote Sensing and GIS, Peking University, Beijing 100871, China
[3] Institute of Scientific and Technological Informatics, Heilongjiang Academy of Land
Reclamation Sciences, 150038, Harbin, China
[4] Institute of Remote Sensing Applications, Chinese Academy of Sciences,
Beijing, 100101, China

Abstract. We designed and implemented an integration framework of enterprise
business information and multi-source heterogeneous geospatial information
based on composition of semantic geospatial Web service. It can integrate
business information and geospatial information seamlessly and effectively in
enterprise workflow. An example of tobacco planting spatial analysis this
framework was described to verify the practicability and feasibility of this
framework.

Keywords: Semantic, Geographic information service, Automatic services
composition, Enterprise geographic information integration.

1 Introduction

Since1990s, geographic information system(GIS) has came into enterprise applying
stage gradually[1]. It has been applied widely in many fields such as land management,
intelligent transportation, business analysis, and played a more and more important
role. GIS promotes revolutionary changes in work patterns of these fields. With the
accumulation of spatial information, data becomes more and more, and data sources
become more and more complicated. Meanwhile, the association between GIS and
general business data becomes stronger. But traditional GIS runs outside enterprise
context data as an independent system which seriously limits the application of GIS in
enterprise. Especially, the unique spatial decision support functions of GIS are often
isolated in enterprise decision-making. In order to utilize the advantages of GIS in
spatial decision support, how to realize the sharing and interoperability of
heterogeneous geospatial information, and how to integrate them with enterprise
context data seamlessly have been the current focus in geographic information
research.

---

* Corresponding author.

The sharing and integration of geographic information mainly goes through four stages [2]: file management stage, spatial database management stage, metadata information exchange stage and the current integration and interoperability of spatial information on the semantic level. The semantic Web can realize the sharing and thoroughly interaction of the resource on the internet on the semantic and knowledge level, especially the exact expression, sharing and interoperability of Web information and knowledge on the semantic level. The spatial data have the following characteristics: multiple sources, complicated structure, structure, mass data, implicit expression and complicated process.

So it is only the semantic Web service that can achieve the knowledge sharing and interoperability of spatial data on the syntax and semantic level.

Many foreign universities, research institutions and large companies have set up special project teams to promote the development of semantic Web technology. Among these project teams, W3C, DAML,SDK WSMO, LSDIS lab in Georgia University and the European Union IST research projects are representative in the semantic Web research. In 2002, the semantic Web technology was listed in the National 863 Program as a key support project. The NSF has also supported many research programs about the geospatial services on the semantic level. Which mainly includes 'Digital Government: Geospatial Knowledge Ontology' research programs developed by Findler and Malyankar, 'Spatial thinking and inference' research program hosted by Reginald G. Golledge, and 'IGERT: Geographic Information Science Training Scheme' hosted by David M. Mark and Barry Smith. Werner Kuhn, a researcher in the institute of geographic information science in Munster University, Germany, led the semantic interoperability laboratory and enhanced the availability of geospatial information through the implementation of semantic interoperability. It provided a semantic reference body for geospatial information. In addition, the application of geographic body in the transportation, environment and emergency systems was also studied. The CSISS Center at George Mason University studied the knowledge discovery of heterogeneous remote sensing images based on intelligent Web services under the support of NGA. The state key laboratory of remote sensing information engineering in Wuhan University carried out spatial information grid research based on geographic ontology.

2   The Basis of Semantic Geospatial Services Technology

2.1   Web Services

Web services are software systems which are used to support the network interoperability of the software interaction, whose interface can be described by the format (usually is WSDL) being automatically processed by machine. The other systems use SOAP message mechanism to communicate with it to the specified network services. Usually, they use the HTTP protocol and the combination of the XML serialization objects and other network standards to communicate message.

Web services have the following characteristics [7]:

1) Good encapsulation. Web services are the objects which are deployed on the Web. So they have good encapsulation. The user can only need and see the function list provided for objects.

2) Loosely coupled. This feature stems from the object/component technology. As long as the call interfaces of Web services do not change, any changes in the implementation of Web services are all transparent. Even when the implementation platform of Web services migrates from J2EE to .NET or in the opposite direction, the user may know nothing about it.

3) Using standard protocol specifications. All public Web services protocols use the open standard protocols. In general, the majority of protocol norms will eventually use W3C or OASIS as the publisher and maintenance side of the final version.

4) Highly integrated capability. Web services use simple and intelligible standard protocols as the norms of the component interface description and collaboration description. It shields the differences of different software platforms. So CORBA, DCOM or EJB can use a standard protocol for interaction operation to achieve highly integrated in the current environment.

2.2   Semantic Web Services

Tim Berners-Lee presented the concept and the architecture of semantic Web in 2000. The extensible markup language (XML), resource description framework (RDF) and ontology are at the core of this system [8]. The ontology as semantic layer helps human and machine to communicate clearly and supports exchange on the semantic level, not just on the syntax level. This layer describes the relation among various resources, and the ontology reveals more complex and richer semantic information of resource itself as well as among resources. Thus, it can separate structure from content of information and describe information formally to make the information online have computer-readable semantics. So the ontology plays an important role in the semantic Web. Meanwhile, it is the basis of solving the sharing and exchange of Web information on the semantic level.

The semantic Web services provide machine-readable description by adding semantic annotations for Web services. It can make automatic discovery, composition, invocation and monitoring services better in open, complex and chaotic network environment [9]. The Web Ontology Language for Service (OWL-S) is recommended by the World Wide Web Consortium (W3C) to describe the inputs and outputs of Web services, predefined conditions and the ontology layer in function and processing model. The OWL-S is based on the ontology Web language (OWL) which is a W3C recommended standard and presented by the Web-Ontology work group. The OWL is a markup language which can distribute and share data on the Internet by the ontology. Through OWL, ontology library can be created for any field so the instantiated ontology description can be provided to sites in any field. In addition, the OWL supports logical inference so it can ensure the logical relation and logical order of the ontology expression by automated inference. The W3C does not give a specific definition of OWL-S, and just gives its main task to achieve [10]:

1) Automatic Web services discovery: The automatic Web service discovery can automatically position and provide the specified functions. Meanwhile, it can meet the constraints of the client. Through the services utilizing the OWL-S semantic markup, it can standard out the information required by Web service using computer-readable semantic markup at Web services site. Thus, the enhanced search engine of service registry center or the ontology can mark up automatic positioning services by these semantic markups. In addition, service providers can also utilize the OWL-S to pre-advertising its services at registry center (that is middle agent). The service requester can find these services when they query the registry center. Therefore, the OWL-S supports the service properties of automatic service discovery and statement advertising of functions.

2) Automatic Web services invoking: The automatic Web service call just needs a declarative description of service. The application program or agent can automatically call the service and does not need to bind a particular service in pre-programmed way. The execution of Web service is similar to a series of remote call procedures. The OWL-S markup of Web services provides a declarative computer-readable API. It includes parameters and the semantics of the information returned after execution which is required during executing call. Software agents should be able to explain the inputs and the information returned required by calling service. The OWL-S combines with the domain ontology written in OWL and provides declarative API to automatically execute Web services.

3) Automatic Web services composition and interoperability: The automatic Web services composition and interoperability is to implement the automatic selection, composition and interoperability of Web services involved in a complex task after given abstract description of a task. Through the semantic markup of Web services, the information required by the selection and composition services can be encoded at the Web services site. Software can automatically process the description of task and complete the task. In order to support these functions, the OWL-S not only provides the prerequisite required by each atomic services and declarative definition of result information, but also provides a kind of descriptive services combination and data flow interaction language.

More and more companies begin to transform some their business processes into Web services which makes the company's business relatively independent and more conducive to maintenance. At the same time, the needs of enterprises become more and more complex. A single atomic Web services often can not meet the complicated and individualized requirements of enterprises. So the requirements of the Web services composition emerge in response to. the situation requires. The Web services composition aims to let the seamless integration of cross-enterprise processes and the life cycle of transaction have same type, and to be used for many Web services.

At present, business process execution language for Web services (BPEL4WS), Web services modeling ontology, OWL-S, Petri nets and pacification [11] are the commonly used methods in the Web services composition. The BPEL4WS is a static services composition mode. It executes the pre-binding workflow file using workflow processing engine. The advantage of this mode is the stability of the services composition, and its disadvantage is the poor adaptability. The WSMO, OWL-S, Petri

nets and pacification all belong to the dynamic services composition mode. This mode dynamically searches and discovers services during the execution. Besides, it handles business processes using dynamic binding and composite services. It has good flexibility and adaptability. Among these methods, the OWL-S is the Web services composition standard recommended by the W3C. In the paper, this mode is also utilized.

2.3   Geospatial Services

At present, the geospatial services are the easiest way to access geographic data by using common network protocol. It is due to that geospatial services shield the differences between data formats and interfaces, and do not need to utilize traditional data distribution model [12]. The geospatial services have the potential to change the development, access and use of GIS, and make the sharing of geographic information data and functions easier [13]. Using the semantic Web technology, the geospatial services can be understood by other programs or software agents to achieve automatic composition of geographic services according to demand. It can provide more flexible and powerful functions. The ISO/TC211, OGC, INSPIRE and FGDC are all committed to research and develop the standards and norms of the geospatial services. The most important basis of the implementation of the semantic geospatial services is to establish a formal unified geographic ontology.

3   Integration Framework Design of Semantic Geospatial Services

The semantic geospatial services will have many advantages comparing with the existing geospatial services in many aspects. It will provide real-time and interactive geographic information services which also support spatial reasoning. The semantic geospatial services will support the seamless integration with general business management system. So the spatial information can be seamlessly managed as a part of a larger enterprise management system. This completely changes the island status of the geographic information in enterprise application.

The basis of achieving the semantic geospatial services integration is to establish the ontology library which meets the application requirements. According to the level of production, the ontology can be divided into three layers: Top-level ontology, domain ontology and application ontology. The top-level ontology usually describes the general concepts independent of any fields, such as event and object. The domain ontology is usually a concept in a specific field, such as the geographic information field and business logistics field. The application ontology is usually an application project oriented concept developed for a specific field and application.

Figure 1 shows the overall framework of the designed geospatial services integration in this paper. The basis of integration is to establish application ontology library

according to the application requirements of project. This ontology library consists of agriculture, tobacco, logistics, commercial and much other ontology. Figure 2 shows the tobacco-growing geographic information application ontology object. The geospatial services and enterprise services achieve the sharing and interoperability of information using the semantics provided by this ontology library.

In figure 1, the service manager is implemented by a Web service. It responses to the inputs of user and parses the user's request into a service list. This service list contains the name and parameters of input and output of the required service, and receives the result data returned from the service composition manger. The service composition manger is also a Web service and implemented by Jena + OWL-S. Its main function is to combine service list sent by service manger into concept-level service composition chain. The service chain repository is used to save the executed service chain information which includes the composition state of service chain and the reference state of result data generated by executing service chain. The function of inference service and matching service is to change concept-level service composition chain into executive instance-level service composition chain through the semantic integration service. The semantic integration service provides the unified entrance for accessing geospatial services and enterprise application services.

A typical user request and response process is as follows:

1) User submits a request to the service manager by client;

2) Service manger parses the request into service list;

3) Service manger sends the service list to service composition manger. Service composition manger generates a concept-level service composition chain according to the input parameters.

4) Service composition manager looks for the instance of this service chain from instance service chain library. If it is found, directly return data reference address to the service composition manager. Service composition manager returns the address to service manager. Service manager executes this reference and return the result data to user.

5) If the instance of service chain can not be found, service composition manager calls the matching service and inference service. Through semantic integration services, the two services call the corresponding geospatial services and enterprise services to generate the instance of this service chain according to according to semantic information provided by the ontology library. After the instance service chain executes, the instance is returned to the service composition manager and is stored to service chain library. The service composition manager returns the address reference of result data generated by the execution of the instance service chain to service manager.

Before the execution of each atomic service of the instance service chain generated by inference service and matching service, the instance service chain library is checked first. If the instance of this atomic service is found, the result address reference of the instance is directly called. Otherwise, execute the specific service process after finding this instance service.
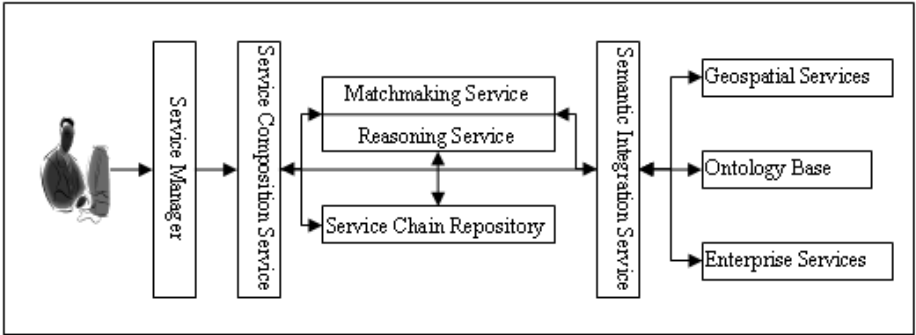
Fig. 1. Geospatial Service Integration Framework
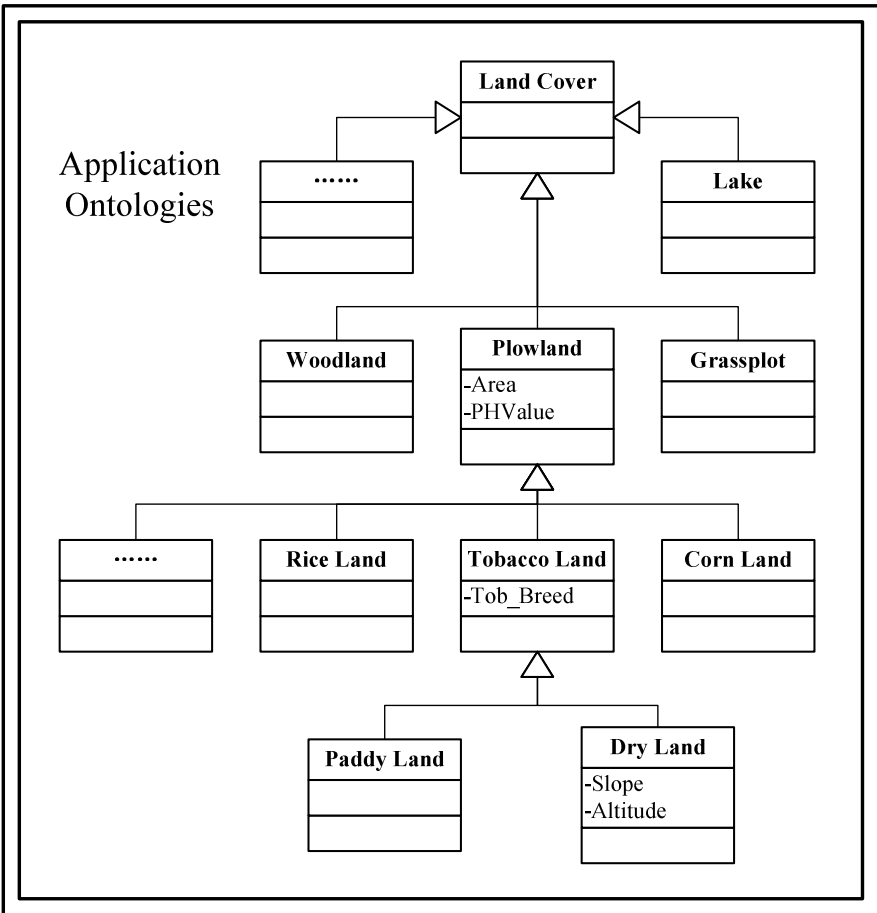
Fig. 2. Schema of Geospatial Ontologies in Tobacco Planting

4   An Example of Application

Since China has joined the international "Framework Convention on Tobacco Control" in 2005, the State Tobacco Monopoly Bureau has begun to implement the strict "dual control" policy gradually, so the rational tobacco farming planning has become very important to tobacco manufactures. In the last two years, tobacco manufacturers in every district have begun to implement basic tobacco farming planning. Most of the basic data have been collected manually. The spatial data have been stored in paper maps and the attribute data have been stored in ordinary relational databases.

Due to the subjectivity and arbitrariness of the manual planning, companies now need to determine whether tobacco farm is suitable for growing tobacco. That is the 'whether a land suits to grow tobacco' problem.

Table 1. Conditions for Tobacco Planting

Project	Description of the appropriate conditions
Altitude	1200-1900 meters
Slope	less than 15 degree
Soil type	Red soil, yellow soil, red and yellow soil
The type of land use	Land, land on a field whose slope is less than 15 degree
Soil PH	5.5-7
Sunlight	To large land, Sunlight duration is up to 500-700 hours; sunshine rate is greater than or equal to 40% ; Mining grilled period needs 280-400 hours; sunshine rate is greater than or equal to 30%。
Rainfall	Lager land requires that the monthly average rainfall is 100-130 mm
Temperature	The average daily temperature is less than 20 centigrade from July to September. The optimal temperature is 20-25 centigrade. The accumulated temperature is less than 10 centigrade in growth period

The temperature, humidity, sunshine, rainfall and other climatic conditions of tobacco growing have certain requirements. Meanwhile, tobacco growing needs appropriate soil type, land cover, slope, elevation and other geographic conditions. The climate data are often the statistical data of each administrative area issued by meteorological department. The geographic data often derives from the raster of vector map data issued by territory department and surveying and mapping department.

Whether a farm is suitable for the cultivation of tobacco depends not only on the elevation, slope, soil type, soil PH, water conservancy facilities and other geographic data, but also on sunshine, rainfall, temperature and other statistical data. Table 1 shows the details.

The elevation and slope data are generated from the DEM data. The soil type distribution maps, PH value of soil distribution maps and land use distribution maps are obtained from the relevant departments, respectively. The sunshine, rainfall and temperature statistical data of each administrative area can be obtained from the meteorology bureau. Because the study area in this paper does not provide online data services, the data obtained in this paper are the original data files. In actual, the corresponding data services are established in the internal network. Table 2 shows the service list.

Table 2. Services used in the Application

Content of service	Description of service
MapServ	Vector map service
FeatureServ	Feature service
ImageServ	Image map service
Overlay	Map overlay service
Buffer	Buffer service
SunlightComp	Sunlight statistical data service
RainComp	Rainfall statistics service
TemperatureComp	Temperature statistical data service

Assume that the user selects a piece of land from the client (usually a map page, also can be a data report) and then it is submitted to services manager for querying whether this land is suitable to grow tobacco. The process flow of the service request handled by the service manager is as follows: The existing service manager divides user request into atomic service list. The atomic service lists are combined into service list included two sub-service chain by the service composition manager. A sub-service chain is to determine whether the land conforms to the related spatial information conditions. Another sub-service chain is to determine whether the township where the land locates

conforms to the related statistical conditions. Figure 3 shows the details of service chain process. Finally, the return value of the whole service(ie, appropriate level) is calculated according to the results returned by the two sub-service chains. The appropriate level can be calculated according to the following equation.

$$\begin{cases} Q_a = \sum k_i * q_i \\ Q_b = \sum k_j * q_j \\ R = 100\% * (1 - Q_a - Q_b) \end{cases}$$

Where ki ,qi , kj ,qj represent the relevance degree and weights returned using the queried land as input parameters in a Web service, respectively; R is the appropriate level of growing tobacco in the queried land. Figure 4 shows the GUI of the system.
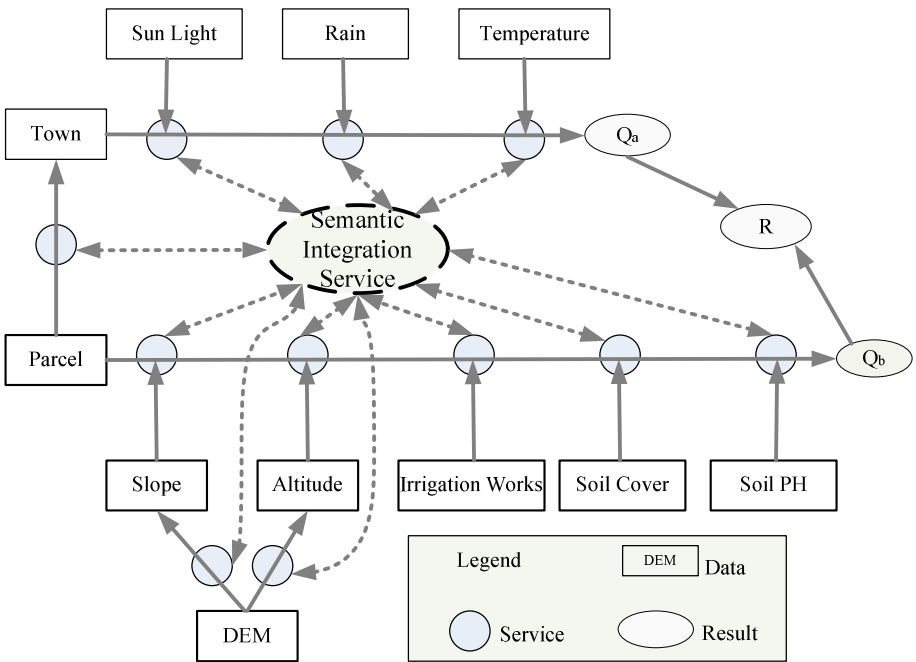
Fig. 3. Data Process Workflow of Business Instance
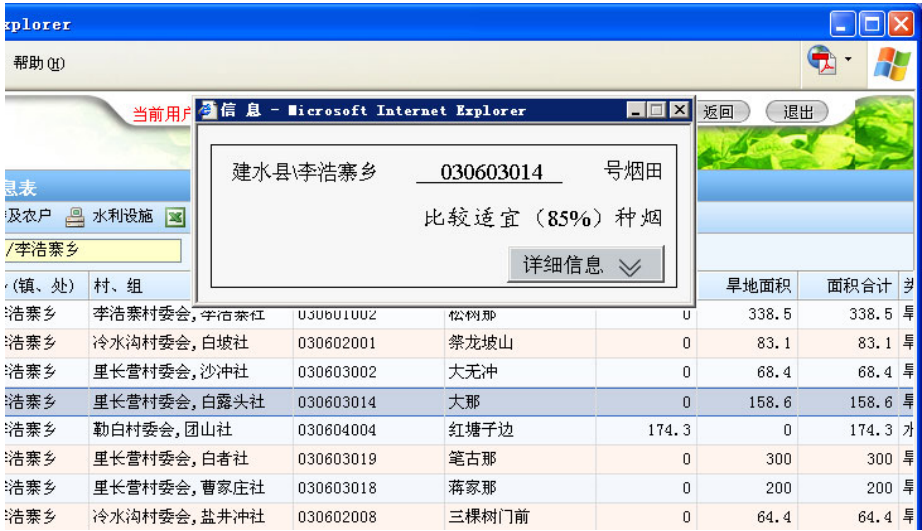
Fig. 3. A Snapshot of System UI

5   Conclusions

In this paper, the seamless integration and interoperability of the geospatial services in tobacco-growing enterprise application system is implemented under the framework of the geospatial services theory and technology. It promotes the application of the geospatial services in enterprise decision-making and meets the requirements of the enterprise geographic information application well.

The semantic geospatial services have wide prospect in the enterprise application. It is the development trend of the geospatial services enterprise integration. But it still has few mature applications at present. The further research and practice in this aspect will be conducted.

Reference

[1] Song, G., Zhong, E., Liu, J., et al.: A Study on Seamless Integration of Multi-sources Spatial-data (SIMS). Progress in Geography (02) (2000)
[2] Wu, M.: Study on Ontology-driven Multi-Source Heterogeneous Spatio-temporal Data of Agricultural Geographical Information Taxonomy and Query. IRSA,CAS, 5 (2007)

[3] Guide for Hi-tech Research and Development Program of China - Semantic Web Technology and Ontology [EB/OL] (May 8, 2002),
http://www.863.org.cn/863_105/applygide/applygide2/informati
on2_app/200205080043.html

[4] Geospatial Knowledge Discovery via Intelligent Web Services[EB /OL] (October 10, 2007), http://www.laits.gmu.edu/geo/nga/index.html

[5] Li, D., Cui, W.: Geographic Ontology and SIMD. Acta Geodaetica Et Cartographica Sinica (02) (2006)

[6] Web Services Architecture [EB/OL] (February 11, 2004),
http://www.w3.org/TR/ws-arch/

[7] James, S.: Implementing Web Services with the WSTK [EB/OL]. IBM developerWorks (May 30, 2001)

[8] Yang, A.: Research on Geo Web Services Based on Ontology. Wuhan University (2005)

[9] Henriksson, R.: Semantic Web Service Composition [EB/OL] (November 26, 2006),
http://www.cs.helsinki.fi/u/thruokol/opetus/2006/sose/papers
/henriksson_sws-composition.pdf

[10] OWL-S: Semantic Markup for Web Services [EB /OL] (November 22, 2004),
http://www.w3.org/Submission/OWL-S/

[11] Charif, Y., Sabouret, N.: An Overview of Semantic Web Services Composition Approaches. Electronic Notes in Theoretical Computer Science 85(6) (2005)

[12] Breitman, K.K., Casanova, M.A., Truszkowski, W.: Semantic Web: Concepts, Technologies and Applications. Springer, London (2007)

[13] Tang, W., Selwood, J.: Connecting Our World: GIS Web Services. ESRI Press (2003)

[14] Rana, S., Sharma, J.: Frontiers of Geographic Information Technology. Springer, Berlin (2006)

[15] Yue, P., Di, L., Yang, W., et al.: Semantics-based automatic composition of geospatial Web service chains. Computers & Geosciences 33, 649–665 (2007)

The Estimation of Winter Wheat Yield Based on MODIS Remote Sensing Data

Linsheng Huang[1,2], Qinying Yang[1,2], Dong Liang[1], Yansheng Dong[2], Xingang Xu[2], and Wenjiang Huang[2,*]

[1] Key Laboratory of Intelligent Computing & Signal Processing,
Ministry of Education, Anhui University, Hefei, 230039, P.R. China
[2] Beijing Research Center for Information Technology in Agriculture,
Beijing, 100097, P.R. China
`linsheng0808@163.com, huangwj@nercita.org.cn`

Abstract. A yield estimation method by remote sensing was used to estimate the yield of winter wheat in Jiangsu province, China. The first step of this study was to extract the planting area of winter wheat from environmental satellite images and land -use map of Jiangsu province, meanwhile, correlation analyses were performed by using 8-day of composite Leaf Area Index (LAI) data from Moderate Resolution Imaging Spectroradiometer (MODIS) and statistical yield of corresponding counties. Secondly, the average LAI was calculated at the optimal growth period, and the statistical yields of wheat for all counties were collected, in which the former was chosen as the independent variable and the latter was the dependent variable, and the regression model was established. Finally, the accuracy and stability of the regression model were validated using the data of another year. The results indicated that the yield estimation model at provincial level was reliable, the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE) of the model was 12.1% and 9.7%, respectively. In addition, the yield estimation system of winter wheat in Jiangsu province was constructed and published based on ArcMap and ArcGIS Server.

Keywords: LAI, remote sensing, yield estimation, regression model, ArcGIS Server.

0 Introduction

With the increase of population and the reduction of arable land, food security has become a very serious and prominent issue around the world, especially in China. Winter wheat is one of the major crops in China, its planting area accounts for one-fifth of the total crops area nationwide. Timely and accurate estimation of the winter wheat yield has an important significance for a country to enhance the risk management of food security as well as to develop appropriate market policies[1-3].

---

* Corresponding author.

Currently, a range of methods have been used for estimating crop yield[4-8]. These methods, however, have their own drawbacks respectively. The statistical surveys methods are time-consuming and require a lot of manpower and material resources; agricultural weather forecasting methods and agricultural forecasting methods are not suitable for large-scale crop yield estimation.

As satellite technologies improved significantly in recent years, remote sensing has become an important method to monitor and gather agricultural information in practice. Compared with these methods mentioned above, remote sensing yield estimation methods, by means of particular advantages such as large coverage, short monitoring cycle, strong instantaneity as well as low cost, have provided new technological means for the realization of large-scale crop yield estimation. Remote sensing yield estimation methods include crop growth model method, semi-empirical and semi-mechanistic model method and empirical model method. They have their own advantages and disadvantages, crop growth model method requires a lot of input parameters (such as meteorological parameters, management parameters and crop genetic parameters, etc.), which limits the use in large regions because some of the parameters are not easy to obtain in these areas. Also, it is difficult to simulate some parameters in semi-empirical and semi-mechanistic model method. Empirical model method though has some limitations, is a simple and suitable method for large-scale crop yield estimation under certain accuracy conditions. By using this method, many scholars at home and abroad had made outstanding achievements in crop yield estimation[9-10]. Ren et al. (2010) estimated regional winter wheat yield based on the inversion of LAI, the yield of winter wheat could be accurately forecasted 20-30 days prior to harvest, the results indicated that LAI and crop yield have a good correlation[9]. Becker-Reshef et al. (2010) utilized the maximum of Normalized Difference Vegetation Index (NDVI) in crop growing season to estimate yield, the error was less than 10%[10]. Mkhabela et al. (2011) developed a production-NDVI model to estimate crop yield, in which the NDVI values were selected at the optimal estimation period (the period when correlation coefficient of NDVI values and the yields is largest) [11]. Based on previous studies, this study estimated the winter wheat yield with LAI at the optimal estimation period in Jiangsu province.

1   Study Area and Data Processing

1.1   Study Areas

Jiangsu province is located in eastern coast of China mainland which is the transition zone of subtropical to tropical region, between longitude 116°18'-121°57 ' and north latitude 30°45' -35°20',the geographic range shown in Figure 1(a). Its average temperature is 13°C -16°C and annual rainfall is about 1000 ml, the area of arable land is 4,902 million hectares, accounting for 3.97% of nationwide. It is one of the most important food sources in China, growing areas of winter wheat shown in Figure 1(b).

(a)   (b)
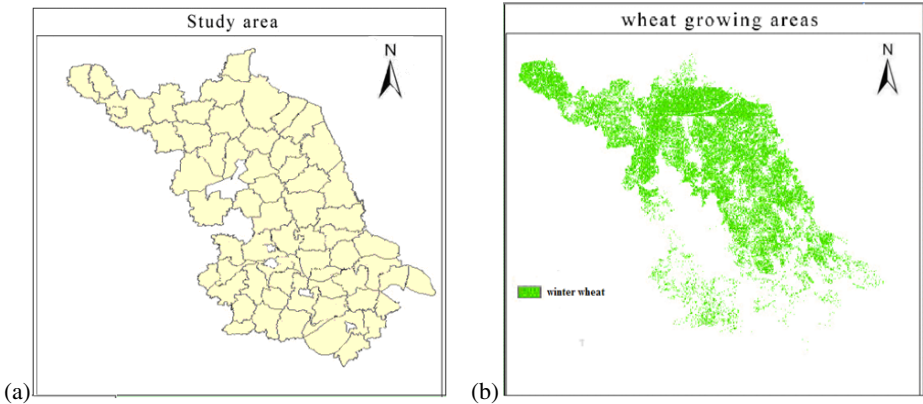
Fig. 1. The study area *(a)* and growing areas of winter wheat *(b)* selected in this study

Winter wheat in Jiangsu Province was sowed at the period of early to mid October, matured at about late May to early June, the phenophase of winter wheat is shown in Table 1.

Table 1. Phenophase of winter wheat in Jiangsu province

Month		Feb		Mar		Apr		May	Jun
Date	early	mid-late	early-mid	late	early	mid-late	early-mid	late	early
Stage	wintering	turning green	erecting	jointing	booting	heading	flowering	maturity	

1.2   Data Processing

This study utilized the data of environmental satellite and the land use map of 2009 to extract the information of winter wheat growing areas and got the average LAI of county scale from MODIS products. By using the remote sensing images of environmental satellite, the information of winter wheat growing areas was extracted in 16 April and 18 April, 2009. The process of preprocessing included radiometric calibration, atmospheric correction as well as geometric correction. The first step was using the land use map of Jiangsu province to mask remote sensing images which had been preprocessed by which environmental satellite images within the scope of cultivated land obtained, and then through the process of supervised classification for satellite images (which had been masked) to extract the winter wheat growing areas. The principle of winter wheat extraction is: there were almost no other crops grown in Jiangsu province in April, therefore the green vegetation in arable land can be regarded as winter wheat. The next step was stitching the winter wheat growing areas of the images, which captured in April 16 and April 18, 2009 to form an integrated image of Jiangsu Province. In order to match the pixels of MODIS-LAI data, this study resampled the spatial resolution of 30m to that of 1km in these areas. Also, it

assumed that the wheat growing areas varied slightly at the county scale on the condition that the spatial resolution was 1km. Meanwhile, the wheat growing areas of 2009 could be used as that of 2006 and 2007, as shown in Figure 1(b). LAI data used in this study came from MOD15A2 which is a product of MODIS. The advantages of MODIS data as below: short repetition period, large coverage area and strong access ability. After the process of smoothing and denoising, MOD15A2 data was re-projected with MODIS Reprojection Tool (MRT), and the projection types were geographic latitude-longitude projection and WGS84 coordinates system. In order to get data of wheat growing areas in Jiangsu province, h27v05 and h28v05 (track number) were stitched in ENVI software in this study. Next, the LAI data in wheat growing areas would be reached by utilizing the wheat areas to mask the stitched images. A total of 64 scenes of MOD15A2 data were adopted in this study, the time range covered the seven critical growth stages of winter wheat (turning green, erecting, jointing , booting, heading, flowering and milky mature) from 2006 to 2007.

2   Estimation Methods and Results Analysis

2.1   Estimation Methods

As different crops have different spectral information, firstly satellite sensors are used to record information of the surface and identify the type of crops, then extract various vegetation indices to establish the relationship between vegetation index and crop yield, and finally use vegetation indices for estimation, this method is called yield estimation [12]. LAI is defined as the ratio the projected area of the leaves to that of the total area when the sunlight is direct. Crop LAI is an important parameter in the process of material exchange with the environment, and it directly relate to photosynthesis, respiration, transpiration and biomass accumulation of crop and has significant impact on the yield of crop [13-14]. LAI used in this study was the average number in the county scale, which was got through the statistical computation of MOD15A2 product in wheat growing areas. Also, the LAI of winter wheat growing areas in each county and the number of pixels corresponding to these areas were counted in ArcMap. Total number of LAI in winter wheat growing areas divided by the number of pixels will get the average LAI of the winter wheat in this county. Due to the correlation between LAI at different periods and crop production is different, firstly through the correlation analysis of average LAI at county-scale and statistical yield in a county in 2006, then selected the period which correlation coefficient was largest as the optimal estimation period. Furthermore, this study utilized the average LAI of this period and statistical yield for linear regression to obtain a provincial-scale estimation model. Estimated yield of 2007 would be obtained when substituted the average LAI of each county at the optimal estimation period in this year into estimation model. Finally, accuracy of the model would be got through the comparison and validation of estimated yield with statistical yield.

RMSE and MAE are the two indicators for validate the accuracy of yield estimation model. RMSE and MAE are defined as follow:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - M_i)^2} \qquad (1)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}| P_i - M_i | \qquad (2)$$

where n is the number of counties for estimation in Jiangsu Province, Pi is the estimated yield and Mi is the statistical data. The percentage values of RMSE and MAE can be obtained when the values of them divided by the average value of statistical yield. RMSE can well reflect the accuracy of predicted values, when RMSE is less than 10% indicates that the consistency of predicted and actual values is very good, when the number is less than 10% and greater than 20% shows the predicted results are good, between 20% and 30% means predicted results are general, greater than 30% indicates large deviations[11].

2.2   Result Analysis

At the period from turning green to maturity stage of winter wheat in 2006, through the correlation analysis on the average LAI every 8 days and the statistical yields in each county, the results shown in Figure 2(a). The correlation coefficient of winter wheat LAI and yield was maximum in the first 97 to 121 days of the year, up to 0.6. It also showed that the booting to heading stage was the optimal yield estimation period for winter wheat in Jiangsu province because this period was very important in the process of reproductive growth and yield formation.
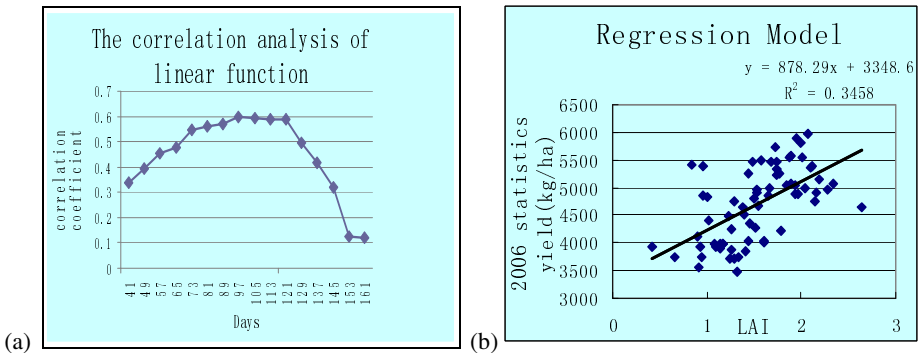
Fig. 2. The correlation analysis *(a)* and relationship between wheat yield and LAI *(b)*

In this study, through the linear regression of the LAI at heading stage and statistical yield of winter wheat in each county in 2006, the yield estimation model will be reached, as shown in Figure 2(b). The regression equation is:

$$y = 878.29x + 3348.6 \qquad (3)$$

where *y* is the yield of winter wheat, *x* is the average LAI at heading stage. As can be illustrated from Figure 2(b) that LAI and yield are correlated remarkably.

When substituted the average LAI at heading stage in 2007 into regression model, the yield estimation of each county and the spatial distribution of winter wheat yield in 2007 will be reached, as shown in Figure 3(a). In order to validate the results, this study compared the estimated yield with statistical yield in Figure 3(b). RMSE and MAE of wheat yield estimation was 12.1% and 9.7% respectively, it showed that the accuracy of yield estimation model was satisfactory.

(a)   (b)
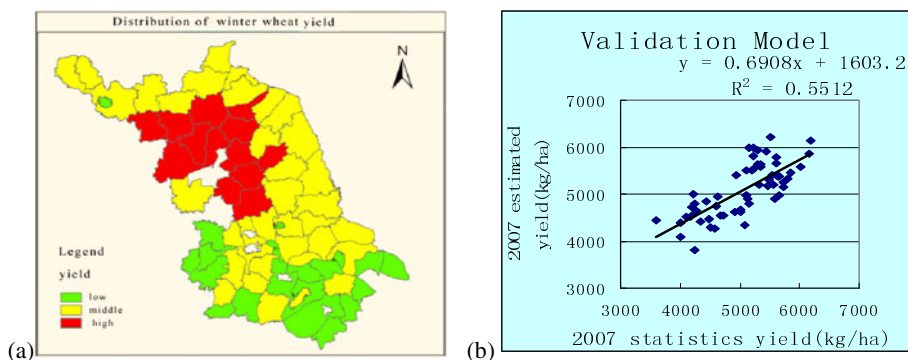
Fig. 3. Distribution of winter wheat yield *(a)* and accuracy validation of yield estimation model in 2007 *(b)*

3   Implementation of Web Publishing System

Web services make it possible for service providers to sell their services by publishing their availability over the Internet. They provide a standard means of interoperating between different software applications, running on a variety of platforms or frameworks. ArcGIS Server is developed by ESRI based on ArcObjects (AO) components; the aim of it is to build enterprise-class Geographic Information system (GIS). AO components provide a wealth of features, including access to GIS data, maps performance and spatial analysis, etc[15]. Map Service function of ArcGIS Servers are mainly used in this study. Before calling the estimation data, the data must first be published as the configuration information of map layer. The map layer is designed in ArcMap, the basic functions of it including zoom, layer stack as well as inserting data to the layer, the output map is saved as mxd format. A service object will be created in the interface of ArcGIS Server Manage by double-clicking the Add New Service, entering the name of the service object and the folder where the published map located, and the published map is a active layer[16].

This study uses the JavaScript language and calls the components of ArcGIS through API to build the published system. The system is developed in the .net framework of Microsoft Visual Studio 2008, where js file (a script file of JavaScript) can be directly executed. The function modules of the system including basic map functions, layer management as well as search functions. The function menus and the operation interface of the system showed in Table 2 and Figure 4, respectively.

Table 2. function menus of the system

Function modules	Specific functions	Function description
basic map functions	zoom, roam, view	facilitate the access of map
layer management	add &delete layer	highlight layer information
search functions	search for yield data	in favor of data sharing
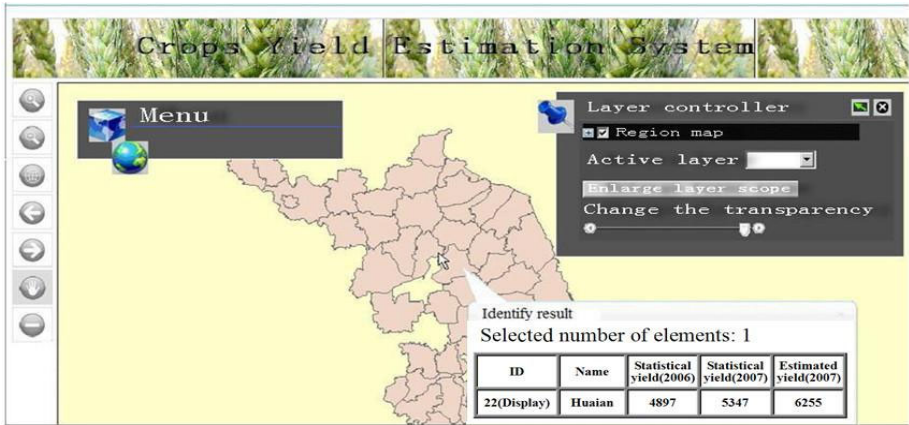
Fig. 4. Operation interface of yield estimation system

4   Discussion and Conclusions

Based on the data of MODIS products, this study established the model of winter wheat yield estimation by selecting the average LAI at the optimal estimation period and statistical yield data (2006-2007) of each county in Jiangsu province. The accuracy of the model was quite satisfactory through the validation of detailed data. Winter wheat yield can be estimated one to two months prior to harvest, which provide the scientific basis for the local government to make appropriate food policies. In addition, for the purpose of information sharing, this study designed a Web publishing system to publish the data of wheat yield.

Despite these good results, there are still some inadequacies in this study:

1, the current model was developed based on the 2 years of data available and the accuracy of the model was validated only by 1 year of data, the results were lack of universality because of the small amount of data;

2, the constructed environment of the model was relatively simple in which the impact of natural disasters and other factors on yield were not carefully considered;

3, the functions of the Web publishing system were not perfect and needed to further improvement.

Consequently, further studies are needed to resolve these drawbacks mentioned above.

References

1. Li, M.Z., Li, P., Wang, Q., Fang, J.Q., Wang, M.H.: Development of an Intelligent Yield Monitor for Grain Combine Harvester. In: Artificial Intelligence Applications and Innovations. IFIP, pp. 663–670 (2005)
2. Huang, W.J., Wang, J.H., Wang, Z.J., Zhao, C.J., Liu, L.Y., Wang, J.D.: Inversion of Foliar Biochemical Parameters at Various Physiological Stages and Grain Quality Indicators of Winter Wheat With Canopy Reflectance. Int. J. Remote Sens. 25, 2409–2419 (2004)
3. Xu, X.G.: Research of Crop Yield Models. Chinese Academy of Sciences, Beijing (2007)
4. Liu, L.Y., Wang, J.H., Huang, W.J., Zhao, C.J., Zhang, B., Tong, Q.X.: Improving Winter Wheat Yield Prediction by Novel Spectral Index. Trans. Chin. Soc. Agric. Eng. 20, 172–175 (2004)
5. Li, R., Li, C.J., Xu, X.G., Wang, J.H., Yang, X.D., Huang, W.J., Pan, Y.C.: Active Solar Heating System With Soil Heat Storage for Plastic Film Greenhouse and Its Effects. Trans. Chin. Soc. Agric. Eng. 25, 114–117 (2009)
6. Li, W.G., Wang, J.H., Zhao, C.J., Liu, L.Y.: A Model of Estimating Winter Wheat Yield Based on TM Image and Yield Formation. J. Triticeae Crops. 27, 904–907 (2007)
7. Tang, Y.L., Wang, J.H., Huang, J.F., Wang, R.C.: Yield Estimation by Hyperserctral Data of Rice Canopies in Mature Stages. Acta Agron. Sin. 30, 780–785 (2004)
8. Wang, T., Zhang, L.D., Lao, C.L., Zhao, L.L., Li, J.H., Wang, Z.Y., Huang, W.J., Wang, J.H., Yan, Y.L.: Study on NPLS Model with Three-Dimensional Hyperspectrum for Assessing Per-Mu-Yield of Winter Wheat. Spectrosc. Spectral Anal. 26, 1915–1917 (2006)
9. Ren, J.Q., Chen, Z.X., Zhou, Q.B., Tang, H.J.: LAI-based Regional Winter Wheat Yield Estimation by Remote Sensing. Chin. J. Appl. Ecol. 21, 2883–2888 (2010)
10. Becker-Reshef, I., Vermote, E., Lindeman, M., Justice, C.: A Generalized Regression-Based Model for Forecasting Winter Wheat Yields in Kansas and Ukraine Using Modis Data. Remote Sens. Environ. 114, 1312–1323 (2010)
11. MKhabela, M.S., Bullock, P., Raj, S., Wang, S., Yang, Y.: Crop Yield Forecasting on the Canadian Prairies Using MODIS NDVI Data. Agr. Forest Meteorol. 151, 385–393 (2011)
12. Deng, R., Huang, J.F., Wang, F.M., Sun, H.S., Peng, D.L.: Research on Yield Estimation of Rice Based on Remote Sensing Using Moderate-Resolution Imaging Spectroradiometer(MODIS) Data:A Case Study of Jiangsu Province, China. Chin. J. Rice Sci. 24, 87–92 (2010)
13. Chen, J.S., Huang, J.X., Lin, H., Pei, Z.Y.: Rice Yield Estimation by Assimilation Remote Sensing Into Crop Growth Model. Sci. China Ser. F. 40(suppl.), 173–183 (2010)
14. Bai, L., Wang, J., Jiang, G.Y.: Study State and Developing Trends of Hyperspectral Remote Sensing Technology in Estimating Crop Yield. Mod. Agr. 1, 30–33 (2006)
15. Zhao, P., Liu, G., Zhang, X.J., Li, M.Z.: A Method to Develope WebGIS Based on ArcIMS and Application. App. Res. Comput. 22, 258–260 (2005)
16. Yang, F.P., Li, L., Feng, J.F., Wu, Z.T.: Research on integration method of GIS-based application model based on web service composition. Computer Engineering and Des. 1, 133–137 (2011)

Prediction of Wheat Stripe Rust
Based on Neural Networks

Haiguang Wang[*] and Zhanhong Ma

Department of Plant Pathology, China Agricultural University, Beijing 100193, China
wanghaiguang@cau.edu.cn

Abstract. Stripe rust caused by *Puccinia striiformis* f. sp. *tritici*, is a devastating wheat disease in the world. The prediction of this disease is very important to make control strategies. In order to figure out suitable prediction methods based on neural networks that could provide accurate prediction information with high stability, the predictions of wheat stripe rust by using backpropagation networks with different transfer functions, training functions and learning functions, radial basis networks, generalized regression networks (GRNNs) and probabilistic neural networks (PNNs) were conducted in this study. The results indicated that suitable backpropagation networks, radial basis networks and GRNNs could be used for the prediction of wheat stripe rust. Good fitting accuracy and prediction accuracy could be obtained by using backpropagation networks with trainlm, trainrp or trainbfg as training function. Radial basis networks had more power than backpropagation networks and GRNNs to predict wheat stripe rust. GRNNs were easier to be used than backpropagation networks. New methods based on neural networks were provided for the prediction of wheat stripe rust.

Keywords: Wheat stripe rust, prediction, neural networks, epidemics.

1 Introduction

Wheat stripe rust caused by *Puccinia striiformis* f. sp. *tritici*, is a very important wheat disease in the world [1]. This disease is widely distributed in eastern and central Asia, western Europe and the pacific coast of North America [1], [2], [3]. The epidemics of this disease could also occur in Oceania, northern and eastern Africa, and South America [1]. In China, in terms of distribution range and destructivity, stripe rust is the most important among the three wheat rusts including leaf rust and stem rust. Stripe rust could occur in all main wheat areas and could cause severe losses in wheat production in China. It is critical to the national food security of China. The epidemics of wheat stripe rust are related to the long distance transports of the pathogen driven by winds. Accurate prediction of this disease is the key to make strategies for effective control.

The prediction of plant disease is to estimate the prevalence of plant disease after a certain time using expert experience or system simulation methods based on epidemic

---

[*] Corresponding author.

law of the disease. There have been many reports about the prediction of wheat stripe rust. The methods that were usually used included regression analysis [4], [5], [6], [7], discrimination analysis [8], [9], [10], Markov forecast method [11], principal component analysis [12], grey model forecast method [13], [14] and neural networks [15], [16], [17]. Wheat stripe rust is influenced by various factors, including the amount of pathogen, variety resistance and meteorological conditions. The relationship between wheat stripe rust and these factors is usually nonlinear. Neural networks have good advantages of solving the nonlinear problems. They are composed of interconnecting artificial neurons. They have good adaptive learning ability and nonlinear mapping ability. Neural networks have been successfully applied in various fields, including pattern recognition, image analysis and control systems. Although neural networks have been used for the prediction of wheat stripe rust [15], [16], [17], individual backpropagation network or individual function combination was be used in most cases. To the authors' knowledge, there is not any report about the application of radial basis networks to the prediction of wheat stripe rust. Furthermore, the comparison between prediction performances of different neural networks and the comparison between prediction performances of backpropagation networks combined with different transfer functions, training functions and learning functions have not been reported yet.

In order to select prediction methods based on neural networks with good prediction performance and prediction accuracy and provide disease information for making control strategies and taking control measures, the predictions of wheat stripe rust based on backpropagation networks with different transfer functions, training functions and learning functions, radial basis networks, and two variants of radial basis networks, generalized regression networks (GRNNs) and probabilistic neural networks (PNNs), were conducted and their prediction effects were compared in this study.

2 Materials and Methods

2.1 Materials

The data used for analysis in this study were colleted from [18]. They consisted of the data of disease prevalence of wheat stripe rust and related meteorological data in Hanzhong in Shannxi Province during the period from 1974 to 1997. In detail, they included the amount of stripe rust pathogen in autumn, the amount of stripe rust pathogen in spring, average temperature in January, average temperature in February, average temperature in March, average temperature in April, average temperature in November, average temperature in December, precipitation in January, precipitation in February, precipitation in March, precipitation in April, precipitation in November, precipitation in December, the proportion of susceptible variety area and disease prevalence in adult stage of wheat. Among these factors, the amount of stripe rust pathogen in autumn and related data in November and December referred to the data in the preceding year. The fifteen influencing factors were normalized. The data set contained the fifteen normalized factors and disease prevalence of wheat stripe rust

was regarded as Data set 1. In the data set, disease prevalence of wheat stripe rust was divided into five classes represented by 1, 2, 3, 4 and 5, respectively. The data set was regarded as Data set 2 while Class 1, 2, 3, 4 and 5 were expressed as (0 0 0 0 1), (0 0 0 1 0), (0 0 1 0 0), (0 1 0 0 0) and (1 0 0 0 0), respectively.

In the literature [18], the amount of stripe rust pathogen in spring, the amount of stripe rust pathogen in autumn, the proportion of susceptible variety area, precipitation in April and average temperature in April were selected as the main factors influencing disease prevalence from the fifteen factors by using stepwise regression analysis. The five selected factors were normalized. The data set contained the five normalized factors and disease prevalence of wheat stripe rust represented by 1, 2, 3, 4 or 5, was regarded as Data set 3. The data set was regarded as Data set 4 while the disease prevalence of wheat stripe rust was represented by (0 0 0 0 1), (0 0 0 1 0), (0 0 1 0 0), (0 1 0 0 0) or (1 0 0 0 0).

2.2   Methods

Under MATLAB environment, backpropagation networks were used to process Data set 2 and Data set 4, radial basis networks and GRNNs were used to process Data set 1, Data set 2, Data set 3 and Data set 4, and PNNs were used to process Data set 1 and Data set 3. During data processing, disease prevalences were used as outputs and the other variables were used as inputs. For the data sets, the data in 1974-1993 were considered as the training sets and the data in 1994-1997 were considered as the testing sets. The prevalences of wheat stripe rust in 1974-1993 in Hanzhong in Shannxi Province were fitted and that in 1994-1997 were predicted. Firstly, the training sets were trained to acquire optimized neural networks. And then the neural networks were used to predict output values of the testing sets, and meanwhile the training sets were also fitted. The fitting and prediction results obtained were compared with actual disease prevalences and then fitting accuracy and prediction accuracy were obtained, respectively.

One-hidden-layer backpropagation networks were constructed. Three common transfer functions including logsig, tansig and purelin were used in the networks. Since the outputs of the networks were between 0 and 1, the output layer used the log-sigmoid transfer function (logsig). The networks used trainlm, traingd, traingdm, trainrp and trainbfg for training. Learngd and learngdm were used as learning functions. Different combinations of transfer functions, training functions and learning functions were recorded as BP1, BP2, BP3, … , BP30, and were shown as in Table 1. For Data set 2, fifteen neurons were used in the input layer and five neurons were used in the output layer. For Data set 4, five neurons were used in the input layer and five neurons were used in the output layer. Maximum number of epochs to train for the backpropagation networks was 5000. The neuron number in the hidden layer was assumed to be 1 to 20 with step size 1.

Radial basis networks, GRNNs and PNNs were used to fit the training sets and predict the disease prevalences using the testing sets. Spreads of radial basis functions for radial basis networks, GRNNs and PNNs were assumed to be 0.1 to 2.0 with step size 0.1.

Table 1. Different combinations of transfer functions, training functions and learning functions for backpropagation networks

Tag	Transfer function		Training function	Learning function
	Hidden layer	Output layer		
BP1	tansig	logsig	trainlm	learngdm
BP2	purelin	logsig	trainlm	learngdm
BP3	logsig	logsig	trainlm	learngdm
BP4	tansig	logsig	trainlm	learngd
BP5	purelin	logsig	trainlm	learngd
BP6	logsig	logsig	trainlm	learngd
BP7	tansig	logsig	traingd	learngdm
BP8	purelin	logsig	traingd	learngdm
BP9	logsig	logsig	traingd	learngdm
BP10	tansig	logsig	traingd	learngd
BP11	purelin	logsig	traingd	learngd
BP12	logsig	logsig	traingd	learngd
BP13	tansig	logsig	traingdm	learngdm
BP14	purelin	logsig	traingdm	learngdm
BP15	logsig	logsig	traingdm	learngdm
BP16	tansig	logsig	traingdm	learngd
BP17	purelin	logsig	traingdm	learngd
BP18	logsig	logsig	traingdm	learngd
BP19	tansig	logsig	trainrp	learngdm
BP20	purelin	logsig	trainrp	learngdm
BP21	logsig	logsig	trainrp	learngdm
BP22	tansig	logsig	trainrp	learngd
BP23	purelin	logsig	trainrp	learngd
BP24	logsig	logsig	trainrp	learngd
BP25	tansig	logsig	trainbfg	learngdm
BP26	purelin	logsig	trainbfg	learngdm
BP27	logsig	logsig	trainbfg	learngdm
BP28	tansig	logsig	trainbfg	learngd
BP29	purelin	logsig	trainbfg	learngd
BP30	logsig	logsig	trainbfg	learngd

3   Results and Analysis

The data sets were processed using the methods as shown above. The results of the neural networks with fitting accuracy ≥75% and prediction accuracy ≥50% were picked out and were shown in Table 2, Table 3, Table 4, Table 5, Table 6 and Table 7, respectively.

Using the backpropagation networks, for either Data set 2 or Data set 4, the selected combinations with good fitting accuracy and prediction accuracy were BP1, BP2, BP3, BP4, BP5, BP6, BP19, BP20, BP21, BP22, BP23, BP24, BP25, BP28, BP29 and BP30. Among these combinations, the training functions of BP1, BP2, BP3, BP4, BP5 and BP6 were trainlm, that of BP19, BP20, BP21, BP22, BP23, BP24 and BP25 were trainrp, and that of BP28, BP29 and BP30 were trainbfg. When the training function was traingd or traingdm, good fitting accuracy and prediction accuracy could not be obtained using the backpropagation networks. So it is important to choose suitable training function for backpropagation networks.

As Table 2 shown, for Data set 2, the fitting accuracies were all 100% and the prediction accuracies were all 75% while using BP1 with the number of neurons in the hidden layer equal to 3 or 11, BP3 with the number of neurons in the hidden layer equal to 10, 12 or 17, BP6 with the number of neurons in the hidden layer equal to 16, BP19 with the number of neurons in the hidden layer equal to 4, BP21 with the number of neurons in the hidden layer equal to 5, BP24 with the number of neurons in the hidden layer equal to 10 or 19, and BP28 with the number of neurons in the hidden layer equal to 5.

For Data set 4, the fitting results and the prediction results of the backpropagation networks were shown as in Table 3. Fitting accuracies were all 100% and prediction accuracies were all 75% while using BP4 with the number of neurons in the hidden layer equal to 9 and BP22 with the number of neurons in the hidden layer equal to 8 besides the combinations as the same as those with the same number of neurons in the hidden layer for Data set 2.

The results indicated that backpropagation networks could be used for the prediction of wheat stripe rust. When using this method, the inputs could be the data of all the factors obtained in the investigations or be the data of selected main factors. Otherwise, it would take a longer time to build suitable networks when using the data of all the influencing factors obtained.

The fitting results and the prediction results of radial basis networks were shown as in Table 4 and Table 5. The fitting accuracies were all 100% for the selected networks. The results implied that radial basis networks have good adaptive learning ability and stable prediction ability. For Data set 1, the prediction accuracy was 75% when the value of spread was 0.8, 0.9, 1.0 or 1.1. For Data set 2, the prediction accuracies were all 50% while using the selected networks. For Data set 3, the lowest prediction accuracy was 50% and the highest prediction accuracy was 100%. The prediction accuracy was 75% when the value of spread was 0.3, 0.4, 0.5, 1.5, 1.6, 1.7, 1.8, 1.9 or 2.0. The prediction accuracy was 100% when the value of spread was 0.6 to 1.4. For Data set 4, the prediction accuracy was 75% when the value of spread was 0.3, 1.2, 1.3, 1.4 or 1.5. The results indicated that radial basis networks have high prediction ability for the disease. In practical applications, it is better to choose the key influencing factors as the inputs of the networks to improve the run speed of the programs, prediction accuracy, and performance stability.

Table 2. The fitting results and the prediction results of the backpropagation networks for Data set 2

Tag	Number of neurons in the hidden layer	Fitting accuracy	Prediction accuracy	Tag	Number of neurons in the hidden layer	Fitting accuracy	Prediction accuracy
BP1	1	80%	50%	BP19	4	100%	75%
BP1	4, 5	85%	50%	BP20	10, 11, 13, 15, 16, 17, 18	100%	50%
BP1	6	95%	50%	BP21	2, 4, 14, 15, 16, 17, 19, 20	95%	50%
BP1	8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20	100%	50%	BP21	3, 6, 7, 8, 9, 10, 11, 12, 13, 18	100%	50%
BP1	7	95%	75%	BP21	5	100%	75%
BP1	3, 11	100%	75%	BP22	2, 4, 5, 10, 12, 17, 18	95%	50%
BP2	1, 2, 4, 13	85%	50%	BP22	6, 7, 9, 11, 13, 14, 15, 16, 19, 20	100%	50%
BP2	3	95%	50%	BP22	8	100%	75%
BP2	6	100%	50%	BP23	10	90%	50%
BP3	1, 7	85%	50%	BP23	5, 7, 12, 13, 14, 15, 17	95%	50%
BP3	9	90%	50%	BP23	4, 6, 8, 9, 11, 16, 18, 19, 20	100%	50%
BP3	3, 4, 11, 13, 16	95%	50%	BP24	1	85%	50%
BP3	2, 5, 8, 14, 15, 18, 19, 20	100%	50%	BP24	2, 3, 4, 9, 15, 16, 20	95%	50%
BP3	10, 12, 17	100%	75%	BP24	6, 8, 11, 12, 13, 14, 17, 18	100%	50%
BP4	7	75%	50%	BP24	10, 19	100%	75%
BP4	3, 4, 5, 17	85%	50%	BP25	15	80%	75%
BP4	2, 6, 8, 9, 10, 11, 16, 18, 19, 20	100%	50%	BP25	7	85%	50%
BP5	1	75%	50%	BP25	4, 18	95%	50%
BP5	2, 8	80%	50%	BP25	9, 11, 14, 17, 20	100%	50%
BP5	3, 5, 9	85%	50%	BP28	7, 11, 19	85%	50%
BP5	4, 20	90%	50%	BP28	1	90%	75%
BP5	6, 13, 15	95%	50%	BP28	4, 15	95%	50%
BP5	7, 10, 11, 17, 19	100%	50%	BP28	12, 13, 14, 17	100%	50%
BP6	11, 19	90%	50%	BP28	5	100%	75%
BP6	13	90%	75%	BP29	5	85%	50%
BP6	4	95%	50%	BP29	10	100%	50%
BP6	3, 5, 6, 7, 9, 10, 12, 14, 15, 17, 18	100%	50%	BP30	20	85%	50%
BP6	16	100%	75%	BP30	7, 13, 16	90%	50%
BP19	2	90%	50%	BP30	8	95%	50%
BP19	3, 8, 13	95%	50%	BP30	3	100%	50%
BP19	5, 6, 7, 9, 10, 11, 12, 14, 15, 16, 17, 20	100%	50%				

Table 3. The fitting results and the prediction results of the backpropagation networks for Data set 4

Tag	Number of neurons in the hidden layer	Fitting accuracy	Prediction accuracy	Tag	Number of neurons in the hidden layer	Fitting accuracy	Prediction accuracy
BP1	1	80%	50%	BP19	4	100%	75%
BP1	4, 5	85%	50%	BP19	5, 6, 7, 9, 10, 11, 12, 14, 15, 16, 17, 20	100%	50%
BP1	6	95%	50%	BP20	10, 11, 13, 15, 16, 17, 18	100%	50%
BP1	7	95%	75%	BP21	2, 4, 14, 15, 16, 17, 19, 20	95%	50%
BP1	8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20	100%	50%	BP21	3, 6, 7, 8, 9, 10, 11, 12, 13, 18	100%	50%
BP1	3, 11	100%	75%	BP21	5	100%	75%
BP2	1, 2, 4, 13	85%	50%	BP22	2, 4, 5, 10, 12, 17, 18	95%	50%
BP2	3	95%	50%	BP22	6, 7, 9, 11, 13, 14, 15, 16, 19, 20	100%	50%
BP2	6	100%	50%	BP22	8	100%	75%
BP3	1, 7	85%	50%	BP23	10	80%	50%
BP3	2, 5, 8, 14, 15, 18, 19, 20	100%	50%	BP23	5, 7, 12, 13, 14, 15, 17	95%	50%
BP3	9	90%	50%	BP23	4, 6, 8, 9, 11, 16, 18, 19, 20	100%	50%
BP3	3, 4, 11, 13, 16	95%	50%	BP24	1	85%	50%
BP3	10, 12, 17	100%	75%	BP24	2, 3, 4, 9, 15, 16, 20	95%	50%
BP4	2, 6, 8, 10, 11, 16, 18, 19, 20	100%	50%	BP24	6, 8, 11, 12, 13, 14, 17, 18	100%	50%
BP4	7	75%	50%	BP24	10, 19	100%	75%
BP4	3, 4, 5, 17	85%	50%	BP25	15	80%	75%
BP4	9	100%	75%	BP25	7	85%	50%
BP5	1	75%	50%	BP25	4, 18	95%	50%
BP5	2, 8	80%	50%	BP25	9, 11, 14, 17, 20	100%	50%
BP5	3, 5, 9	85%	50%	BP28	7, 11, 19	85%	50%
BP5	4, 20	90%	50%	BP28	1	90%	75%
BP5	6, 13, 15	95%	50%	BP28	4, 15	95%	50%
BP5	7, 10, 11, 17, 19	100%	50%	BP28	12, 13, 14, 17	100%	50%
BP6	11, 19	90%	50%	BP28	5	100%	75%
BP6	13	90%	75%	BP29	5	85%	50%
BP6	4	95%	50%	BP29	10	100%	50%
BP6	3, 5, 6, 7, 8, 9, 10, 12, 14, 15, 17, 18	100%	50%	BP30	20	85%	50%
BP6	16	100%	75%	BP30	7, 13, 16	90%	50%
BP19	2	90%	50%	BP30	8	95%	50%
BP19	3, 8, 13	95%	50%	BP30	3	100%	50%

The fitting results and the prediction results of GRNNs were shown as in Table 6. For Data set 1 including the fifteen influencing factors, the fitting accuracy was 100% and the prediction accuracy was 75% while the value of spread was equal to 0.3. For Data set 3 including the five selected factors, the fitting accuracy was 100% and the prediction accuracy was 75% while the value of spread was equal to 0.1 or 0.2. And for Data set 3, there was one exception that the fitting accuracy was only 80%, but the prediction accuracy was 75% while the value of spread was equal to 0.4. For Data set 2 including the fifteen influencing factors, the fitting accuracy was 100% and the prediction accuracy was 50% while the value of spread was equal to 0.2, 0.3 or 0.4. For Data set 4 including the five selected factors, the fitting accuracy was 100% and the prediction accuracy was 50% only when the value of spread was 0.1. The results showed that GRNNs could be used as a method to predict wheat stripe rust and that the prediction ability when disease prevalence of wheat stripe rust was expressed as 1, 2, 3, 4 or 5 was higher than that when disease prevalence was expressed as (0 0 0 0 1), (0 0 0 1 0), (0 0 1 0 0), (0 1 0 0 0) or (1 0 0 0 0). Therefore, disease prevalence could be expressed as 1, 2, 3, 4 or 5 when the prediction of wheat stripe rust is conducted using GRNNs.

Table 4. The fitting results and the prediction results of radial basis networks for Data set 1 and Data set 2

Data set	Spread	Fitting accuracy	Prediction accuracy
Data set 1	0.1	100%	50%
Data set 1	0.2	100%	50%
Data set 1	0.3	100%	50%
Data set 1	0.4	100%	50%
Data set 1	0.6	100%	50%
Data set 1	0.8	100%	75%
Data set 1	0.9	100%	75%
Data set 1	1.0	100%	75%
Data set 1	1.1	100%	75%
Data set 1	1.2	100%	50%
Data set 1	1.3	100%	50%
Data set 1	1.4	100%	50%
Data set 1	1.5	100%	50%
Data set 1	1.6	100%	50%
Data set 1	1.7	100%	50%
Data set 1	1.8	100%	50%
Data set 1	1.9	100%	50%
Data set 1	2.0	100%	50%
Data set 2	0.1	100%	50%
Data set 2	0.2	100%	50%
Data set 2	0.3	100%	50%
Data set 2	0.4	100%	50%
Data set 2	0.6	100%	50%
Data set 2	1.9	100%	50%
Data set 2	2.0	100%	50%

When PNNs were used to process the data of wheat stripe rust, the highest fitting accuracy was 100% and the highest prediction accuracy was 50% (as shown in Table 7). PNNs showed relatively low prediction ability. For Data set 1, the fitting accuracy was 100% and the prediction accuracy was 50% when the value of spread was 0.1, 0.2, 0.3 or 0.4. For Data set 3, the fitting accuracy was 100% and the prediction accuracy was 50% when the value of spread was 0.1. The fitting accuracy had a tendency to decrease with the value of spread increasing.

Table 5. The fitting results and the prediction results of radial basis networks for Data set 3 and Data set 4

Data set	Spread	Fitting accuracy	Prediction accuracy	Data set	Spread	Fitting accuracy	Prediction accuracy
Data set 3	0.1	100%	50%	Data set 4	0.1	100%	50%
Data set 3	0.2	100%	50%	Data set 4	0.2	100%	50%
Data set 3	0.3	100%	75%	Data set 4	0.3	100%	75%
Data set 3	0.4	100%	75%	Data set 4	0.4	100%	50%
Data set 3	0.5	100%	75%	Data set 4	0.5	100%	50%
Data set 3	0.6	100%	100%	Data set 4	0.6	100%	50%
Data set 3	0.7	100%	100%	Data set 4	0.7	100%	50%
Data set 3	0.8	100%	100%	Data set 4	0.8	100%	50%
Data set 3	0.9	100%	100%	Data set 4	0.9	100%	50%
Data set 3	1.0	100%	100%	Data set 4	1.0	100%	50%
Data set 3	1.1	100%	100%	Data set 4	1.1	100%	50%
Data set 3	1.2	100%	100%	Data set 4	1.2	100%	75%
Data set 3	1.3	100%	100%	Data set 4	1.3	100%	75%
Data set 3	1.4	100%	100%	Data set 4	1.4	100%	75%
Data set 3	1.5	100%	75%	Data set 4	1.5	100%	75%
Data set 3	1.6	100%	75%	Data set 4	1.6	100%	50%
Data set 3	1.7	100%	75%	Data set 4	1.7	100%	50%
Data set 3	1.8	100%	75%	Data set 4	1.8	100%	50%
Data set 3	1.9	100%	75%	Data set 4	1.9	100%	50%
Data set 3	2.0	100%	75%	Data set 4	2.0	100%	50%

Table 6. The fitting results and the prediction results of GRNNs

Data set	Spread	Fitting accuracy	Prediction accuracy
Data set 1	0.2	100%	50%
Data set 1	0.3	100%	75%
Data set 1	0.4	100%	50%
Data set 1	0.5	95%	75%
Data set 1	0.6	95%	50%
Data set 2	0.2	100%	50%
Data set 2	0.3	100%	50%
Data set 2	0.4	100%	50%

Table 6. (*Continued*)

Data set 2	0.5	95%	50%
Data set 2	0.6	95%	50%
Data set 2	0.7	85%	50%
Data set 2	0.8	85%	50%
Data set 3	0.1	100%	75%
Data set 3	0.2	100%	75%
Data set 3	0.3	90%	75%
Data set 3	0.4	80%	100%
Data set 4	0.1	100%	50%
Data set 4	0.2	95%	50%
Data set 4	0.3	95%	50%
Data set 4	0.4	80%	50%
Data set 4	0.5	80%	50%

Table 7. The fitting results and the prediction results of PNNs for Data set 1 and Data set 3

Data set	Spread	Fitting accuracy	Prediction accuracy
Data set 1	0.1	100%	50%
Data set 1	0.2	100%	50%
Data set 1	0.3	100%	50%
Data set 1	0.4	100%	50%
Data set 1	0.5	95%	50%
Data set 1	0.6	95%	50%
Data set 1	0.7	85%	50%
Data set 1	0.8	80%	50%
Data set 3	0.1	100%	50%
Data set 3	0.2	95%	50%
Data set 3	0.3	95%	50%
Data set 3	0.4	80%	50%
Data set 3	0.5	80%	50%

4   Conclusion and Discussion

In this study, backpropagation networks with different transfer functions, training functions and learning functions, radial basis networks, GRNNs and PNNs were applied to predict wheat stripe rust, and comparison of prediction effects of these methods were also conducted. The results showed that good fitting accuracy and prediction accuracy could be achieved using suitable backpropagation networks, radial basis networks and GRNNs, and that the methods based on these three algorithms of neural networks could be used for the prediction of wheat stripe rust. In contrast, radial basis networks had more power to predict wheat stripe rust than backpropagation networks and GRNNs, and the programs of GRNNs were simpler than backpropagation networks and needed less time to run. The backpropagation

networks using traingd or traingdm as the training function could not fit the data of wheat stripe rust well and could not make accurate predictions. The studies on using neural networks to predict wheat stripe rust reported before mainly used backpropagation networks [15], [16], [17]. This study provides new approaches based neural networks for the prediction of this disease.

Many studies on the prediction of wheat stripe rust have been conducted and many methods have been applied for the prediction of this disease. However, some kinds of problems such as small sample size are often met in the process of making disease prediction and could not be solved very well. Support vector machine (SVM), a new kind of machine learning method proposed by Vapnik [19], could solve these problems with many advantages. SVM is an algorithm based on VC dimension theory and structural risk minimization principle. It could solve the small sample, nonlinear problems, high dimension and local minimum points and other practical issues. SVMs are currently being used in many fields, such as text classification [20], [21], image recognition [22], remote sensing information analysis and processing [23], [24], disastrous weather forecast [25], and so on. The prediction models based on SVMs to predict wheat stripe rust timely and accurately could be established.

In the studies on the prediction of wheat stripe rust, many prediction models have been established. However, each model generally has some certain application conditions. Therefore, suitable model should be selected for application according to the actual conditions. Prediction model system for wheat stripe rust could be built to collect the prediction models into one computer system through programming the models. It could make easier to choose suitable model. Integrated prediction model could be established using different models and methods according to the needs in the future.

References

1. Li, Z.Q., Zeng, S.M.: Wheat Rusts in China. China Agriculture Press, Beijing (2002) (in Chinese)
2. Chen, X.M.: Epidemiology and Control of Stripe Rust [*Puccinia striiformis* f. sp. *tritici*] on Wheat. Can. J. Plant Pathol. 27, 314–337 (2005)
3. Line, R.F.: Stripe Rust of Wheat and Barley in North America: a Retrospective Historical Review. Annu. Rev. Phytopathol. 40, 75–118 (2002)
4. Zeng, S.M.: Interregional Spread of Wheat Yellow Rust in China. Acta Phytopathologica Sinica 18, 119–223 (1988) (in Chinese)
5. Yang, Z.W., Shang, H.S., Pei, H.Z., Xie, Y.L.: Dinamic Forecasting of Stripe Rust of Winter Wheat. Scientia Agricultura Sinaca 24, 45–50 (1991) (in Chinese)
6. Hu, X.P., Yang, Z.W., Li, Z.Q., Deng, Z.Y., Ke, C.H.: Studies on the Prediction of Wheat Stripe Rust Epidemics in Hanzhong District of Shaanxi Province. Acta Univ. Agric. Boreali-occidentalis 28, 18–21 (2000) (in Chinese)
7. Fan, S.Q., Xie, X.S., Li, F., Yin, Q.Y., Zheng, W.Y.: Forecast Model for Prevalent Stripe Rust in Winter Wheat in Shanxi Province. Chinese Journal of Eco-Agriculture 15, 113–115 (2007) (in Chinese)

8. Chen, G., Wang, H.G., Ma, Z.H.: Forecasting Wheat Stripe Rust by Discrimination Analysis. Plant Protection 32, 24–27 (2006) (in Chinese)
9. Chen, G., Wang, H.G., Zhang, L.D., Wang, T., Ma, Z.H.: Preliminary Research on the Regional Relationship of Epidemic of *Puccinia striiformis* in China. Chinese Agricultural Science Bulletin 22, 415–420 (2006) (in Chinese)
10. Yun, X.W., Wang, H.G., Ma, Z.H.: Forecast of Wheat Stripe Rust by Upper-air Wind. Chinese Agricultural Science Bulletin 23(8), 358–363 (2007) (in Chinese)
11. Qiang, Z.F.: Markov Forecast of Wheat Stripe Rust in Qinghai in 1998. Plant Protection 25, 19–22 (1999) (in Chinese)
12. Yuan, L., Li, S.Q.: Application of Principal Component Analysis of Wheat Stripe Rust. Computer Engineering and Design 31, 459–461 (2010) (in Chinese)
13. Pu, C.J.: On Periodic Epidemic Regular Pattern and Prediction of Wheat Stripe Rust in Gansu Province. Acta Phytopathologica Sinica 28, 299–302 (1998) (in Chinese)
14. Liu, R.Y., Ma, Z.H.: The Prediction Methodology of Wheat Stripe Rust Using Combination Model Based on GM(1,l). Journal of Biomathematics 22, 343–347 (2007) (in Chinese)
15. Hu, X.P., Yang, Z.W., Li, Z.Q., Deng, Z.Y., Ke, C.H.: Prediction of Wheat Stripe Rust in Hanzhong Area by BP Neural Network. Acta Agriculturae Boreali-occidentalis Sinica 9, 28–31 (2000) (in Chinese)
16. Zhang, J.: Research on Chaotic Characteristics of the Disaster Rate of Crops and Its GA-BPNN Forecasting Model. Jour. of Northwest Sci-Tech Univ. of Agri. and For(Nat. Sci. Ed.) 34, 63–66 (2006) (in Chinese)
17. Jin, N., Huang, W.J., Jing, Y.S., Wang, D.C., Luo, J.H.: Long-term Meteorological Prediction of Country Wide Wheat Stripe Rust by Genetic Neural Network. Chinese Journal of Agrometeorology 30, 243–247 (2009) (in Chinese)
18. Ma, Z.H.: Plant Disease Epidemiology. Science Press, Beijing (2010) (in Chinese)
19. Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer, New York (1995)
20. Bao, J., Ji, M., Feng, J.: Text Categorization Method Based on Fuzzy Support Vector Machine. Journal of Liaoning Technical University (Natural Science) 29, 974–977 (2010) (in Chinese)
21. Liu, X.L., Ding, S.F., Zhu, H., Zhang, L.W.: Appropriateness in Applying SVMs to Text Classification. Computer Engineering & Science 32, 106–108 (2010) (in Chinese)
22. Liu, S.: Image Recognition Based on SVM Information Fusion and DSP Parallel Realization. Computer Engineering and Applications 45, 168–170 (2009) (in Chinese)
23. Wang, H.G., Ma, Z.H., Wang, T., Cai, C.J., An, H., Zhang, L.D.: Application of Hyperspectral Data to the Classification and Identification of Severity of Wheat Stripe Rust. Spectroscopy and Spectral Analysis 27, 1811–1814 (2007) (in Chinese)
24. Fu, J.E., Su, Q.X., Pan, S.B., Lu, J.X.: Support Vector Machine Based Groundwater Level Monitoring Model by Using Remote Sensing Images. Journal of Geo-information Science 12, 466–472 (2010) (in Chinese)
25. Wang, D.C., Wang, C.X., Zhu, T.Y., Qin, J.: Support Vector Machines Based Algorithm for the Disastrous Weather Forecast. Journal of Wuhan University of Technology 32, 121–124 (2010) (in Chinese)

Research on Surface Sampling for Determination of Pesticide Residues in Pome Fruit

Yunxia Luan[1,2], Hua Ping[1,2], Zhihong Ma[1,2], and Ligang Pan[1,2]

[1] Beijing Research Center for Agrifood Testing and Farmland Monitoring, Beijing, China
[2] National Engineering Research Center for Information Technology in Agriculture, Beijing, China
`luanyx@nercita.org.cn`

Abstract. A surface sampling method for analysis of pesticide residues in pome fruit was studied by layer-by-layer scan from pericarp to kernel. Phoxim residues of apples, pears and papayas was determined by GC-MS, respectively, in core, different layers of pulp, peel ,as well as the whole fruit. Phoxim residue content was decreased when depth of the layer increased. It was inferred from experimental data the the appropriate depth of surface sampling is 2.5mm, 2mm and 3.5mm in apples, pears and papayas. Although the ratio of the pesticide in pericarp and the total (Cp/Ct) ranged from 6.241 to 9.262 in apple, 4.254 to 5.069 in pears, and 4.39 to 5.037 in papayas in different groups, the ratio of Cp/Ct and quality of specific group is essentially unchanged for a certain kind of fruit. According to the results, a conversion formula of the content of pesticide residues in whole fruit and pericarp was set, based on surface sampling method.

Keywords: Pome fruit, Pesticide residues, Phoxim, Surface sampling.

1 Introduction

One of the major roles of public health agencies is to ensure safe products for consumers through analysis of residual agricultural chemicals and veterinary drugs in foods. The use of agricultural chemicals and veterinary drugs in agriculture is necessary to improve the quality of the food produced. They play a beneficial role in providing a plentiful, low-cost supply of high-quality food. On the other hand, as a consequence of this use, the presence of residues in food that is a critical element of overall public health is unavoidable, and residues in food are of great importance in the evaluation of food quality. Organophosphorus (OP) pesticides are the most widely used agricultural pesticide. More than 70% of that are hyper-toxic or high-toxic, and normally are forbidden to use in planting fruits and vegetables. OP insecticides are routinely applied to fruit crops (e.g., apple, peach, pear) that are subject to consumption as single servings [1]. Consumption of individual fruits containing high OP insecticide residues may result in high exposure over a short period of time. Residue testing for the presence

of pesticides, however, is routinely performed on composites of 5 to 10 individual or unit samples rather than samples of individual fruit [2].

Experimental study of the fruit for pesticide residues in agricultural products exist in the form, that is attached to the fruit into the fruit surface and the distribution of pesticides within the fruit. It is necessary to provide a representative sample to the laboratory for analysis [3]. No matter how carefully laboratory analyses are carried out, the results will be of limited value unless the sample provided is truly representative of the sample being tested [4]. The uneven distribution of pesticide residues among the treated objects leads to an inevitable variability of pesticide residue levels measured in the samples, which may significantly contribute to the combined uncertainty of the analytical results. Data generated in the UK have indicated that pesticide residue levels can be highly variable between the individual fruit from the same crop or lot in trade. Many fruit or vegetables in trade are mixed after harvest to form combined lots [5]. Analysis of composite samples provides a good indication of average residues but, where the lot has been mixed, such average values are potentially misleading [6]. Residues monitoring is the best means available for general control of pesticide use and consumer exposure, but new strategies for sampling and analysis are required to address the combined effects of residues variability and mixing of lots.

Development of analytical methods for pesticides in agri-food requires precise method of sampling. Pome fruit peel is the main remaining parts, parts of the peel residue is a measure of its ability to meet the safety standard limit fruit key factor. The apple pesticide residue question has become the chief obstacle in affecting the apple industry internationalization of our country [7]. At present, samples of pome fruit pesticide residue mainly from the entire apple or some edible parts, because the results from these have the generalization. However, the pesticide residue mainly exists in the pericarp. Therefore, the careful research to pesticide residue of the pericarp provides the theory instruction and the design basis to the pesticide residue and the elimination research.

However, limited data have been reported that directly compare the sampling procedures against one another. Even less information is published regarding the most appropriate part of the material to use in conjunction with a sampling procedure for pesticides detection, which is particularly important to the portable rapid detector of contamination in fruits and vegetables [8]. In our Previous studies, the temporal and spatial variations of phoxim residue in apples were studied. The results show that half-life of phoxim in apples was 1.64 days. The phoxim residue order in different position of apple is: pericarp >entire apple> sarcocarp> kernel and the ratio of residue in pericarp and the entire apple ranged from 6.241 to 9.262 in groups with different specific surface area [9]. The results indicate that, there are significant differences in the contents of residues between the pericarp and entire apple, which provide evidence in the theory instruction and design basis to the sampling method of pesticide residue based on pericarp. In this paper, further research was carried out on a surface sampling method for analysis of pesticide residues in pome fruit was studied by layer-by-layer scan from pericarp to kernel, and to explore the relationship between the content of pesticide in pericarp and the whole fruit.

2   Materials and Methods

2.1   Chemicals and Reagents

Phoxim standard was purchased from National Institute of Metrology P.R.China. High-purity solvents cyclohexane, acetone, and acetonitrile were all purchased from Fisher Chemicals [10]. Analytical reagents magnesium sulfate and sodium chloride were purchased from Beijing Chemical Reagents Company (China). High-purity solvents acetonotrile, suitable for liquid chromatography, gas chromatography, and residue analysis used in the extraction and cleanup steps, were purchased from EMD Biosciences Inc(Mississauga, ON). Reagent grade $MgSO_4$ and sodium chloride were also purchased from EMD Biosciences Inc.

2.2   Apparatus

The gel permeation chromatography (GPC) equipped with gas chromatography mass spectrometry system（GC/MS）was used. The GPC consists of two LC-10ADvp pumps, a SIL-10ADvp auto-sampler, a Shodex CLNpak EV-200AC column (2mm i.d.×150 mm) and CTO-10ASvp column oven, a SPD-10Avp UV detector, two FCV-12AH flow channel selectionvalves (RV.A, RV.B) and a SCL 10Avp system controller. GC/MS machine is a Shimadzu GC/MS-QP2010 instrumentation equipped with a PTV-2010 large-volume injection device.

   GC/MS data analysis was triggered by a contact closure start signal from the HPLC controller. Data acquisition was performed using a C-R8A plus data processor. All these parts are the products of Shimadzu (Kyoto, Japan), except the Shodex CLNpak EV-200AC column (Shoko Co., Tokyo, Japan). Acetone/ cyclohexane mixing solvent (3/7, v/v) was used as the mobile phase of GPC [11], and the flow rate was set at 0.1 mL/min. The mobile phase was degassed using DGU-14A degasser (Shimadzu), and the GPC column was kept at 40℃ in the column oven.

2.3   Sample Preparation

All samples were purchased at local markets in Beijing. Samples were extensively collected to achieve good sample homogeneity. The phoxim compounds were applied using an air blast sprayer and application rates guaranteed to contain 50% phoxim emulsifiable with the dilution multiple of 1:500. Individual fruits were cored and sliced into 5 segments (pericarp(P), sarcocarp1(S1), S2,S3,S4 and kernel) using a corer/slicer retailed and peeling machine for domestic use. The first and alternate slices of each apple were taken, chopped manually using a knife, and placed in a plastic bag for storage at -25℃ until extraction and analysis. The remaining segments of individual apples were retained in separate bags for preparation of composite samples. Composites were constructed from the apples prepared from each group by randomly selecting 8 apples for each treat, without regard to the style and weight of each apple. Composite samples were prepared by thoroughly mixing the chopped apple pieces

from the bags containing the retained portions of the individual apples. Composite samples were frozen until extraction and analysis. To prepare each sample, 10 g of a previously homogenized food materialwas transferred into a suitable glass vessel. Then, 10mL acetonitrile was added to each sample using an adjustable-volume solvent dispenser. The glass vessels were capped before vortex mixing for 1 min at maximum speed. Once the initial sample mixing was completed, 1 g NaCl and 4 g anhydrous $MgSO_4$ were added and mixed immediately on a Vortex mixer for 1 min. It is important to note that this step must be taken immediately after the initial mixing step to prevent the formation of $MgSO_4$ conglomerates. To separate phases, samples were centrifuged for 10 min at 1570×g. Using an adjustable repeating pipette, 1.0mL aliquot of upper acetonitrile layer was transferred into a 1.5mL flip-top microcentrifuge vial containing 150 mg anhydrous $MgSO_4$ and 50 mg PSA sorbent. The vial was tightly capped and shaken on a vortex mixer for 1 min before extraction. Then the mixed extraction solution was centrifuged for 5 min to separate solids from solution. The solution was then transferred into an autosampler for GPC-GC/MS analysis.

2.4   Determination of Phoxim in Apple by GPC-GC/MS

Analysis was performed by GPC-GC/MS. The analytical GC/MS equipped with a PTV-2010 large-volume injection device was carried out using deactivated silica tubing (5m×0.53mm i.d.) , a Rtx-5MS pre-column (5m×0.25mm) and a Rtx-5MS column (30m×0.25mm×0.25μm film thickness; coated with 5% phenyl and 95% methylpolysiloxane; Restek Corporation, Bellefonte, Palo Alto, USA). The temperature of the PTV injector was set at 120℃ for the initial 5 min of sampling time, and then increased to 250℃ at 100℃/min. The oven temperature was maintained at 80℃ for 5 min, subsequently increased to 280℃ at a rate of 8℃/min, and then held constant for 10 min [12]. The quadrupole mass spectrometer was operated in the electron impact ion (EI) mode. Ion source temperature and interface temperature were set at 200℃ and 250℃, respectively. The mass spectrometer was operated in an ionizing energy of 70 eV. Injection volumes were 1 μL for all analyses. Helium（99.999%）was used as the carrier gas [13].

2.5   Quality Assurance/Quality Control

Apples from market with no phoxim insecticide application were obtained, prepared the same as the rinsed and peeled apples, and were used to prepare blank apple matrix for quality assurance testing. With each set of samples 10-15 samples per set) extracted and analyzed, two aliquots of blank apple matrix were prepared for separate extraction followed by extraction, cleanup, and analysis as for all other samples. Background levels of some analysis were periodically detected in the blank matrix samples and were used for background subtraction in the determination of recovery from spiked matrix only; residue concentrations in samples were not blank corrected. No traces of the phoxim were observed in any of the reagent blanks analyzed.

3   Results and Discussion

3.1   Determine of the Parameters for Surface Sampling

In order to determine the exact thickness of pericarp and sampling style, the materail was cut by 1.5mm-2mm each piece from the pericarp to kernel, and pesticide residues in fruit peel, kernel and pulp of different depth were determined. Taking the Fuji apple as example, according to dilution ratio of 1:1000 dilution of Phoxim EC, more than fifteen days after spraying pesticide on fruit, pesticide residues in samples was detected according to the method of NY/T 761.1-2004 "vegetables and fruits organophosphorus, organochlorine, pyrethroid and carbamate pesticide multi-residue methods-Part 1 of fruits and vegetables organophosphorus pesticide multi-residue method" and GB/T5009.199-2003 "vegetables of organophosphorus and carbamate pesticide residues in rapid detection." for lab and rapid detection. Each sub-sample was taken from 15 samples of different shape and weight. The results are presented in Figure 1.
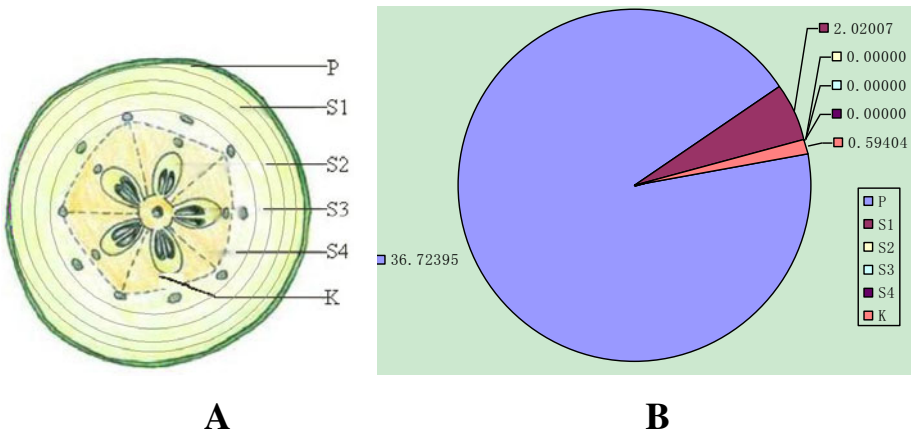
Fig. 1. Compartmentalization and sampling methord of apple (A) and the content of Phoxim in each depth (B)

P: pericarp of apples with 0.5mm pulp, S1: The first layer with 2mm pulp, S2: the second layer with 2mm sarcocarp , S3: the third layer with 2mm sarcocarp, S4: the sarcocarp near kernel, K: kernel or core of apples.

Experimental study of the pesticide residues in pome fruit, which is attached to the fruit surface and the distribution of pesticides within the fruit was carried out. From experiments performed and analysis of pesticide residues in different layer from pericarp to kernel, it is showed that, pesticide was detected maily in layer P and S1, the pesticide in pericarp and the first layer with 2mm pulp account for 98.49% of the whole sample. Through different parts of the fruit residue test results revealed that the major part of remaining in the apple peel, fruit pulp and other parts of the residual is minimal. According to the apple skin to the stone from the stratified analysis, pesticide residues, mainly in the skin and pulp of the following 2.5mm, this part of the sample selected sampling to detect the levels of residues of pesticides, because this part of the high levels of pesticide residues, can increase rapidly screening test accuracy.

According to the experimental results, the standard design of the skin sampling methods, reduction in access to sub-samples the following 2.5mm thick skin and flesh as the test object.

3.2   Study of the Weight Coefficient of Pesticide Residues in Pericarp

Selected based on the standard higher levels of pesticide residues in samples of skin as a target for different types of fruits, pesticide residues on its surface and overall there is a certain amount of pesticide residue conversion relationship [14]. The research was supported by the skin of pesticide content in the trend over time to determine the appropriate sampling time interval after application, and in accordance with the standard parameters of the specific sampling techniques to obtain test samples, the research team through the following experiments, made pome fruit surface pesticide residue content and the overall coefficient of pesticide content in the skin pesticide residues.

Using Apple, pear, and papaya(n=15), sampling the surface homogenized pericarp(with 2.5mm pulp) as detection material, homogenized with a whole apple parallel processing [15], determination of pesticide content, access to pesticide skin factor, diluted with water according to the proportion of 1:1000 dilution phoxim EC, using the method of spraying pesticide on apples [16]. According to NY/T761.1-2004 "vegetables and fruits organophosphorus, organochlorine, pyrethroid and carbamate pesticide multi-residue methods-Part1 of fruits and vegetables organophosphorus pesticide multi-residue method" approach with gas chromatography of whole fruit and pericarp of the accurate detection of pesticide residues pesticide residues in order to obtain skin factor [17].

As the result in Fig2-4, the three kinds of fruit showed various levels of pesticide in pericarp and the whole fruit, though the content in pericarp was much higher than the total homogenate. So, the ratio of the pesticide in pericarp and the whole fruit (Cp/Ct) was got from the three kinds of fruits with different shape and quality as shown in Table1.
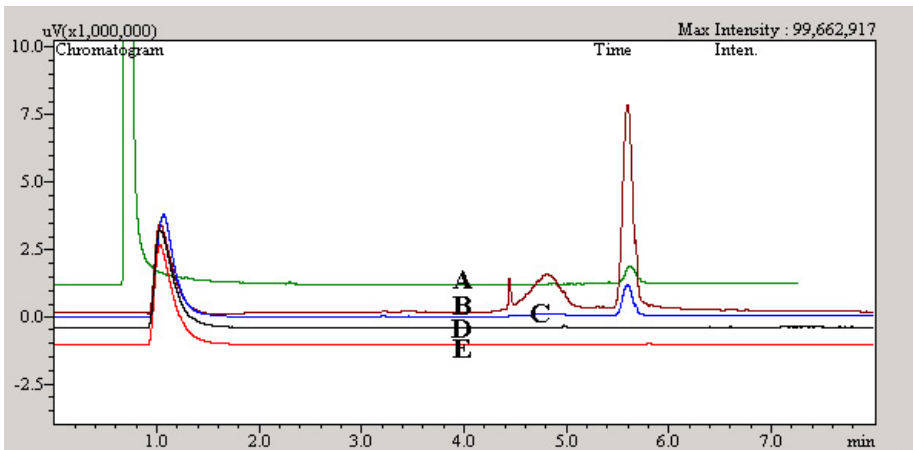
Fig. 2. GC diagram of phoxim in Fuji apple

A phoxim 5.25 mg/kg; B sample of pericarp;C sample of the whole fruit; D and E as CK.
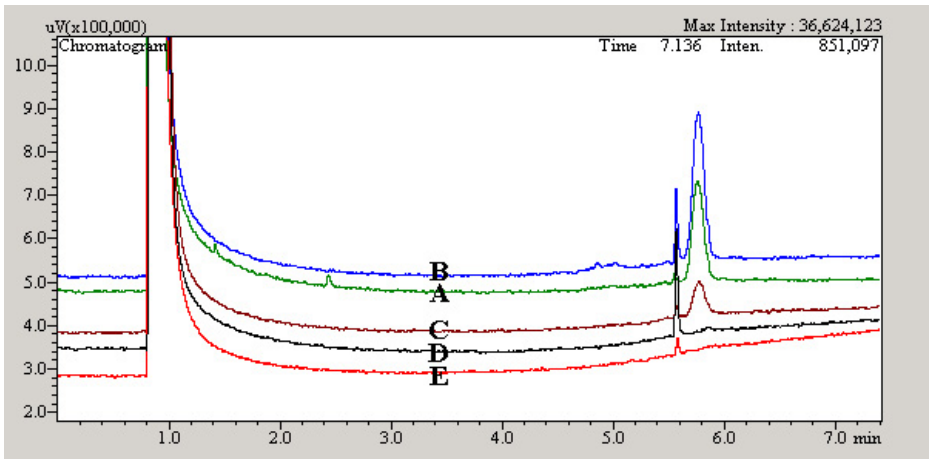
A phoxim 5.25 mg/kg; B sample of pericarp;C sample of the whole fruit; D and E as CK.

Fig. 3. GC diagram of phoxim in pear
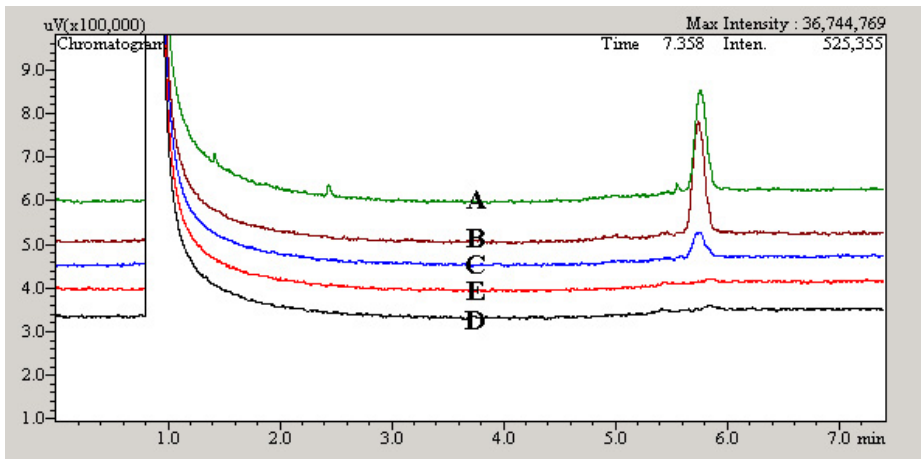
A phoxim 5.25 mg/kg; B sample of pericarp;C sample of the whole fruit; D and E as CK.

Fig. 4. GC diagram of phoxim in papaya

Table 1. Weight Coefficient of pesticide residues in pericarp of common pome fruit

Fruit	Quality(g)	C_p/C_T
Fuji apple	169.605±4.666	6.241
	189.425±4.748	7.963
	209.032±4.015	9.262
Pear	135.098±3.411	4.254
	146.270±3.566	4.516
	159.215±3.652	5.069
Papaya	1060.96±10.21	4.390
	1072.21±10.33	4.815
	1080.67±11.19	5.037

3.3 Formula for Calculating the Amount of Pesticide Residues in the Epidermis

Although the ratio of the pesticide in pericarp and the total (Cp/Ct) ranged from 6.241 to 9.262 in apple, 4.254 to 5.069 in pears, and 4.39 to 5.037 in papayas in different groups of weight, the ratio of Cp/Ct and quality of specific group is essentially unchanged for a certain kind of fruit (Table 2). According to these results, a conversion formula of the content of pesticide residues in whole fruit and pericarp was set based on surface sampling method.

The weight of a pesticide(a) refers to the epidermis and fruit pesticide content ratio of the overall pesticide content(1).

$$a = \frac{C_p}{c_t} \tag{1}$$

a: Weight coefficient of pesticide residues in pericarp,
C_t: total pesticide content,
C_p: pesticide content in the epidermis

Coefficient A refers to the contribution of the content of epidermal pesticide to the whole fruit

$$A = \frac{a}{m} \tag{2}$$

A: surface pesticide coefficient,
a: Weight coefficient of pesticide residues in pericarp,
m: the quality of a single sample

According to the data of Weight coefficient of pesticide residues in pericarp of common pome fruits in Table1, the surface pesticide coefficient could be calculated

based on formula (2). As shown in Table2, the surface pesticide coefficient A of Fuji apple, pear and papaya were 0.041, 0.031 and 0.004 respectively.

So fruit samples with pesticide residues in total mass fraction of w, in units of milligrams per kilogram (mg / kg), according to the measured pesticide residues in skin samples according to the formula (3)

$$w = \frac{w_p}{A \times m} \tag{3}$$

w - the total pesticide residues in fruits, in milligrams per kilogram (mg / kg);

w_p - the amount of pesticide residues in fruit pericarp, in milligrams per kilogram (mg / kg);

m - single sample sample quality of the sample, in grams (g);

A - surface pesticide coefficient

Table 2. Surface pesticide coefficient in common pome fruit

Fruit	A
Fuji apple	0.041
Pear	0.031
Papaya	0.004

3.4   Application of the Surface Sampling Method on Rapid Detector of Pesticides

Select the most commonly used Rapid detector of Pesticides or detector have a fast detector function ,to detect pesticides in food safety rapid detector to verify the sampling method, the results shown in Table 3. Take apples, pears and papaya samples of different quality and shape of the three groups of 15 individuals, according to this standard involve the rapid detection of pesticide residues in fruit skin sampling methods in the technical parameters, take the skin (with 2.5mm pulp) as sample, while the overall homogenized fruit sample, according to GB/T 5009.199-2003 "vegetables of organophosphorus and carbamate pesticide residues in the rapid detection," Determination of pesticide content. It can be seen from the results, obtained in accordance with the samples of fruit skin test results of pesticide residues, pesticide residues on the enzyme that the inhibition rate was significantly greater than the overall sample of inhibition of fruit (see Table 3).

The result of rapid detectors using whole apple sample and surface sample demonstrated that, a surface sampling method for analysis of pesticide residues in pome fruit could increased sensitivity and reduced the average impact of whole fruit homogenates. Combination of the phoxim residues of apples, pears and papayas determined by GC-MS in different layers of the fruit and the formula for calculating the amount of pesticide residues in the epidermis amount of pesticide residues in the epidermis, an efficient surface sampling method was studied and established.

Table 3. Comparision of the results of rapid detectors using different sampling method

Sample	Inhibition rate （%）
Whole apple	34.6±5.31
Peel of apple	84.4±4.89
Whole pear	30.1±5.83
Peel of pear	71.8±3.26
Whole papaya	35.4±2.91
Peel of papaya	91.3±6.03

4   Discussion

According to the reseach of pesticide residues in pome fruit by layer-by-layer scan from pericarp to kernel, the previous solution of the significant differences in the contents of residues between the pericarp and entire apple was demonstrated, further more these research revealed the pesticide residues, mainly remained in the pericarp and pulp of the following 2.5mm thick, this part of the sample could be selected as sample to detect the levels of residues of pesticides, because this part of the high levels of pesticide residues, can increase rapidly screening test accuracy [18]. From the experimental results obtained, the standard design of the skin sampling methods, reduction in access to sub-samples the following 2.5mm thick skin and flesh as the test object.

Scientific and standardized sampling method can reduce the skin samples and pulp composition on the average of the impact test results, reduce false positives and improve accuracy [19]. But different types of fruit, types of pesticides and interval time after spraying should be further studied to verify the coefficient and formula. Effect of surface waxes on the persistence of pesticide in Fuji apples, pears, and papayas would effect the diffusion of pesticide to pulp.

Coefficient A refers to the epidermal skin pesticide pesticide content of the fruit content of the overall contribution of major factors affecting A include the following factors: (1) differences include different types of fruit and fruit surface area for adsorption of pesticides in different penetrate the skin of the fruit impact of pesticides [20]. (2) determination of pesticide residues in skin samples when differences in sampling methods must be standardized to ensure that the skin sample, and with the standard method for sampling the epidermis consistent with the provisions. (3) the time factor after the recommended interval of sampling.

Therefore, this surface sampling method and the corresponding formula can be applied to various methods of detection of pesticide residues, in particular, the portable rapid detection machines. As the rapid development of materials science, biology and chemistry technologies, the surface sampling method would be widely used in sensor and other equipments.

References

1. Hamers, T., van den Brink, P.J., et al.: Estrogenic and esterase-inhibiting potency in rainwater in relation to pesticide concentrations, sampling season and location. Environ. Pollut. 123(1), 47–65 (2003)
2. Fenske, R.A., Bradman, A., et al.: Lessons learned for the assessment of children's pesticide exposure: critical sampling and analytical issues for future studies. Environ. Health Perspect. 113(10), 1455–1462 (2005)
3. Wang, J., Tuduri, L., et al.: Flexibility of solid-phase microextraction for passive sampling of atmospheric pesticides. J. Chromatogr. A 1216(15), 3031–3037 (2009)
4. Wang, K., Shipp, J.L.: Sequential sampling plans for western flower thrips (Thysanoptera: Thripidae) on greenhouse cucumbers. J. Econ. Entomol. 94(2), 579–585 (2001)
5. Ambrus, A.: Estimation of sampling uncertainty for determination of pesticide residues in plant commodities. J. Environ. Sci. Health B 44(7), 627–639 (2009); Chromatogr A 1216(15), 2972–2983
6. Yusa, V., Coscolla, C., et al.: Sampling and analysis of pesticides in ambient air. J. Chemosphere 51(6), 509–513 (2009)
7. Neumann, M., Liess, M., et al.: A qualitative sampling method for monitoring water quality in temporary channels or point sources and its application to pesticide contamination (2003)
8. Shen, J., Liu, J.: Determination of carbendazim residue in orange and soil using high performance liquid chromatography. Se Pu. 27(3), 308–312 (2009)
9. Luan, Y., Ping, H., Pan, L.: Residue dynamics of phoxim in pericarp, sarcocarp and kernel of apple. In: Li, D., Liu, Y., Chen, Y. (eds.) CCTA 2010, Part II. IFIP AICT, vol. 345, pp. 457–464. Springer, Heidelberg (2011)
10. Shen, J., Liu, J.: Determination of carbendazim residue in orange and soil using high performance liquid chromatography. Se Pu. 27(3), 308–312 (2009)
11. Schafer, R.B., Paschke, A., et al.: Performance of the Chemcatcher passive sampler when used to monitor 10 polar and semi-polar pesticides in 16 Central European streams, and comparison with two other sampling methods. Water Res. 42(10-11), 2707–2717 (2008)
12. Schafer, R.B., Paschke, A., et al.: Performance of the Chemcatcher passive sampler when used to monitor 10 polar and semi-polar pesticides in 16 Central European streams, and comparison with two other sampling methods. Water Res. 42(10-11), 2707–2717 (2008)
13. Schafer, R.B., Pettigrove, V., et al.: Effects of Pesticides Monitored with Three Sampling Methods in 24 Sites on Macroinvertebrates and Microorganisms. Environ. Sci. Technol.
14. Bernard, C.E., Berry, M.R., et al.: Sampling household surfaces for pesticide residues: comparison between a press sampler and solvent-moistened wipes. Sci. Total Environ. 389(2-3), 514–521 (2008)
15. Rosenheck, L., Cowell, J., et al.: Determination of a standardized sampling technique for pesticide transferable turf residues. Bull. Environ. Contam. Toxicol. 67(6), 780–786 (2001)
16. Teng, H., Kwigizile, V., et al.: Investigation of the AP-42 sampling method. J. Air Waste Manag. Assoc. 58(11), 1422–1433 (2008)

17. Li, H., Vermeirssen, E.L., et al.: Controlled field evaluation of water flow rate effects on sampling polar organic compounds using polar organic chemical integrative samplers. Environ. Toxicol. Chem. 29(11), 2461–2469
18. Rabiet, M., Margoum, C., et al.: Assessing pesticide concentrations and fluxes in the stream of a small vineyard catchment–effect of sampling frequency. Environ. Pollut. 158(3), 737–748
19. Berger, B., von Holst, C.: Pesticide residues in products of plant origin in the European Union. Sampling strategy and results from the co-ordinated EU monitoring programmes in 1996 and 1997. Environ. Sci. Pollut. Res. Int. 8(2), 109–112 (2001)
20. Gaultier, J., Farenhorst, A., et al.: Regional assessment of herbicide sorption and degradation in two sampling years. J. Environ. Qual. 37(5), 1825–1836 (2008)

Hyperspectral Discrimination and Response Characteristics of Stressed Rice Leaves Caused by Rice Leaf Folder

Zhanyu Liu[1,2,*], Jia-an Cheng[2], Wenjiang Huang[3], Cunjun Li[3], Xingang Xu[3],
Xiaodong Ding[1], Jingjing Shi[4], and Bin Zhou[1]

[1] Institute of Remote Sensing and Earth Sciences, Hangzhou Normal University,
Hangzhou 311121, China
[2] Department of Plant Protection, Zhejiang University, Hangzhou 310029, China
[3] Beijing Research Center for Information Technology in Agriculture, Beijing 100097, China
[4] Institute of Agricultural Remote Sensing & Information Technology, Zhejiang University,
Hangzhou 310029, China
`liuzhanyu@zju.edu.cn`

Abstract. Detecting plant health condition plays an important role in controlling disease and insect pest stresses in agricultural crops. In this study, we applied support vector classification machine (SVC) and principal components analysis (PCA) techniques for discriminating and classifying the normal and stressed paddy rice (*Oryza sativa* L.) leaves caused by rice leaf folder (*Cnaphalocrocis medinalis* Guen). The hyperspectral reflectance of paddy rice leaves was measured through the full wavelength range from 350 to 2500nm under the laboratory condition. The hyperspectral response characteristic analysis of rice leaves indicated that the stressed leaves presented a higher reflectance in the visible (430~470 nm, 490~610 nm and 610~680 nm) and one shortwave infrared (2080~2350 nm) region, and a lower reflectance in the near infrared (780~890 nm) and the other shortwave infrared (1580~1750 nm) region than the normal leaves. PCA was performed to obtain the principal components (PCs) derived from the raw and first derivative reflectance (FDR) spectra. The nonlinear support vector classification machine (referred to as *C*-SVC) was employed to differentiate the normal and stressed leaves with the front several PCs as the independent variables of *C*-SVC model. Classification accuracy was evaluated using overall accuracy (OA) and Kappa coefficient. OA of *C*-SVC with PCA derived from both the raw and FDR spectra for the testing dataset were 100%, and the corresponding Kappa coefficients were 1. Our results would suggest that it's capable of discriminating the stressed rice leaves from normal ones using hyperspectral remote sensing data under the laboratory condition.

Keywords: Hyperspectral remote sensing, Rice crop, Rice leaf folder, Principal components analysis (PCA), Support vector classification (SVC).

---

* Corresponding author.

1   Introduction

Detecting plant health condition plays an important role in controlling disease and insect pest stresses in agricultural crops, which always lead to yield loss and poor quality [1]. Although pest damage is predominantly concentrated in patches around original foci, the common protecting practice is still to spray agrochemicals indiscriminately over the entire fields [2]. To minimize economic loss and reduce ecological and environmental pressure due to the superfluous input of agrochemicals, it's necessary to accurately know the pest distribution locations and damage degrees caused by disease and insects [3] so that the most economical amount of agrochemicals can be applied to the infestation patches. It's a pity that the most common method of damage quantification evaluation is still visual survey [4]. However, this technique is subjective, labor-intensive, and time-consuming, and it is impossible to accurately estimate the damaged areas and severity over a wide range [5]. Recently, remote sensing techniques have been proved to be effective to monitor insect infestation and disease epidemic in agricultural crops and other plants from ground to satellite platforms, especially in the site-specific crop management of precision agriculture [6].

The current studies have shown that plant spectral properties at visible and near-infrared (NIR) regions vary with the development of specific signatures for a specific stress of a given plant species [2]. The reflectance or digital number of healthy plants is low in the visible region (400~700 nm) due to strong pigment absorption, is high in the NIR region (750~1300 nm) for cellular and subcellular structure, and is variable in the shortwave infrared (SWIR) region (1300~2500 nm) because of the presence of water vapor, proteins, and other biophysical and biochemical elements [2-3, 5]. In addition to raw reflectance, derivative spectral reflectance and various vegetation indices can also be effective in detecting plant stresses [6-7].

Furthermore, various statistical methods, including cluster analysis [8], decision tree [9], partial-least square regression [10], and principal component analysis (PCA) [2], have been used with various remote sensing data for detecting plant stresses. Among them, PCA is the most widespread method for hyperspectral remote sensing data because it can solve the high dependence and autocorrelation in adjacent wavebands [11]. Moreover, the application of artificial intelligence methods including artificial neural networks (ANN) and support vector machines (SVM) in plant stress detection has also been reported recently [2,12]. The nonlinear support vector classification machine (viz. C-SVC) had been used to detect the rice health condition at the leaf and canopy levels, which obtained above 85% of the classification accuracy [12].

Rice (*Oryza sativa* L.) is the most important staple food in Asia area and there have been a lot of remote sensing-based researches in rice disease and insect pest damage which harm the leaf lamina, leaf sheath, stem, panicles and other organs [2,12]. However, little study was found to discriminate the normal and stressed rice leaves caused by rice leaf folder [13-14]. Rice leaf folder (*Cnaphalocrocis medinalis* Guen) is widely distributed between 48° N and 24° S latitude and 0° E to 172° W longitude in the rice growing areas of Asia, Oceania, Australia and Africa [15]. The feeding damage caused by rice leaf folder shows up characteristically in pale stripe in the leaf where the chlorophyll-bearing tissues have been eaten away, and only epidermis and cuticle remains. All feeding damage is strictly linear, i.e., occurs in the longitudinal axis of the leaf between the veins.

The purposes of the present study are (1) to analyze the hyperspectral response characteristics of raw and first derivative reflectance (FDR) of normal and stressed paddy rice leaves caused by rice leaf folder, (2) to identify the principal components (PCs) derived from both the raw and FDR spectra, and (3) to examine the capability of SVM in combination with PCA for discriminating normal and stressed rice leaves.

2   Methodology

2.1   Study Sites

The samples of normal and stressed rice leaves caused by rice leaf folder (*Cnaphalocrocis medinalis* Guen) were collected in three paddy rice fields at the active tillering stage. One field was located near Yundong village (30°12′N, 120°28′E), and the second in Zhangpanqiao Village (30°06′N, 120°15′E), Xiaoshan District, and the last in the Experiment Farmland of Zhejiang University (30°14′N, 120°10′E), China. The sampling date in the former two paddy fields were on July 27th, 2007 and the sampling number of normal and stressed rice leaves in all was as equal as 19. The sampling date in the latter and last paddy field was on August 14th, 2007 and the sampling number of normal and stressed rice leaves was as equal as 35. The rice cultivar names in the three sites were unknown, but all of them belonged to hybrid and Japonica rice.

2.2   Foliar Hyperspectral Reflectance Measurement

The normal and stressed rice leaves caused by rice leaf folder (*Cnaphalocrocis medinalis* Guen), which were verified by plant protection expert, were cut from the rice individuals in the paddy fields. The samples were put into an incubator (25℃) and were transported to the laboratory for further measurements. The foliar hyperspectral reflectance of paddy rice was measured with a portable spectrophotometer (Analytical Spectral Devices Inc., Boulder, CO, USA) through the full wavelength range from 350 nm to 2500 nm. The spectral resolution of the instrument varied from 3 nm less than 1000 nm to 10nm more than 1000 nm, while the spectra were interpolated by the spectrometer software in 1 nm intervals.

Each leaf was positioned on a dark background. The fiber optic sensor with 8° instantaneous field of view (IFOV) vertically pointed to the upper leaf surface about 3 cm height. An incidence angle of 45° was maintained at a standard distance (50 cm) . The light source used was one 50 watt halogen lamp. For each paddy rice leaf, a sampling spectrum was made consisting of five readings and four sampling spectra were measured. The average hyperspectra derived from four sampling spectra represented each leaf using ViewSpec Pro (version 5.6.10). All spectral measurements were made relative to a BaSO4 standard reference-panel, which was observed immediately before and after the spectral measurements of rice leaves.

2.3   Foliar Pigment Concentration Measurement

Immediately after foliar hyperspectral reflectance measurements have been taken, the leaves were chipped, weighted and then dipped in 20 milliliter solution (acetone:

ethanol: distilled water = 4.5:4.5:1) for 24 hours to extract pigment. The optical density of the extraction solution was measured at 440 nm, 645 nm, and 663 nm by spectrophotometer (Shimadzu UV 2550, Tokyo, Japan). Pigment consists of chlorophyll a (ab. Chl a), Chlorophyll b (ab. Chl b), total chlorophyll (ab. ChlT, which is equal to Chl (a+b)) and Carotenoid (ab. Car) in the study. The formula of pigment concentration (mg/g) was referenced the literature of Tang *et al*. [16]. The statistical information of pigment on rice leaves was shown in Table 1.

2.4   Data Preprocessing

Firstly, the hyperspectral reflectance was smoothed with a seven step moving average to suppress the instrumental and environmental noise before these data were further analyzed [5]. Next, the raw hyperspectral reflectance less than 400 nm and more than 2400 nm were deleted due to severe instrument and system noise. The effective spectral region of FDR ranged from 450 nm to 850 nm. Furthermore, the total dataset was consciously divided into two parts, namely training and testing datasets. The training dataset included the normal and stressed rice leaves caused by rice leaf folder, viz. those leaves which were collected on August 14th, 2007. The testing dataset was consisted of those leaves which were collected on July 27th, 2007.

Derivative spectroscopy is proposed to tackle analogous problems such as background signals and to solve overlapping problems in spectral features. In practice, the simplest method of finding derivatives is by dividing the difference between successive spectral reflectance by the wavelength interval separating them [7]. The FDR spectra can provide the slope information on the rate of change reflectance with respect to wavelength [17]. FDR were calculated to analyze the hyperspectral response characteristics of stressed rice leaves caused by rice leaf folder in this study.

2.5   Analytical Techniques

Principal component analysis (PCA) is capable of preserving the total variance and minimizing the mean square approximate errors, and it is also used to identify the dominant modes of the original dataset such that the first principal component (PC_1) accounts for the maximum possible proportion of the variance information of the original dataset, and subsequent principal components (PC_2, PC_3 …) account for the maximum proportion of the unexplained residual variance, and so forth [17]. The application of PCA has been reported in many studies as an effective data compression technique especially for hyperspectral remote sensing data because it can solve the high dependence and autocorrelation in adjacent wavebands [11]. PCs were calculated to represent the hyperspectral features of rice leaves in this study.

Support vector classification (SVC), which is a popular technique for data classification, is one of the most important parts of support vector machine (SVM). The aim of SVC is to devise a computationally efficient way of learning 'good' separating hyperplanes in a high dimensional feature space based on the generalization and optimization theory [18]. SVC was conducted in LIBSVM (Library for Support Vector Machines) developed by Chang and Lin [19] in our study. The goal of LIBSVM is to help users to easily use SVM as an attached library

for SVM. LIBSVM is consisted of *C*-SVC, *v*-SVC, distribution estimation (one-class SVM), and other models. Thereinto *C*-SVC was adopted in this study. The key parameters of LIBSVM such as kernel function γ and penalty parameter *C* and other parameters are provided with default selections available.

There were two assessment parameters to evaluate the accuracy of C-SVC combing with PCA in this study. One is Kappa coefficient which is one measure of the level of agreement. A Kappa coefficient of 1 denotes a full agreement and 0 denotes full disagreement between the actual and predictive classification [20]. The other is overall accuracy (OA) which denotes the agreement probability between the predictive and actual classification for each random sample.

3   Results

3.1   Hyperspectral Response Characteristics of Paddy Rice Leaves

Applying hyperspectral remote sensing to precise crop integrated pest management (IPM) is on the basis of spectral response characteristics of paddy rice leaves to crop insect stress in Fig.1. Fig.1 showed the hyperspectral difference was obvious between normal and stressed paddy rice leaves caused by rice leaf folder at the nanometer level because there were evident changes in pigment concentration (Table 1), cell and texture arrangement structure and foliar water content due to the feeding damage caused by rice leaf folder [15]. Furthermore, six wavelength intervals (Table 2) had been calculated to simulate the high resolution geometric (HRG) and vegetation (VGT) sensors of Systeme Probatoire d'Observation dela Tarre (SPOT) satellite based on arithmetic mean of hyperspectral reflectance for universal application of SPOT's image in vegetation monitoring all over the world. The simulated six wavelength intervals clearly indicated the hyperspectral difference between normal and stressed paddy rice leaves from visible spectrum region (PA: 460~690 nm, B0: 430~470 nm, B1: 490~610 nm and B2: 610~680 nm) to NIR (B3: 780~890 nm) and SWIR (1580~1750 nm).

As shown in Fig.1 (a), the raw hyperspectral feature of the normal paddy rice leaves was different from the stressed leaves through the full spectra range (400~2400 nm). Because paddy rice foliar pigment concentration decreased obviously (Table 1), hyperspectral reflectance of stressed leaves increased 2.9, 1.9 and 3.3 times than the normal ones in the visible blue (B0), green and near red (B1) and red (B2) light region (Table 2), respectively. However, the hyperspectral reflectance of stressed leaves decreased 20.7% and 9.7% than the normal ones in NIR (B3) and one SWIR (2.08~2.35μm) region, respectively. But in the other SWIR region, the hyperspectral reflectance increased 28.9% in comparison with the normal ones.

The amplitude of FDR for stressed paddy rice leaves was smaller than that of normal ones as shown in Fig.1 (b). Maybe it's because the stressed paddy rice leaves looked like 'white' color after rice leaf folder fed; the hyperspectral reflectance of the stressed paddy rice leaves had lower change rage with respect to wavelengths. Both of the peak wavebands of normal and stressed leaves in the blue light region were located at 521 nm. Same phenomenon occurred in the green light region, the peak waveband (569 nm) of stressed leaves was equal to normal leaves. The red edge position of normal leaves was at 702 nm, but the stressed leaves shifted toward short wavebands about 5 nm.

(a)   (b)
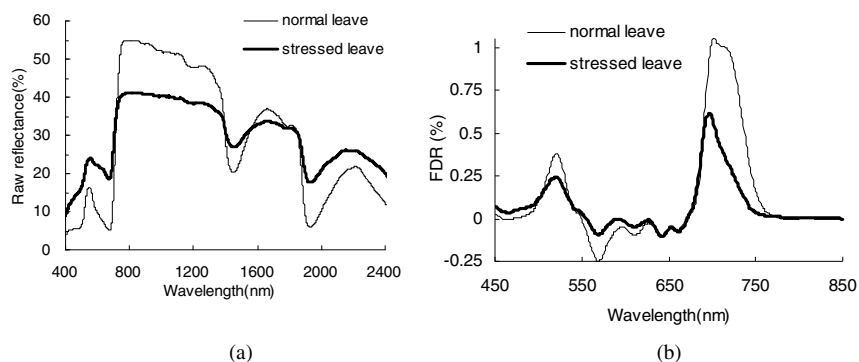
Fig. 1. Average hyperspectral reflectance of rice leaves collected on August 14th, 2007

Table 1. Statistical description of paddy rice foliar pigment concentration (Unit: mg/g)*

Statistical parameters	Normal rice leaves, N=35			Stressed rice leaves, N=35		
	Chl a	Chl b	Car	Chl a	Chl b	Car
Min.	1.63	0.57	0.60	0.88	0.36	0.35
Mean	2.06	0.71	0.76	1.35	0.52	0.54
Max.	2.53	0.83	0.96	1.90	0.73	0.79
Std. dev	0.19	0.06	0.08	0.25	0.10	0.10
VC	9.2%	8.0%	10.1%	18.7%	18.7%	19.4%

* The paddy rice leaves were collected on August 14th, 2007.

Table 2. Average canopy hyperspectral reflectance of rice crops at six wavelength intervals representing the High Resolution Geometric (HRG) and Vegetation (VGT) of SPOT (Unit: %)*

Spectra range	July 27th, 2007		August 14th, 2007	
	Normal(n=19)	Stressed(n=19)	Normal(N=35)	Stressed(n=35)
Pan: 490~690 nm	10.1c	22.6a	9.6c	20.9b
B0: 430~470 nm	5.2c	15.0a	5.5c	13.4b
B1: 490~610 nm	12.3c	22.8a	11.6c	21.4b
B2: 610~680 nm	6.8c	22.3a	6.7c	20.1b
B3: 780~890 nm	57.9a	45.9c	54.7b	41.1d
SWIR: 1580~1750 nm	39.2a	35.4b	35.7b	33.2c

* Reflectance values within a row followed by the same letter are not significantly different from each other by Duncan test at α = 0.05.

3.2   Principal Component Analysis

The first principal component (PC1) explained 72.1% of the variance of the raw hyperspectral reflectance in rice leaves for the training dataset. Fig.2 showed both eigenvector of PC1 and the spectral reflectance difference between the stressed and normal leaves. PC1 had a loading of 0.958 to the reflectance difference of raw spectra between the stressed and normal rice leaves.
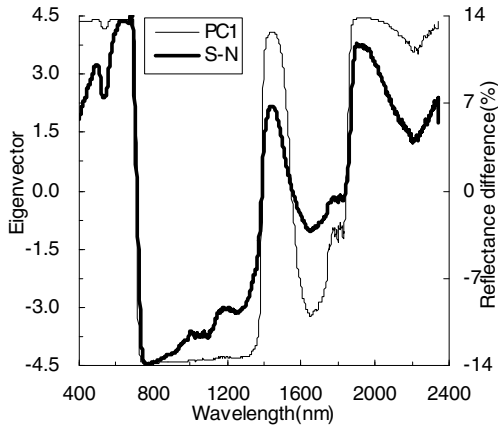
Fig. 2. Loading plot of the first principal component (PC1) derived from the raw hyperspectral reflectance. S-N meant the hyperspectral reflectance difference between the stressed and normal paddy rice leaves.

The results of the principal component analysis (PCA) revealed that the cumulative contribution of the front three components of the raw hyperspectral reflectance and FDR for the training dataset were 98.1% and 66.2%, respectively (Table 3). Generally speaking, when the cumulative contribution of the front several components is above 80~85%, the remaining components always are omitted in further analysis [17]. So the front eighteen components could explain 85.6% of the variance of FDR for training dataset, which were regarded as the principal component spectra (PCs) in this study.

Table 3. Percentage of explained variance for the front three PCs (Unit: %)*

Spectral types	Training dataset (August 14th, 2007)				Testing dataset (July 27th, 2007)			
	PC1	PC2	PC3	CC	PC1	PC2	PC3	C
Raw	72.1	24.6	1.4	98.1	70.4	26.8	1.6	98.8
FDR	47.9	14.1	4.2	66.2	68.4	11.1	7.6	87.2

* CC denotes the cumulative contribution.

3.3   Results of Support Vector Classification Machine

The nonlinear support vector classification machine (referred to as *C*-SVC) was used to automatically detect the health condition of paddy rice leaves. The *C*-SVC model was built to differentiate the stressed rice leaves from normal ones by using the training dataset consisting of dependent (i.e. normal and stressed rice leaf categories) and independent (i.e. PCs) variables for calibration purpose. The *C*-SVC model was run repeatedly with multifarious sets of values of γ (kernel function) and *C* (penalty parameter) at least three times for cross-validation in order that the optimum classification result could be obtained. When γ was RBF (radial basis function) and *C*

was 1, the *C*-SVC model was able to discriminate paddy rice leaves categories without any error. Then the *C*-SVC model was validated by the testing dataset and the classification results were given in Table 4. Not only for the raw hyperspectral reflectance, but also FDR spectra, none of normal or stressed paddy rice leaf was classified mistakenly. The stressed rice leaves had obviously spectral difference from normal ones after the rice leaf folder fed (Fig. 1 and Table 2), which resulted in the perfect classification result by *C*-SVC model integrated with PCA.

Table 4. Error matrix for C-SVC from the testing dataset with PCs derived from raw and FDR spectra

Actual classification	Predictive classification					
	Raw spectra (OA=100%, Kappa coefficient=1)			FDR spectra (OA=100%, Kappa coefficient=1)		
	Normal	Stressed	Total	Normal	Stressed	Total
Normal	19	0	19	19	0	19
Stressed	0	19	19	0	19	19
Total	19	19	38	19	19	38

4   Discussions

In this study, *C*-SVC was employed to detect the health condition of paddy rice leaves by integrating it with PCA using raw hyperspectral reflectance and FDR. The results indicated that the stressed rice leaves caused by rice leaf folder presented a higher reflectance in the visible (430~470 nm, 490~610 nm, 610~680 nm) and one shortwave infrared (2080~2350 nm) region, and a lower reflectance in the NIR (780~890 nm) and the other SWIR (1580~1750 nm) region than the normal leaves (Fig.1a and Table 2). These results were consistent with our previous research [13], and were similar to other's results for geenbug-damaged wheat at the leaf or canopy level [6]. However, our results were not in agreement with the report from Huang *et al*. [14] who concluded that the spectral reflectance of rice leaf decreased in the green (530~570 nm) region with the increase of the damage degrees. Perhaps because in our study the rice leaf folder almost fed on the whole upper surface leaf, there were little pigment concentration in the remaining downer surface leaf than the stressed leaves caused by greenbug [6] or rice leaf folder [14] in others reports.

FDR spectra were also analyzed from a different view. Perhaps because the stressed paddy rice leaves looked like 'white' color after rice leaf folder fed, the amplitude of FDR for stressed leaves was smaller than that of normal ones (Fig.1 b). Not only the stressed paddy rice leaves, but also the normal ones, the peak wavebands of blue and green spectrum region were located at 521 nm and 569 nm, respectively. But the stressed leaves shifted toward short wavebands about 5 nm, which suggested that the red light spectrum region was more sensitive to plant pest damage than the blue and green regions. Similar results occurred in other results [6].

The front several PCs were found to be sensitive to the reflectance and digital number of healthy and nonhealthy plants as well as their linear combinations according to current research [2]. Same results were also found in this study. PC1

could explain 95.8% of variance in the raw reflectance difference between the stressed and normal rice leaves (Fig.2). The front three PCs would reveal above 85% of cumulative contribution except the FDR spectra of training dataset (Table 3). Subsequently, the classification was almost perfect with the combination of *C*-SVC classifier and PCA for discriminating the stressed rice leaves from normal ones with the obtaining PCs (Table 4). Maybe it's because the stressed rice leaves had obviously spectral difference from normal ones after the rice leaf folder fed (Fig. 1 and Table 2).

5   Conclusions

Even though the present study would suggest that it's capable of discriminating the stressed rice leaves from normal ones using hyperspectral remote sensing data under the laboratory condition, this technique is probably still not very practical for detecting the health condition of agricultural crops under the field conditions [6]. More external factors, including atmospheric (e.g., illumination, cloudy shadow), edaphic (e.g., soil type, soil moisture), and biotic (e.g., crop variety, leaf area index) conditions as well as field management strategies, coexist at the canopy level [2], which consequentially bring about too much noise in hyperspectral reflectance data and lower classification accuracy for stressed crops and their organs. Therefore, more studies are needed to explore the spectral response characteristics of rice crops under fungi and insect stresses of different levels in outdoor conditions.

The samples of rice leaves only had two levels, namely, normal and stressed rice leaves caused y rice leaf folder. However, the rate of rolling leaf is employed to determine the damage severity caused by rice leaf folder in practice. Future study should include more stress levels in order to discriminate not only the normal rice leaves from the stressed ones but different stress status.

References

1. Everitt, J.H., Escobar, D.E., Summary, K.R.: Using airborne video, global positioning system, and geographical information system technologies for detecting and mapping citrus blackfly infestations. Southwest Entomol. 19(2), 129–138 (1994)
2. Liu, Z.Y., Wu, H.F., Huang, W.J.: Application of neural networks to discriminate fungal infection levels in rice panicles using hyperspectral reflectance and principal components analysis. Comput. Electron. Agr. 72(2), 99–106 (2010)
3. Pedigo, L.P.: Closing the gap between IPM theory and practice. J. Agr. Entomol. 12(4), 171–181 (1995)

4. Moshou, D., Bravo, C., West, J., Wahlen, S.: Automatic detection of 'yellow rust' in wheat using reflectance measurements and neural networks. Comput. Electron. Agr. 44(3), 173–188 (2004)
5. Kobayashi, T., Kanda, E., Kitada, K.: Detection of rice panicle blast with multispectral radiometer and the potential of using airborne multispectral scanners. Phytopathol. 91(3), 316–323 (2001)
6. Yang, Z., Rao, M.N., Ellitott, N.C., Kindler, S.D.: Differentiating stress induced by greenbug and Russian wheat aphids in wheat using remote sensing. Comput. Electron. Agr. 67(1-2), 64–70 (2009)
7. Demetriades-Shah, T., Steven, M., Clark, J.: High- resolution derivative spectra in remote sensing. Remot. Sens. Environ. 33(1), 55–64 (1990)
8. Li, B., Liu, Z.Y., Huang, J.F.: Hyperspectral identification of rice diseases and pests based on PCA and PNN. Trans. CSAE 25(9), 143–148 (2009)
9. Goel, P.K., Prasher, S.O., Patel, R.M.: Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn. Comput. Electron. Agr. 39, 67–93 (2003)
10. Huang, J.F., Apan, A.: Detection of Sclerotinia rots disease on celery using hyperspectral data and partial least squares regression. J. Spat. Sci. 52(1), 131–144 (2006)
11. Panda, S.S., Hoogenboom, G., Paz, J.: Distinguishing blueberry bushes from mixed vegetation land use using high resolution satellite imagery and geospatial techniques. Comput. Electron. Agr. 67(1-2), 51–58 (2009)
12. Liu, Z.Y.: Monitoring the rice disease and insect stress with remote sensing. Doctor Thesis, pp. 79–87. Zhejiang University, Hangzhou, China (2008)
13. Shi, J.J., Liu, Z.Y., Zhang, L.L.: Hyperspectral recognition of rice damaged by rice leaf roller based on support vector machine. Rice Sci. 23(3), 331–334 (2009)
14. Huang, J.R., Sun, Q.H., Liu, X.D.: Spectral characteristics of rice leaves damaged by rice leaf roller. Sci. Agri. Sin. 43(13), 2679–2687 (2010)
15. Khan, Z.R., Barrion, A.T., Litsinger, J.A.: A bibliography of rice leaf-folders (Lepidoptera: Pyralidae). Insect Sci. Appl. 9(2), 129–174 (1988)
16. Tang, Y.L., Wang, R.C., Huang, J.F.: Relations between red edge characteristics and agronomic parameters of crops. Pedo. 14(4), 467–474 (2004)
17. Holden, H., LeDrew, E.: Spectral discrimination of healthy and non-healthy corals based on cluster analysis, principal component analysis, and derivative spectroscopy. Remot. Sens. Environ. 65(2), 217–224 (1998)
18. Cristianini, N., Shawe-Taylor, J.: An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods, pp. 93–112. Cambridge University Press, London (2000)
19. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. (2001) Software available at, http://www.csie.ntu.edu.tw/~cjlin/libsvm
20. Congalton, R.G.: A review of assessing the accuracy of classification of remotely sensed data. Remot. Sens. Environ. 37(1), 35–46 (1991)

Image Segmentation of Pseudo-foreign Fibers in Cotton on the Basis of Improved Genetic Algorithm

Lulu Ge[1], Daoliang Li[2,*], Liu Yang[1], and Wenzhu Yang[3]

[1] College of Engineering, China Agricultural University, Beijing, China
gelulugelulu@126.com, yangliu@cau.edu.cn
[2] College of Information and Electronic Engineering,
China Agricultural University, Beijing, China
li.daoliang@163.com
[3] College of Mathematics and Computer Science, Hebei University, Baoding, China
wenzhuyang@163.com

Abstract. In the foreign fibers cleaning process, pseudo-foreign fibers are often mistaken for foreign fibers, this result not only seriously affects the detecting precision of foreign fibers cleaning machine, but also doubles the time of cleaning up lint. As for false identification problem of pseudo-foreign fibers in cotton, this paper proposes a new approach for fast segmentation of pseudo-foreign fibers in cotton on the basis of improved genetic algorithm. This improved genetic algorithm reduced the searching range for calculating optimal threshold from 0~255 to 100~220. The calculating speed in this stage was improved more than twice in average. The fitness amendments formula is also proposed to improve genetic algorithm disadvantage, at the same time, this solved issues of "premature", and converging to global optimal solution difficultly in the traditional algorithm. The results show that the algorithm has high speed, accuracy, anti-interference and so on.

Keywords: Cotton, Pseudo-foreign fibers, Genetic algorithm, Image segmentation, Threshold.

1 Introduction

Sorting out foreign fibers in cotton is an urgent solved problem [1]. Foreign fibers have affected the quality of cotton products seriously [2]. Automated visual inspection (AVI) system is a main tool at present for real time foreign fibers detection in lint. The theory of this system is color differences of image processing technology [3-4]. The concept of color difference is based on that impurities of lint cotton and the cotton background have different colors. Hence, the impurities which has similar colors and shapes with foreign fibers were thought as foreign fibers to be removed, which is called as "pseudo-foreign fibers in cotton", also referred to be "pseudo-foreign fibers".

Pseudo-foreign fibers in cotton have various kinds, such as cottonseed, grass blade, boll shell, cotton leaf, cotton stick, weed, stringy cotton, stringy cotton, colored cotton and oiled cotton, etc. [5]. These pseudo-foreign fibers are very similar to the foreign fibers in color, size and shape. However, compared with foreign fibers, pseudo-foreign fibers harm hardly to the cotton into yarn, dyeing, bleaching and other aspects. On the new standard GB1103-2007 of cotton, these pseudo-foreign fibers in cotton are ruled to follow the normal process of impurities to remove, and then determine level according to standards cotton classification [6].

Meanwhile, the traditional methods of image segmentation, such as Otsu method, and the watershed method, are tested to process the pseudo-foreign fibers' image. The histogram information generated by these methods has only one peak, and the peak threshold is more concentrated, and segmentation effect is not satisfactory [7].

Due to false identification by cleaning machine of foreign fibers, pseudo-foreign fibers affected seriously the detection accuracy of the foreign fibers cleaning machine, also affected the efficiency of the foreign fibers cleaning machine. The time of cleaning up lint cotton is thousand times to increase, and also dragging processing progress of lint cotton.

At present, there is not pseudo-foreign fibers' testing equipment in the world. Meanwhile, the pseudo-foreign fibers in cotton make cotton production and processing more difficultly. Hence, China Agricultural University and China Cotton Machinery & Equipment Co., Ltd. are jointly developing detection equipment on line of foreign fibers. Thereinto, Image segmentation is the key of improving the equipment measurement accuracy and reducing false recognition rate, so a suitable image segmentation algorithm of pseudo-foreign fibers determines the latter target recognition and classification of good or bad.

In order to overcome the shortcomings of the traditional methods and segment pseudo-foreign fibers target quickly, this paper puts forward to a new image segmentation method on the basis of improved genetic algorithm.

2   Samples and Methods

2.1   Sample Preparation

Due to the wide of pseudo-foreign fibers is very large, totally 150 kinds, so the typical pseudo-foreign fibers were chosen to do this research which included grass blade, cottonseed, stringy cotton, boll shell, cotton leaf, cotton stick, weed, stringy cotton, ginned dead cotton, waste paper, colored cotton and oiled cotton. Adequate pure cotton lint with no pseudo-foreign fibers was also prepared for making the lint layer. All these pseudo-foreign fibers' samples were provided by China Cotton Machinery & Equipment Co., Ltd.

2.2   Image Acquisition

Colorful images are captured by the color line scan camera of DALSA under high-brightness LED lighting. Shooting frame rate is 30fps @ 4000×500 color images.

Image processing used the Pentium-M processor (CPU), clocked at 2.0GHz, 2GB of memory for the IPC, and software development tools for Visual C++ 6.0.

The steps of image acquisition are as follows. An opening machine is performed to make pseudo-foreign fibers away from the cotton layer as far as possible which is 80 cm wide and 2 mm thick, and then the images are acquired by the DALSA line scan camera. The capture card saves the images following the order of cotton layer passing by.

2.3   Method Selection

There are many image segmentation methods. According to the different treatment strategies, segmentation algorithms can be divided into serial and parallel algorithm. The serial algorithm determines the current point as edge depending on edge detection to get the result of the previous point, but the parallel algorithm only depends on the current point and its neighborhood points. Therefore, the edge detection operator can act on each pixel of the image in parallel algorithm, but the serial algorithm depends on start point and the previous point. That means parallel algorithm's judgments and decisions can be made independently and simultaneously. In the serial algorithm, the results of early treatment can be used by subsequent processing. Therefore, serial segmentation algorithm needs more time to process data and process control is more complicated.

Speed and accuracy of an algorithm are key factors for the online visual inspection system. Hence, parallel algorithm is chosen to process pseudo-foreign fibers' images more suitably. Meanwhile, due to pseudo-foreign fibers have many kinds and colors, and pixels spread are extensive, so all of these specifications are fit for highly adaptability and flexibility of parallel algorithm. Here genetic algorithm is the most appropriate one. Using this algorithm to process threshold, it ensures segmentation's accuracy, and faster speed of this algorithm is more attractive.

2.4   Traditional Genetic Algorithm

Genetic algorithm is a bionic algorithm, the living beings in nature have undergo elimination and evolution continuously from the initial single-celled organisms to multicellular organism, and eventually human beings who have high intelligence and high adaptability of primates [8-9]. The process of evolution totally proved Darwin's "natural selection, survival of the fittest" theory [10- 12]. It told us the living beings who adapt to environment will have chance to breed, and others who can not accustom environment will be eliminated. Though iteration of dozens of generations, the quality of biological communities will improve [13]. Genetic algorithm used this principle to encode the parameter space, also do the steps of selection, crossover and mutation and other operations, then use the fitness function value the to evaluate images, the image information which is evaluated up to standard will be preserved into the next generation, the unqualified image information will be eliminated [14].

In addition, the genetic algorithm has many advantages for image segmentation of pseudo-foreign fibers. It is a probabilistic, iterative global search algorithm for

groupment [15], the search space does not need restrictive assumptions as other image segmentation algorithms, such as: continuous, derivative exists, histogram single peak characteristics, etc. [16]; the searching is not blind, it gains on the best value step by step to finish the "evolution".

2.5 Improved Genetic Algorithm

As the traditional genetic algorithm has appeared "throwback" phenomenon in the process of choosing individual [17], in other words, offspring's fitness function value is less than the parent generation; meanwhile, in order to improve the processing speed of genetic algorithm, this paper did several improvements.

2.5.1 Improvement of Searching Range for Calculating Optimal Threshold

The genetic algorithm is a parallel global optimization algorithm, although the processing speed has obvious advantages than other algorithm, the selection of groupment is random, so it will waste some time. After tested a lot of samples, the distribution of the best segmentation threshold of the pseudo-foreign fibers images is concentrated from 120 to 200. In order to avoid the limitations of the tests and not rule out some special pseudo-foreign fibers which have special color, finally, the optimal threshold distribution is conformed from 100 to 220. This algorithm has reduced the range of searching, greatly decreasing the optimal threshold search time, the average processing time is tens of milliseconds, and the calculating speed in this stage was improved more than twice in average.

2.5.2 Combination of Elitist Strategy and Roulette Wheel Method

As the genetic algorithm is a random selection of individuals, it is possible that the individual whose fitness value is higher has not been selected, but the individual whose fitness value is lower is selected. This is "throwback" phenomenon. According to the principle of bionics analogy, a conclusion can be acquired that the individual whose fitness value is lower may produce the next generation whose fitness is not high, that means these individuals have poor quality. To overcome the drawbacks of these individuals, elitist strategy is led in to choose some individuals who have high fitness value to come into the next generation directly, and other individuals use roulette wheel method to select according to the share of individual fan-shaped area of the roulette selection probability.

2.5.3 Fitness Amendments Formula

In the initial generations of the genetic algorithm may produce some individuals whose fitness value are high, finally these individuals may eventually fill the entire group, which would make the crossover which is used to generate new individuals lose its meaning, even more importantly, the convergence of the algorithm may converge the part optimal value ahead of time, rather than the global optimal value, this is called "population premature" problem. Hence, Goldberg linear stretch formula is fixed to improve the fitness function, which reduces the differences of fitness value between

individuals with higher fitness with other individuals; it can maintain population diversity effectively, balance maximum and average fitness value as the optimal fitness function. The amendments formulas are as follows:

$$H^*(s) = \frac{0.5H_{avg}(s)}{H_{max}(s) - H_{avg}(s)} H(s) + \frac{H_{max}(s) - 1.5H_{avg}(s)}{H_{max}(s) - H_{avg}(s)} H_{avg}(s) \qquad (1)$$

Where $H^*(s)$ is amendments fitness function; $H_{avg}(s)$, $H_{max}(s)$ is average entropy of the histogram for the current generation of individuals, is the maximum entropy of histogram for the current generation of individuals; $H(s)$ is the entropy of histogram for the current generation of individuals. The algorithm model is:

$$S = \arg \max H(s). \qquad (2)$$

$$H(s) = H_A(s) + H_B(s) = \ln P(1-P) + \frac{H_s}{P} + \frac{H_T - H_s}{1-P}. \qquad (3)$$

$$H_T = -\sum_{i=0}^{n-1} p_i \ln p_i. \qquad (4)$$

$$H_S = -\sum_{i=0}^{s} p_i \ln p_i. \qquad (5)$$

$$P = \sum_{i=0}^{s} p_i. \qquad (6)$$

Where S is the segmentation threshold, $H(s)$ is the image histogram entropy which is also the fitness function, $H_A(s), H_B(s)$ is the total entropy for the target A and background B, H_T, H_S is the entropy of all pixels and s the following thresholds (including s), P, p_i is the probability of gray level for the following threshold values of s (including s) and the probability of every gray level. Threshold S which can make $H(s)$ reach the maximum threshold is defined as the optimal threshold for image segmentation, $H(s)$ is also the best fitness value.

2.5.4   Determination of the Optimal Crossover Probability

Crossover is the main method of genetic algorithm to generate new individuals, it is similar to the homologous chromosome synapsis in biological evolution, in other words, a certain bit which comes from an individual's binary code exchanges with another individual corresponding bit to produce a new individual. Crossover

probability value is very important. the value is the larger the better normally, the range between 0.4 to 1. Different targets, different values. If the value is too large, perfect pattern of group activities will be destroyed, the evolution of computing will be interrupted; if the value is too small, the speed of generating a new individual is too slow. Therefore, this paper improved Crossover probability value as follows: in the first 30 generations, the crossover probability is designed as 0.6, each generation has six individuals to executing crossover, other individuals only copied themselves directly into the next generation; in the post 70 generations, the crossover probability is designed as 0.8, because the latter generations' fitness value is relative concentration, this crossover probability can increase group diversity, each generation has eight individuals to executing crossover, and the remaining two individuals simply copied themselves directly into next generation.

2.6   Image Segmentation Steps on the Basis of Improved Genetic Algorithm

The image segmentation steps on the basis of improved genetic algorithm are as follows:

(1) The establishment of the fitness function

 According to formula (2) - (6).

(2) The amendment of the fitness function

 According to formula (1).

(3) The establishment of populations and coding

Several individuals are selected randomly, then they are encoded binary code for the gray value of the individual, each binary code likes a chromosome, and these individuals' chromosomes are as the first generation. After calculating the fitness function value for each corresponding individual, the sum of group fitness value can be calculated, and then the fitness value of every individual also can be obtained, finally following the elite strategy and roulette wheel method, the individuals can be extracted and copied.

(4) Crossover

 Which is followed the new crossover probability proposed in the above.

(5) Mutation

The process modeled the process of organism's gene mutation: a gene on chromosomal locus may mutate as its allele, it usually caused certain phenotype changes. Here this refers to each individual's binary code from 0 to 1 on certain part, or from 1 to 0. For converging to the maximum fitness value as quickly as possible, the mutation probability is set as follows in this paper: the first thirty generations is 0.02; from the thirty to fifty generation is 0.03; the latter fifty generations is 0.02. Then the new individuals are chosen to compose a new generation instead of the original individuals.

(6) Termination conditions

When the algorithm reaches the maximum generations, or after 10 generations the group's highest fitness value has not changed, the algorithm stops.

(7) Image segmentation

The group's highest fitness value is defined as the optimal threshold to perform image segmentation.

3   Results and Analysis

The paper processed twelve kinds of pseudo-foreign fibers' image, and compared segmentation effect and the processing time with the Otsu method and the traditional genetic algorithm. Partial results are shown in Fig. 1-3 and Table 1-2. It can be seen from the results:

(1) This segmentation algorithm can split the pseudo-foreign fibers precisely, and retain a good edge of pseudo-foreign fibers, the segmentation images are clear, which are shown in Fig. 3. As the pseudo-foreign fibers' images distributed un-evenly, target has a wide range of gray scale, by using the Otsu method calculate the optimal threshold, the between-class variance becomes smaller, the probability of faulting the target into the background increases, take Fig. 1(a) as an example, the bottom of the grass blade was mistaken into the white background, that means the pseudo-foreign fibers target are retained half-baked, Fig. 1(b), 1(c) also exist such situation. Traditional genetic algorithm has very nearly the same segmentation effect, and can split basically the pseudo-foreign fibers target, but the overall effect is not as good as the algorithm of this paper, and the effect is shown in Fig. 2.
(2) It can remove a lot of noise, including Gaussian noise produced by the optical components and impulse noise of environment. Three kinds of segmentation me-thods have similar anti-interference capabilities.
(3) Processing time of improved genetic algorithm is only tens of milliseconds, high processing speed. However, the traditional genetic algorithm and Otsu method need relatively long time, which are shown in Table 1.
(4) The optimal segmentation threshold distribution of improved genetic algorithm is between 100 to 220. Specific data is shown in Table 2. Table 2 shows that the op-timal segmentation thresholds which were the following three segmentation me-thods calculated are within a reasonable range, that is the optimal segmentation threshold should not exceed the target ceiling's gray-scale 220; the optimal seg-mentation threshold by using Otsu method obtained is too large or too small, a larger fluctuation range, sometimes closes to the limit, the target image information has been lost, the optimal segmentation threshold by using the genetic algorithm has low threshold range, the target image has also been retained completely.
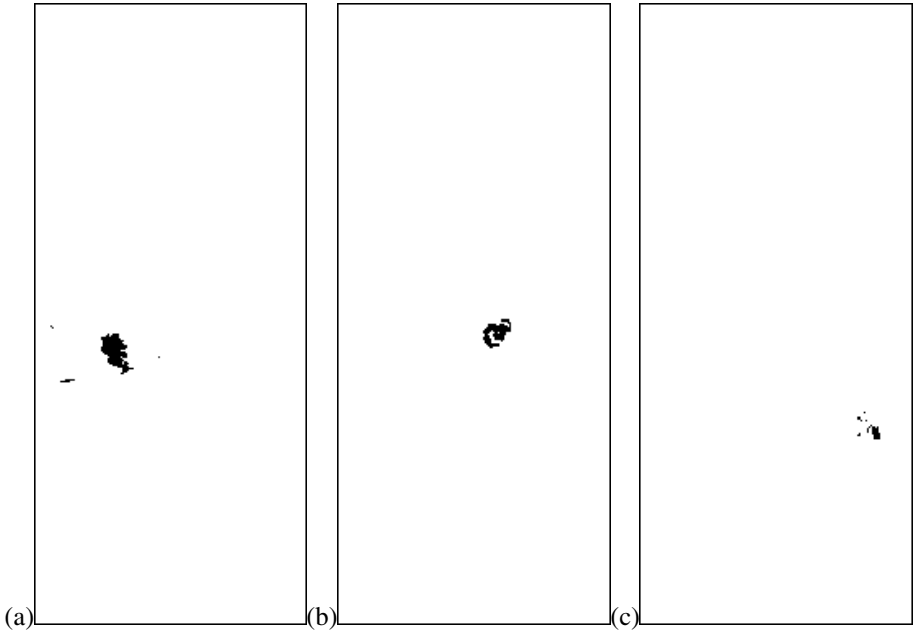
Fig. 1. Segmentation results of the original gray images using the Otsu's method
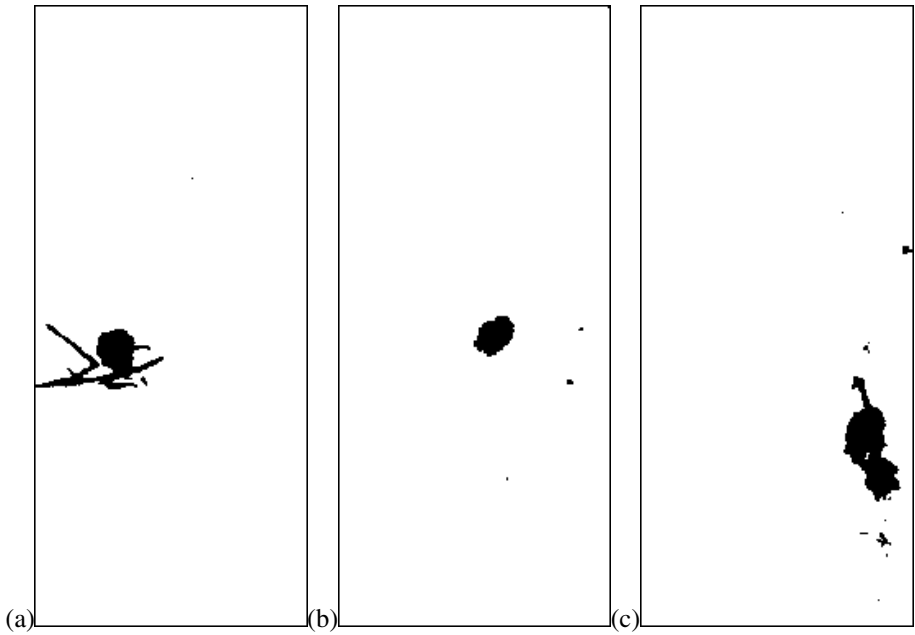
Fig. 2. Segmentation results of the original gray images using the traditional genetic algorithm
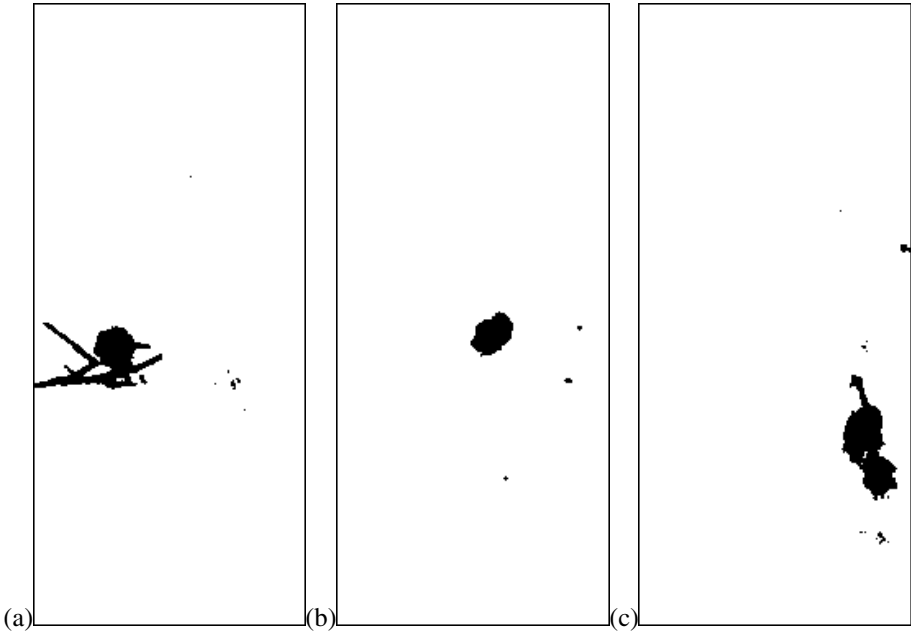
Fig. 3. Segmentation results of the original gray images using the improved genetic algorithm

Table 1. Image segmentation processing time comparison of pseudo-foreign fibers /ms

	grass blade	cottonseed	stringy cotton
Otsu method	2942	406	437
traditional genetic algorithm	31	47	265
improved genetic algorithm	15	43	79

Table 2. The optimal segmentation threshold comparison of pseudo-foreign fibers in cotton

image	the optimal segmentation threshold S		
	Otsu method	traditional genetic algorithm	improved genetic algorithm
grass blade	159	175	184
cottonseed	155	169	174
stringy cotton	200	179	178

4 Conclusion

By using digital imaging technology, this paper has processed the images of pseudo-foreign fibers in cotton, gray-scale characteristics of pseudo-foreign fibers and background are chosen as a starting point, presented a new image segmenting method on the basis of improved genetic algorithm. After the experiment contrast test, the method is verified that it can segment image precisely and clearly, also remove a lot of noise. Meanwhile, it avoided the target segmenting incompletely for a long time and other shortcomings of traditional methods. This new method reduced the searching range for calculating optimal threshold from 0~255 to 100~220. The calculating speed in this stage is improved more than twice in average. This paper proposed the fitness amendments formula, and also solved "premature", converging to global optimal solution difficultly issues of the traditional algorithm; this paper combined elitist strategy and roulette method to improve the algorithm. The results show that the algorithm has high-speed, accuracy, anti-interference and so on.

References

1. Ji, R., Li, D., Chen, L., Yang, W.: Classification and identification of foreign fibers in cotton on the basis of support vector machine. Mathematical and Computer Modelling 51, 1433–1437 (2010)
2. Yang, W., Lu, S., Wang, S., Li, D.: Fast recognition of foreign fibers in cotton lint using machine vision. Mathematical and Computer Modelling 54, 877–882 (2011)
3. Li, D., Yang, W., Wang, S.: Classification of foreign fibers in cotton lint using machine vision and multi-class support vector machine. Computers and Electronics in Agriculture 74, 274–279 (2010)
4. Yang, J., Yuan, J., Wang, Z.: Study on a foreign fiber detecting system with linear CCD. Progress in Textile Science & Technology 6, 60–62 (2009)
5. Hu, B., Hu, B.: Difficulties of machine vision technology in the foreign fiber removed from the cotton. China Cotton Processing 5, 31–33 (2009)
6. General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China. GB-1103-2007, 6:1–16 (2007)
7. Yang, W., Li, D., Wei, X., Kang, Y., Li, F.: Toward Image Seg-mentation of Foreign Fibers in Lint. Transactions of the Chinese Society for Agricultural Machinery 40(3), 156–161 (2009)
8. Lu, X., Li, N., Chen, S.: Two-dimensional thresholding and genetic algorithms in image segmentations. Computer Applications and Software 12, 57–60 (2001)
9. Wang, L., Shen, T.: Two-dimensional entropy method based on genetic algorithm. Journal of Beijing Institute of Technology 11(2), 184–188 (2002)
10. Li, Z., Hou, J.: License plate image segmentation based on modified genetic algorithms. Computer Engineering and Design 26(9), 2455–2457 (2005)

11. Wu, Y., Wang, N., Liu, Y.: Several populations competed genetic algo-rithm and its property analysis. Journal of Northwest Sci-Tech University of Agriculture and Forestry 33(4), 154–156 (2005)
12. Wu, L., Shen, T., Fang, Z., Wang, F.: An image segmentation method using the entropy of histogram and genetic algorithms. Acta armamentarii 20(3), 255–258 (1999)
13. Li, H., Sheng, L., Chen, L., Li, G.: Image Thresholding Seg-mentation Based on 2D Maximum Entropy Principle and Improved Genetic Algorithm. Computer and Modernization 2, 34–37 (2007)
14. Hou, Z., Ma, S., Pei, X., Pan, X.: A Method of Marrow Cell Image Seg-mentation Based on GA. Computer Engineering & Science 28(10), 63–65 (2006)
15. Li, Y., Sun, W., Zhang, Z.: Chromosome Biomedicine Image Processing Based on Genetic Algorithm. Research and Exploration in Laboratory 27(5), 23–25 (2008)
16. Zheng, Y., Zhu, H.: Image Segmentation Approach Based on Improved Ge-netic Algorithm. Journal of Wuhan University of Technology 28(3), 436–438 (2004)
17. Goldberg, D.: Genetic algorithms in search, optimization and Machine Learning, 7 -10, pp. 59–308. Addison -Wesley Publishing Company, USA (1988)

Research and Design of an Agricultural Scientific Instruments Classification and Code Management System

Wenshen Jia[1,2,3], Ligang Pan[2,3], Jihua Wang[2,3,*], Wenfu Wu[1],
Yang Li[2,3], and Yuange Qian[1,2,3]

[1] School of Biological and Agricultural Engineering, Jilin University, Changchun 130022
[2] Beijing Research Center for Agrifood Testing and Farmland monitoring,
Beijing 100097, China
[3] National Engineering Research Center for Information Technology in Agriculture,
Beijing 100097, China
`jiaws09@mails.jlu.edu.cn`

Abstract. As China enlarged the investment in agriculture, the agricultural scientific instruments promoted rapidly. It required us to face the questions scientifically such as the procurement, management, resources sharing, evaluation and etc. in agricultural scientific instruments. Because of the lack of unified classification and code standards, it greatly limited the scientific procurement, efficient management, resources sharing evaluation in agricultural scientific instruments. Therefore, the existing classification and code standards cannot meet the demand also. This paper studies and develops the agricultural scientific instruments classification and the code management system based on B/S (Browser/Server) structure and ASP.*NET* technology. The database used the Microsoft SQL Server, the Server application used C#, the browser based on XHTML + JavaScript. This system on the basis of database, realized the inquiring mode in the fuzzy sense of four kinds of different ways to query data in agricultural scientific instruments. And it also supports the Chinese character automatically converts spell brief-code etc. This system has a significant effect on agricultural science instrument information management and sharing .It helps to improve the assets utilization rate and can regulate the state-owned assets management system also.

Keywords: Instruments, agriculture, information systems, Browser/Server model.

1 Introduction

The agricultural scientific instruments development has four notable features: Firstly, the amount is large. The total industry demand and the total agricultural scientific instruments quantity are huge. It had been increased on agricultural scientific instruments investment in recent five years. The constructive agricultural financial

---

investment funds had invested 60% of the agricultural scientific instruments[1]. Secondly, the range is complete. Modern agriculture is a complex system, covering a complete range of equipment which involved in complex and diverse. It is related to heat, electricity, light, chromatography and other large instrument[2]. Thirdly, the agricultural scientific instruments developed rapidly. According to statistics, the instruments imported about 20 billion totally from 2008 to 2009, of which, the agricultural scientific instruments occupied a large proportion. The agricultural scientific instruments presented a quick development [3]. Fourthly, the feature is distinctive. The agricultural scientific instruments fully embody agricultural characteristics, the instruments and equipments covered a multitude of special features. For example, canola speed measuring device, pesticide residues speed measuring device [4], yellow aspergillus toxin speed measuring device, etc..In Chinese agriculture application, the speed measuring device provides the necessary technical support by preferential agricultural policies. It is a prominent characteristic. Due to the four characteristics of agricultural scientific instrument, it put forward the new age for purchase and management.

This paper will combine the agricultural industry business with modern high-tech information services to realize the agricultural scientific instruments name and classification code fast converting speed, apparatus, the ownership of the classification of quick inquiry. Greatly improved the agricultural scientific researchers work efficiently.

2   Research Status

Because of lacking the unified classification and code standards in agricultural scientific instruments, it greatly limited the agricultural scientific instruments procurement, scientific efficient management, sharing and evaluation. The nation invested vastly in agriculture in recent years. The rapidly investment mainly reflects in the level of the instruments. Since everyone lacks of common standard in agricultural scientific instruments classification and code, and a similar instrument used in the different studies, may name differently. The direct result is each individual experts should have a very high professional judgments in the project review process. Also it brought a series of problems. For example, if the government investment grasp accurate? Is it efficient? Whether embodied the instruments value in the research front-line? It lacks of a unified evaluate judgment. According to the purchasing items of Chinese Academy of Agricultural Sciences from 2006 to 2010, the annual funding for the purchase instruments had more than 150 million RMB, and it had gradually increasing year by year. It also had a lot of difficulties in sorting out, managing and evaluating all of these instruments. A lot of instruments cannot manage in information, sharing the large instruments become impossibly. The lack of consistent classification and coding standards, became the bottleneck in agricultural scientific instruments for purchase, management and sharing [5].

The current classification and code standards cannot meet the agricultural scientific instruments procurement, management and evaluation of demand. In recent years, the national science and technology drafted 'National classification and coding large-scale scientific instruments table'. In the investigation of the application, it for 500000 RMB of above large-scale scientific instruments are classified and coded. In agricultural field application, applying the classification standard, more than 80% of the agricultural scientific instruments are classified in the quality inspection instrument. It can not reflect the characteristics of agriculture. Agricultural research can not reflect the level completely. It should have no value reflect to come out. The classification standard cannot meet the agricultural field. In 1992, the environmental industry drafted environmental protection instrument classification and named standards, but the standard is simple relatively, which of instrument coverage is small. Therefore the standard cannot adapt to the information management needs.

3   System Design and Architecture

3.1   Agricultural Scientific Instruments Classification Method

The system uses the line classification, classification of agricultural scientific instruments, according to instruments type, characteristics and standard. This system is divided into categories of agricultural instruments, categories, sub-category, groups and type of four levels. In the hierarchy, it is divided into the categories of upper class, which divided into the categories of lower class. Directly into a category by the next level out of all kinds of projects, it known as parity class. As a kind of division, it is the same as a benchmark for parallel relationship between the classes. Lower class and upper class categories are subordinate relationship. As a kind of relationship, it sort between not cross, cannot repeat, and only the upper class. Classification in turn, should free layer or additional layers.

The first level instruments of 'category' with the first, second digital code, the second level instruments of 'sub-category' with the third, fourth digital code, the third level instruments of 'group' with the fifth, sixth digital code, the fourth level instruments of 'type' with seventh, eighth digital code.

3.2   Agricultural Scientific Instruments Coding Method

According to the current and to be used in quantity of agricultural scientific instruments, it determined by each level two digital code, a total of eight digital level the code structure. Its classification structure shown Fig. 1.

The system of agricultural scientific instruments is used in line classification, at a higher level on each class (group), it only contains all of the following lower-level classes (groups). The code base is each level differences between the characteristics of the object. It is incompatible features on each level. To lower level code is actually higher levels the code and low levels of the code composite code. The third and fourth layer is no longer sub-classification, in its code up '0' until the eighth. To meet the

agricultural scientific instruments enlargement ductility and coding extensibility principle, each layer classification of all equipped for the item, asylum has not been listed s to code 99.

For agricultural machinery classification method directly used for in 'NY/T 1640-2008 Agricultural machinery classification [6], standard of classification method and the classification number. In this system, the second layer of sub-categories of agricultural machinery with the classification code in the same category code. The third layer group classification of agricultural machinery in the last two class code is consistent. The fourth layer type code and classification of agricultural machinery in the last two items is the same code. Tropical crop of mechanical classification method directly used criteria of 'NY/T 1560-2007 Tropical crops mechanical classification [7], classification method . Code encoding method according to the coding system.


```
XX      XX      XX      XX

                                ────────  The fourth layer (type)

                        ────────────────  The third layer (group)

                ────────────────────────  The second (sub-category)

        ────────────────────────────────  The first layer (category)
```
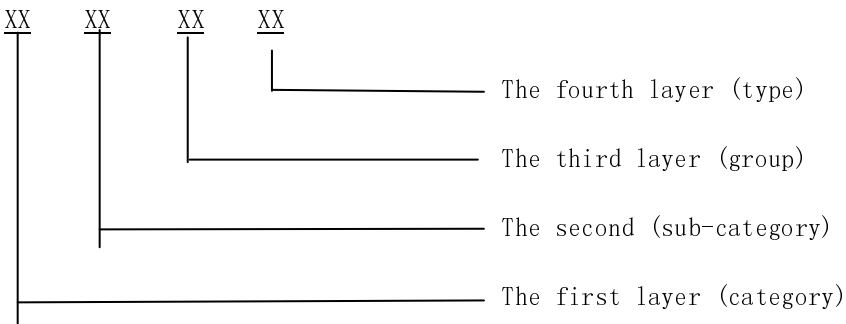
Fig. 1. Coding structure schematic drawing

3.3   System Implementation

System Design. Microsoft.*NET* is to support a new generation of Internet platform. The.*NET* Framework is the foundation of the.*NET* platform, which provides integrated development environment. It uses VB.*NET*, c # or Jscript.*NET* programming language to realize the middle three-layer structure (business logic layer) design, namely using the ASP.*NET* presentation layer to complete the design [8]. The .*NET* development platform application created in Common Language Runtime environment bottom CLR control operation. It is used to load the application, make sure they cannot wrong to perform for the corresponding safety license application, verification, and implementation in operation, finally the completion of the wash them away. Class library provides that applications can be written XML data, on the Internet communications, such as access to database code [9].

In considering the system requirements, the user acceptance and research foundation, after this study adopts the current application extensive Windows Server 2003 +.*NET* + Microsoft SQL Server 2005 system. As for data management and information service network system, and its structure using Browser/Server mode is the best choice. Not only convenient for users to use, but also saving a lot of development and operation of investment. For different needs, it based on the underlying database query detailed information model. In the development process,

highlighting the technical characteristics of humanity, it is from the perspective of user-friendly features and interface design.

System Architecture. Agricultural scientific instruments classification and code management system structure using current applied widely and technology mature Browser/Server mode. If there is Internet access and a browser installed on the computer, this mode can be any user data access and query, which it is the biggest characteristic. Comparing client /server mode to Web page instead of the client software, it is the so-called 'thin client'. Which does not require client software maintenance and upgrades, there is no cross-platform issues. Thereby reducing the burden of agricultural research workers, reducing the cost of data maintenance, but also it makes any changes, effective immediately, to ensure information timeliness, accuracy and completeness [10]. Provide data query, the server program run by the router and the network firewall connected to the Internet, through the access Internet users of computer, login inquires the interface to access and data query. In the function, the information system have three layers of logical structure, including the presentation layer, function and data layer. As a result of a three-layer structure, it is making the deployment, maintenance and upgrades easier and flexible. As shown in Fig. 2. is the main function modules of the system. This system contains query,
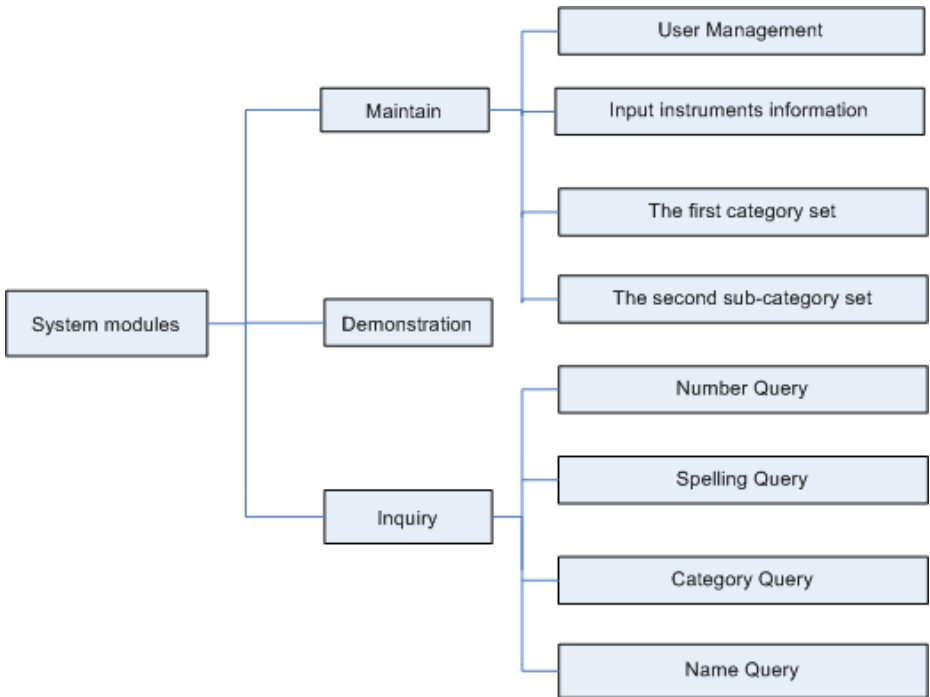
Fig. 2. System Diagram

demonstration and maintenance three function modules, which the query module mainly focus on the serial number query, the spelling query, the category query and the name query. Maintenance module contains the users to access classification management, instrument code entry, instrument code ownership by category and sub-category, etc.

4   Development Operation Environment

4.1   Hardware Environment

An IBM X3650 servers and their associated system of the support system run the hardware environment. The IBM X3650 server have double road Intel Xeon E5606 four nuclear CPU, 2.13 GHz frequency, 4 GB DDRII double channel memory, 146 G hard disk storage system of SAS server, double 1 000 M network adapter.

4.2   Software Environment

Development Platform. In Windows XP operating system environment SP3, it use Microsoft Visual Studio 2005 and AdobeDreamweaver CS3 as a development platform. The system had written the XHTML, JavaScript code for browser and the ASP.*NET* (C#) language for server code.

Release Platform. In Windows Server 2003 Standard Edition operating system environment, it uses IIS as the HTTP server to handle static HTML page views. By the ISAPI server extension called server program, and operation results returned to the client browser in the form of static HTML. Server run in the Microsoft.*NET* Framework 2.0 class library, through the SQL Server Authentication visit Microsoft SQL Server 2005 database Server.

5   Key Algorithm

5.1   Agricultural Scientific Instruments Names Fuzzy Query

Normally it used the SQL statement inquires the text fields, in both sides of the query add '%' wildcard, reoccupy 'LIKE' operator perform a search. However table 1 agricultural scientific instruments classification naming is very complex, and usually the inquires the strategy may produce user input contents leak check or the name of the error types. Therefore this study had developed an agricultural scientific instruments name fuzzed algorithm. The algorithm of the operation process: 1) Pretreatment: It remove the input string without actually meaning characters, such as space, parentheses, '-' and '.' ; 2) Digital processing: It analyze string is involved in Chinese or Arabic Numbers, if have, to the Chinese and Arabic Numbers exchange; 3) Fuzzy treatment: It mask '%' insert into the query string and the end of the characters, thus finish and between the whole fuzzed process. Compared with the

usual inquires the strategy, this algorithm is characterized by be inserted into the wildcard between each character, in order to adapt to all kinds of complicated instrument name. Combined with pretreatment and digital processing, this algorithm can successfully lookup table 1 listed in the various types, which can effectively improve inquires the efficiency, reducing leakage and the chance of error.

Table 1. Agricultural scientific instruments classification code

Classification code	Agricultural scientific instruments
70070205	Meteorological satellite issued equipment ground station
70070206	Meteorological satellite receiving equipment ground station
70070207	Meteorological satellite data collection equipment ground station
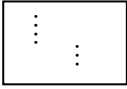
5.2   Chinese Characters Automatically Converts Spell Brief-Code Algorithm

Agricultural scientific instruments classification and code management system database saved thousands of instruments information. The code name translation is one of the important functions. In common Chinese text query string is the input of Chinese characters, but this method requires too many keystrokes. For example, inquires the instrument for 'Meteorological satellite issued equipment ground station' in Chinese, need the Chinese input method input 'qixiangweixingdimianfabuzhanshebei', keystroke at least 34 times, the efficiency is low. The Chinese method of brief-code inquiry can effectively improve the speed of information retrieval equipment. Brief-code is the first Chinese spell brief-code. For example, the same 'Meteorological satellite issued equipment ground station' inquiring record, brief-code inquiring the only need to enter 'qxwxdmfbzsb'. Example shows that the use brief-code inquires to target records quickly, reducing keystrokes, greatly improving the efficiency of query. Then the system uses fuzzy inquiry method. Actually using, you input 'qxw' three letters can quickly lock inquiring the instrument of the relevant information.

The following are Chinese characters automatically converts spell brief-code algorithm.

```
// Create two different encoding object
Encoding unicode = Encoding.Unicode;
// Create GBK code object
Encoding gbk = Encoding.GetEncoding(936);
// Unicode string will be converted to bytes
byte[] unicodeBytes = unicode.GetBytes(UnName);
// Conversion for GBK code
byte[] gbkBytes = Encoding.Convert(unicode, gbk, unicodeBytes);
while (i < gbkBytes.Length)
{
```

```
// If for digital \ letter \ other ASCII characters
if (gbkBytes[i] <= 127)
{
    strResult = strResult + (char)gbkBytes[i];
    i++;
}
//Otherwise the generation of Chinese spell brief-code
else
{
    key = (ushort)(gbkBytes[i] * 256 + gbkBytes[i + 1]);
    if (key >= '\uB0A1' && key <= '\uB0C4')
    {
        strResult = strResult + "A";
    }
    else if (key >= '\uB0C5' && key <= '\uB2C0')
    {
        strResult = strResult + "B";
    }
    else if (key >= '\uB2C1' && key <= '\uB4ED')
    {
        strResult = strResult + "C";
    }
```



```
    else if (key >= '\uD4D1' && key <= '\uD7F9')
    {
        strResult = strResult + "Z";
    }
    else
    {
        strResult = strResult + "?";
    }
    i = i + 2;
}
```

```
    }//end while
    return strResult;
}
```

6   Conclusions

The agricultural scientific instruments classification and code management system will solve the problem of lacking of unified classification and code standards in agricultural scientific instruments. For the convenience of the user's query information and sharing instruments, this study researched and developed the knowledge on the fuzzy query and the Chinese characters. It improved the efficiency of agricultural scientific instruments inquiries, the agricultural scientific instruments' name fuzzy query and Chinese characters automatically converts spelling brief-code algorithm, it also reduced the rate of false query check and the chance of leakage. It is the function of the realization of a certain innovation. The system regulates the flow of agricultural scientific instruments and procedures to share. Thereby it reduced the ambiguity due to the non-standard users information, and it can save the research costs, reduce the unnecessary wear and tear. The current agricultural scientific instruments will be further developed and improved in integrated information systems and the intelligent direction. It will distribute a bar code for each scientific instruments in future, and continue to develop the low-cost hand-held mobile terminals. The bar code scanning equipment can check the information and classification as soon as possible. Thus to reach the aim of sharing the equipment effectively.

References

[1] Fang, S., Zhao, H.: Preliminary Study in Sharing Scientific Instrument and Equipments in China. Technology Management Research 2, 39–41 (2011)
[2] Jiang, H., Xin, L., Cui, J.: Modern agriculture in Chinese the retrospect and prospect. China Venture Capital (2), 32–34 (2011)
[3] Zhuang, S., Xiao, Z.: Development of Science and Technology in Instrumentation, Measurement and Control Area of China. Process Automation Instrumentation (5), 4–9 (2009)
[4] Hui, R.: High intelligence and pesticide residues speed measuring devices to come out. Pesticide Market News (1), 42 (2011)
[5] Xue, F.: Strengthen the Research Funds Management of Agriculture. Journal of Qinghai Normal University(Natural Science Edition) (3), 110–112 (2010)

[6] The agriculture department's agricultural machinery test appraisal station The agriculture department's agricultural machinery maintenance research institute. NY/T 1640-2008. Agricultural machinery classification (2008)

[7] Tropical Chinese academy of agricultural sciences institute of agricultural machinery. NY/T 1560-2007. Tropical crops mechanical classification (2007)

[8] Ferracchiati, F.C., Glynn, J.: NET Data services C# Senior programming. Tsinghua University Press, Beijing (2002)

[9] Rahmel, D.: Use Visual Studio 6 development Web database applications. Tsinghua University Press, Beijing (2002)

[10] Qi, A.: C/S and B/S architecture research. Silicon Valley (22), 55 (2009)

The Self-adaptive Adjustment Method of Clustering Center in Multi-spectral Remote Sensing Image Classification of Land Use

Shujing Wan[1], Chengming Zhang[1,2,*], Jiping Liu[2], Yong Wang[2],
Hui Tian[1], Yong Liang[1], and Jing Chen[3]

[1] School of Information Science and Engineering,
Shandong Agricultural University, Taian, China
[2] Chinese Academy of Surveying and Mapping, Beijing, China
[3] Academy of Shandong BaoLai-LeeLai Bioengineering Co. Ltd

Abstract. As one kind of remote sensing images of land use composed by various categories of surface objects difficult to obtain multi-distribution model of class spectral feature, analyzing the spectral characteristics of LU of multispectral RS imagery, this paper presents a self-adaptive adjustment of clustering center method. Depending on the intercepted situation of the cluster centers between different features to conduct split, the sub-centers obtained are as the sub-category features and the cluster centers assemble to characterize category model which is better to deal with the problems of LU category composed by various surface objects and category model not satisfying multivariate normal distribution. As there are much differences between the many centers features in the unit of category area, so the selection of training area and the determinants of rules are easy. The results of experiment indicate that the LU classification accuracy is increased between 4% and 6% with this method.

Keywords: Multispectral remote sensing imagery, land use classification.

1 Introduction

The LU is classified by the influence degree and the purpose of human activities [1] which refers to the activities of the human imposed on the surface, and the land use categories belong to information categories. The object classification of multi-spectral remote sensing images is based on the spectral characteristics of ground objects, and the objects categories belong to spectral classes. Using multispectral RS imagery for LU classification, a convenient multivariate statistical model is in general not available for the multispectral characteristics of LU class [2-4].

Looking from the existing literature, there are mainly three solutions to solve this problem: firstly, try best to make the samples satisfy the multivariate statistical

---

distribution, such as with the unsupervised clustering method, to choose the dominant and continuous pixels as training samples, then classify with statistical pattern recognition method[4]. Though this solution reduces the differences of samples, the training samples don't contain all ground objects' information. So the difference between classificatory pattern and the ground object is rather great to result in classification error [5, 6].

Secondly, make use of elevations, textures, boundaries and all other secondary information to reduce the dependence on the spectral information, whereas the secondary information add the dimension of the classification feature vector, and the vector multivariate statistical model is more difficult available So we commonly adopt multi-classifier confusion method to classify different types of information, to reduce the classification vector dimensions of single classifier and improving the clustering performance of the classification vector, then fuse the results of several classifiers[7-9]. However, the problem of spectral information statistical distribution is not taken account in these methods, the classification accuracy increases is limited, and even it may lower to some classes [10].

Thirdly, take the un-statistical pattern recognition method to solve multivariate statistical distribution, such as decision tree method. This method makes use of the class subset to form the boundary of branch decision tree [11, 12], but it is difficult to divide subset and to form rules for the method.

Analyzing the multi-spectral feature of LU in RS image, this paper presents a new method, Region Multi-Center (RMC), for LU classification of multi-spectral RS image. In this method, the classification cell is region, classificatory pattern is formed by clustering centers of training samples, and adopt rule-based method to classify.

2   Multi-Spectral Remote Sensing Image Analyses

2.1   Distribution of Within-Class Distance

The theory basis of land use remote sensing image classification is: in the same condition, the same objects should be able to show some internal similarities of theirs, and performing in remote sensing image, the pixels formed by same objects should have the same or similar spectrum information features and space information features. The pixels characteristic vectors which represent the same objects of will distribute in the same feature space area. In fact, although with existing of same objects in different spectral the same kind objects may have greater diversity in the spectrum, that a small class always presents similar spectral features.

Therefore, it is necessary to study the characteristic vector and category of the distribution center distance vector in order to construct a new classifier. In order to give facilities to analysis, confirmed a distance interval analysis unit, first as a threshold to distance of the sample clustering and calculating the clustering center to category center distance, obviously they are consistent, and more meet statistical conditions.

Suppose c class pattern set composed of land-use type $\{x_1, x_2, \ldots\ldots, x_N\}$, the pattern set of class ω_j is $\{x_i^{(j)}; j=1,2\ldots,c; i=1,2,\ldots,n_j\}$, among this superscripts j indicates category，subscripts i indicates the sample model number within the class，n_j is the pattern number of ω_j set, $\sum\limits_{j=1}^{c} nj = N$, ω_j class category center vector with mean vector ith the model indicated，that：$m_j = 1/n_j \sum\limits_{i=1}^{j} xi(j)$，j=1,2, …, c, to ω_j mode of class sample, clustering it by a certain distance threshold，form k Within-class centers（fig.1），they are$\{\omega_{j1}, \omega_{j2}, \ldots, \omega_{jl}\}$,
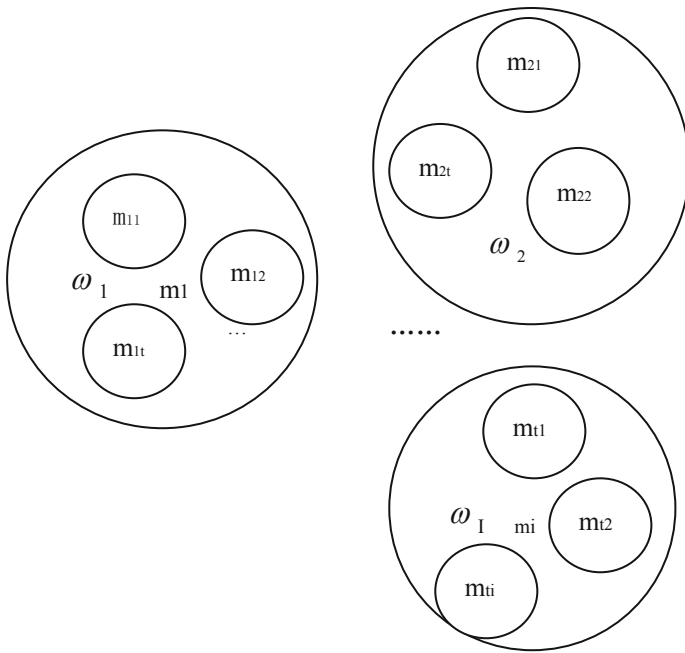
Fig. 1. Category and within-class center schematic

l=1,2,…,k, form the pattern set of within-class center ω_{jl} is $\{x_i^{(jl)}; j=1,2,\ldots,c; l=1,2,\ldots,n_l\}$, ω_{jl} indicates the first within-class center of ω_{j}, n_l is the number of pattern set, $\sum\limits_{l=1}^{k} nl = n_j$, the of mean vectors of within-class center are $\{m_{j1}, m_{j2}, \ldots, m_{jl}\}$, l=1,2, …,k,

$m_{jl} = 1/n_l \sum\limits_{i=1}^{ni} xi(jl)$ ，j=1,2, …, c ; l=1,2，…, k, the distance of within-class is

$d_{jl}=\|m_{jl}-m_j\|$,   mapping with the ordinate of n_l and n_j's percentage and   the abscissa of d_{jl}, called the within-class distance distributions.

2.2   The Analysis of the Land Use Classification of Multi-spectral

The experiments use the area of Shandong province Laiwu city Xueye reservoir with the ETM + images in spatial resolution of 30m, a total of eight bands, the size of 1500 x 1200 pixel and the time is on May 31, 2007.

   The experiments choose samples to analyze the infra-class distance distribution of multi-spectrum in four kinds of land use category as water, residential areas, forest land and bare soil. The rule of selecting samples is that samples should cover as many kinds of pixels in ground objects as possible. The samples should be more and the distribution range of samples should be wide to comply with the above principle. According to the rule, each class respectively chooses the sample 16384 pixels. Fig. 2 shows the rough region of the choosing samples.
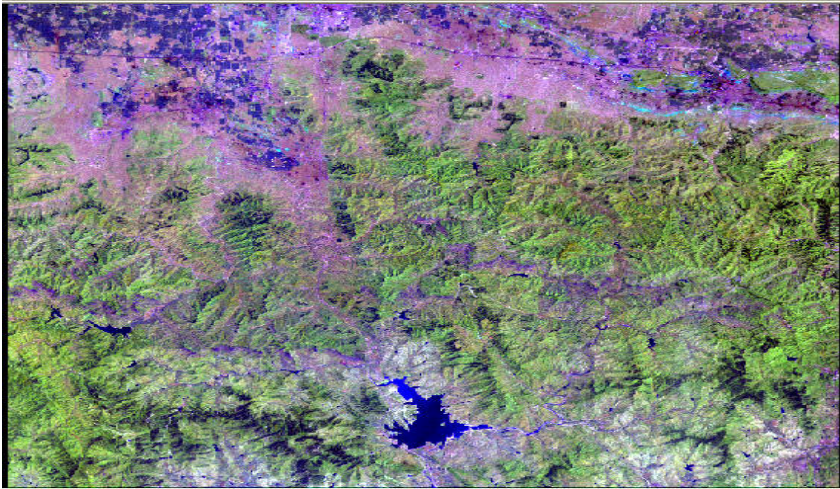
Fig. 2. The remote image of Xueye Reservoir in experiment

   Fig. 2 shows the graph of infra-class distance distributing of LU class, and the distance threshold of max-minimum clustering is 30. From Fig. 2, we can know: (1) The spectral value of pixels in every class has many infra-class centers. (2) Each class exists several relatively concentrated distance intervals, such as 30-40 and 60-70 of water. It represents that one class of LU is composed of several mainly ground material. (3) Spatial distance distribution of water is the most concentrated, and the distribution of residential area is more concentrated than that of woodland and plantation land, but the continuity of urban land is less than woodland and plantation land.
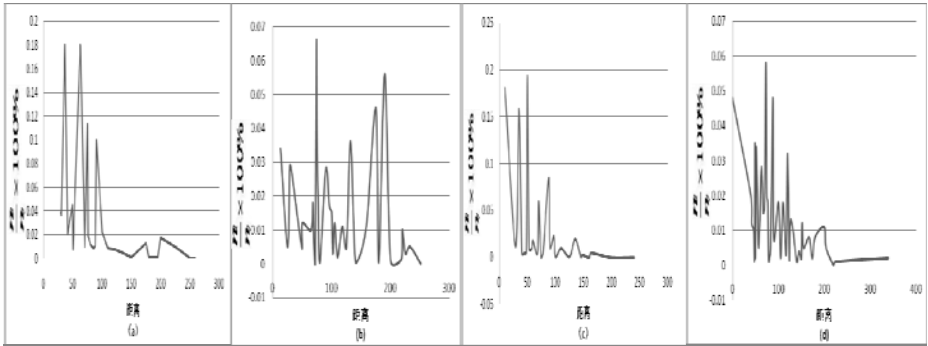
(a) Water   (b) Woodland   (c) bare soil   (d) residential areas

Fig. 3. Graph of special distance distributing of

3   Adaptive Method of Cluster Centers

3.1   The Accuracy Analysis of the Effects of the Phenomenon of the " Same Objects in Different Spectral"

In essence, the " same objects in different spectral " phenomenon is due to in a variety of reasons the value deviation of similar features in some bands or all bands is too large resulting that feature vectors of the class in the feature space distribute excessively decentralization and it is difficult to attribute to the same cluster as we expect. Although it can reach the goals mandatorily by changing the cluster center vectors and the effective radius vectors that it may result in that classify the pixels which not belong to the class to assign to which may reduce the classification accuracy. To better illustrate this problem, we take a classification of the two-band image and for example, assume that the image will be divided into A, B two classes that the distribution of each pixel is as Figure 4.
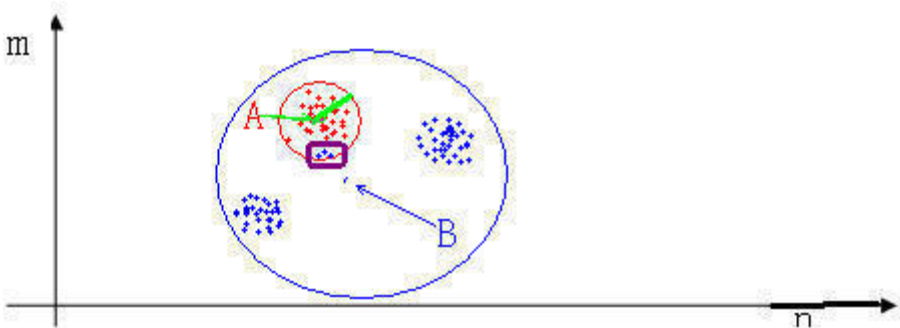
Fig. 4. The schematic of the wrong classifications result

In Figure 4, as the pixels of the class B are distributed in the two regions, the resulting cluster centers and the distance between the two regions is too large, resulting in identification range (ie, blue circle) is too large. As a more extreme case, but another class A area was surrounded in the identification of the B range, surrounded by an orange rectangle should be divided into Class A of the pixel, is divided into Class B wrong. To solve this problem, cluster centers can be used to split ways. After the end of the study sample, the intersection of the cluster center to determine if the two cluster centers according to the radius of the resulting range of the identification part of the intersection, then the cluster centers need to be split.

3.2   Adaptive Step

Adaptive cluster center is split by continued implementation of cluster centers, concrete steps to split is :

(1)During the study of sample, the same category in the normal way to generate cluster centers, cluster centers in the sample vector and the maximum distance as the radius of the cluster centers identified.

(2)Calculate the result cluster centers two by two to determine whether the two cluster centers intersects , in another words, two cluster centers are identified areas of overlap. If they intersected, treat as (3).

(3)Split a larger range between the two cluster centers. Class from the sample to be split to choose one sample, randomly select a distance less than the original value of the largest identified as the threshold, the sample is divided into different subsets, separately for each study, resulting in the respective cluster centers and not split the class with the intersection of the judge, if there is still a subset of the cluster centers with which they have intersected, then re-elect a smaller threshold, re-split. After the split, the original sample a large set of complete small sample to be broken down into multiple sets, as a sub-category treatment. In this way, each "Spectrum with different material"of the surface features can be divided into several sub-categories, each sub-class corresponds to a pattern. The same sub-classes of spectral characteristics are similar.

4   Experiment

Experimental data is shown in figure 2.First, select the body of water, wood land and bare soil this three kinds of quite different categories of land use classification experiment. Various methods using the same training sample selection and accuracy evaluation. Training sample selection principle: the sample should cover as much as possible in a variety of land-use categories pixel surface features as much as possible with the visual way to select a known area of the pixel; accuracy assessment methods: random image classification 100 points, based on visual interpretation of the results as statistical classification error.

4.1   Common Method of Classification
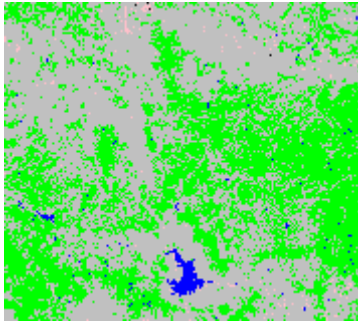
(a) Parallelepiped method
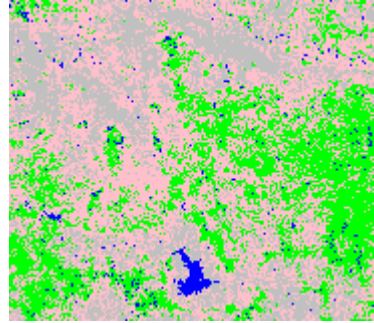
(b) Minimum distance method
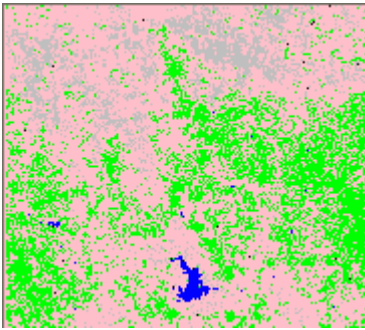
(c) Maximum likelihood
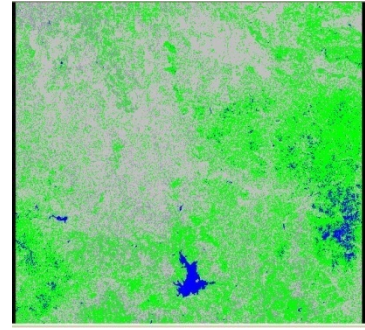
(d) K Neighbor method (k=19)

(e) BP Neural network

water   bare soil   woodland   be classified
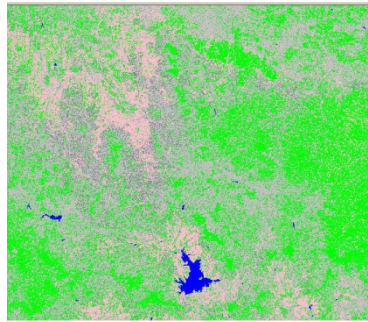
Fig. 5. The results of classification

Looking at the results from the classification point of view, parallelepiped classification method is often a part of the misclassification phenomenon, mainly due to the similarity measure threshold T impact. Through investigation and interpretation,

the method divides a number of unclassified woodland into forest land and bare soil incorrectly. Using minimum distance classification method, the number of pixels to be classified significantly increased, to be classified and sub-phenomenon of mixed woodland is more serious, there is water affected by the shadow of the phenomenon of misclassification. K neighbor method can be a good way to distinguish watershed land and naked body while could not recognize forest, some forest land is divided into body of water mistakenly.

Maximum likelihood method, while classifies water bodies better would make serious mistake in the classification of the forest and bare soil. K-nearest neighbor method can be a good way to distinguish watershed land and naked body while could not recognize forest, some forest land is divided into body of water mistakenly. BP neural network method has better classification of various types of effects, but there is still some vegetation types not distinguished.

Table 1. Accuracy of common classification method (%)

Classification method	water	forest	town	Arable land	bare field	Average
Maximum likelihood	85.6	84.1	76.3	82.9	86.5	83.03
Minimum distance method	87.7	86.8	72.4	85.6	85.3	83.03
K Neighbor method	90.2	85.7	74.6	84.2	85.1	83.9
Neural network	87.0	85.2	81.7	84.1	85.9	84.78

4.2   Adaptive Process of Cluster Centers and the Results of Classification

Using the cluster center adaptive classification method, training samples through the formation of water bodies, residential areas, woodland, bare soil, respectively, the optimal number of model centers 50, 20, 50 and 10, the experimental results shown in Figure 6.

The classification accuracy of different models and types of training samples that shown in Table 2.

Table 2. Classification accuracy of different models and types of training samples

water	forest	town	Arable land	unclassified	Average
88.5	87.8	86.8	89.3	86.2	87.72
88.2	87.6	87.0	88.9	85.3	87.40
87.9	88.3	86.4	89.2	86.0	87.56
86.4	88.5	88.7	89.0	84.1	87.34

water ▮ bare soil ▮ rock ▮ woodland ▮ Residents ▮ lawn ▮ Arable land
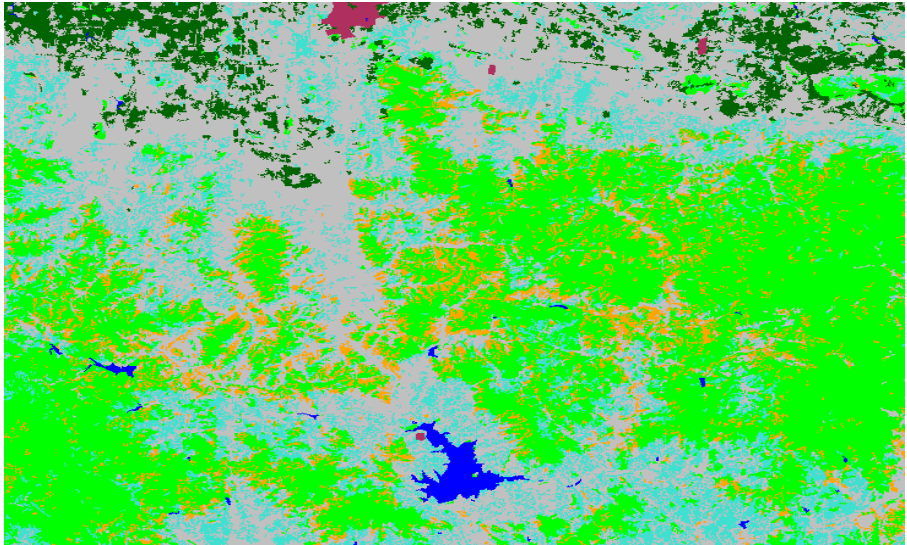
Fig. 6. The extraction results

Comparing different methods of classification results, we can see, the use of this method of classification, the results of the spots decreased; Table 7 Comparison Table 6 Classification accuracy increased by 4 percent.

5   Conclusions

This paper presents a RMC method for LU classification of multispectral RS imagery. The classification cell is region, and the classificatory pattern is represented by infra-class center set of the training samples. The method fully considers the characteristics that a kind of LU class is covered with several ground objects, and the classificatory pattern does not meet multivariate statistical distribution. To RMC method, the multispectral feature is more dispersal, the classification accuracy is improved greatly. There are two classification rules of RMC method, one is the type amount of classificatory infra-class centers which the pixel of the region cell belonged to, the other is the percentage of the pixels belonged to the class from the whole region cell pixels. The rules express sufficiently the basic idea of LU classification system and classification based on the area ratio of ground objects to whole region. Meanwhile, it fuses the region information and the spectral information. Owing to the large difference of RMC feature, it is simple to establish the classification rules and to select training samples.

RMC method which multi-centers confirmed by clustering solve preferably the problems of the division of subset and the confirmation of boundary of decision tree method. Furthermore, the method can be applied to single class recognition and can adopt different classification rules for different classes. RMC method can increase the LU classification accuracy of multispectral RS imagery between 4% and 6%.

References

1. Cingolania, A.M., Renisona, D.: Mapping vegetation in a heterogeneous mountain rangeland using landsat data: an alternative method to define and classify land-cover units. J. Remote Sensing of Environment, 84–97 (2004)
2. Lin, J.: Region multi-center method for land use classification of multispectral RS imagery. Journal of Remote Sensing, 165–172 (2010)
3. Chen, X.: Study on Combined Classifier Based on Error Analysis. Journal of Remote Sensing, 683–691 (2008)
4. Dong, M., Zhu, H.: The Calculation Method of Inverse Camera Response Function of a Single Image based on Kernel Estimation. Computer Engineering and Applications, 1–6 (2010)
5. Chen, J.: Rough set theory based object-oriented classification of high resolution remotely sensed imagery. Journal of Remote Sensing, 1139–1146 (2010)
6. Liu, X., Li, X.: Classification of Remote Sensing Images based on Ant Colony Optimization. Journal of Remote Sensing, 253–262 (2008)
7. Hur, J., Kim, J.: A Hybrid Classification Method Using Error Pattern Modeling. J. Expert Systems with Applications, 231–241 (2008)
8. Lucas, R., Rowlands, A.: Rule-based classification of multi-temporal satellite imagery for habitat and agricultural land cover mapping.   ISPRS Journal of Photogrammetry & Remote Sensing, 165–185 (2007)
9. Mao, D.: Remote sensing image classification based on formal concept analysis. Journal of Remote Sensing, 90–103 (2010)
10. Bian, F.L., Wan, Y.: A novel spatial co-location pattern mining algorithm based on k-Nearest feature relationship. J. Geomatics and Information Science of Wuhan University, 331–334 (2009)
11. Cheng, Q.: Remote sensing image classification method supported by spatial adjacency. Journal of Remote Sensing, 88–99 (2011)
12. Dan, H.: An improved K –means cluster algorithm. Journal of Chongqing Technology and Business University (Natural Science Edition), 144–147 (2009)

A Review of Measurement Methods of Dissolved Oxygen in Water

Haijiang Tai[1], Yuting Yang[1,*], Shuangyin Liu[1,2], and Daoliang Li[1,**]

[1] College of Information and Electrical Engineering, China Agricultural University,
Beijing 100083, P.R. China
[2] College of Information, Guangdong Ocean University,
Zhanjiang Guangdong 524025, P.R. China

Abstract. Some kinds of checking methods and principle of dissolved oxygen in water were summarized. Such as: iodometric method, current determination method (Clark dissolved oxygen electrode), conductance measurement and fluorescence quenching. The advantages and disadvantages of each method were compared, and fluorescence quenching was discussed. The method uses Ruthenium complex as fluorescence sensitive reagent, which emits fluorescence under the excited light. The quenching accords with Stem-Volmer formula, and the density of oxygen could be deduced by checking fluorescence spectrum.

Keywords: Dissolved oxygen, Fluorescence quenching, Environmental monitoring.

1   Introduction

With the rapid development of industry and agriculture, a large number of industrial waste water and agricultural drain is discharged to rivers and lakes. At the same time, about 80% of sewage of city life is discharged directly without handled, towns and rural sewage is discharged in disorder state. As a result, water quality of many locals deteriorating, water pollution and the shortage of water resources increasing seriously. The need of real-time monitoring and effective treatment of wastewater is urgent, and dissolved oxygen content in water quality monitoring is an important indicator.

Dissolved Oxygen (DO) refers to oxygen dissolve in water molecules state, DO is vital for all higher aquatic life (B.A. Cox, 2003; M.A. Best, 2007). A source of DO is atmospheric oxygen infiltrate water body when dissolved oxygen in water is not saturated; another source is released by water plant through photosynthesis. DO is changes as temperature, pressure, salinity changes, generally speaking, the higher of temperature and dissolved salts, the lower of DO content in water; the higher of air pressure, the higher of DO content in water.

---

[*] Yuting Yang, contributed equality with Haijiang Tai, and they are co-first author.
[**] Corresponding author.

DO is consumed by reducibility material, such as sulfide, nitrite and ferrous material. Besides, breathing of microbes in the water and oxidative decomposition of aerobic microbes to organic matter can also consume DO. Consequently, DO is the capital of water, and the indication of water self-purification. DO is approximate to saturated value (9ppm) in natural water, DO content reduce when algae breeding exuberant.

DO content reduce as a result of water polluted by organic matter and reducing material. As to aquaculture, DO content is vital for all aquatic life, such as fish. When DO concentrations are under 4mg/L, fish will be dying by suffocation. As to human, DO content could not be less than 6mg/L. When the rate of Oxygen dissolves into water is lower than DO consumption rate, DO content will be reduce to 0 mg/L, anaerobic bacteria breed and water quality will be worse, so DO content can reflect the contaminated level of water, DO content is an important index of contaminated level of water and also a overall indicator of measuring water quality. Therefore, measurement of DO content has important significance to environmental monitoring, and the development of aquaculture industry.

2   Methods and Principle of DO Measurement

2.1   Iodometric Method (GB7489-87)

Iodometric method (equivalent to international standard ISO 5813-1983) is the benchmark of DO measurement method, it is the first method of DO measurement by using chemical test method, and measurement accuracy is high (Irja Helm, 2009). Its principle is added manganese sulfate and alkaline potassium iodide in water samples, and produces manganic hydroxide. The nature of manganic hydroxide is unstable extremely at the time; it combined with DO quickly, and produces manganese manganic acid:

$$4MnSO_4 + 8NaOH = 4MN(OH)_2 \downarrow +4Na_2SO_4 \qquad (1)$$

$$2Mn(OH)_2 + O_2 = 2H_2MnO_3 \qquad (2)$$

$$2H_2MnO_3 + 2Mn(OH)_3 = 2MnMnO_3 \downarrow +4H_2O \qquad (3)$$
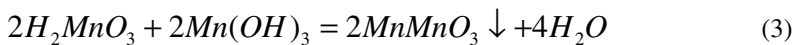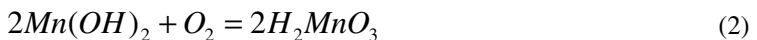
Add in concentrated sulfuric acid and has combined dissolved oxygen (existence in the form of MnMnO3) react with potassium iodide added in solution, and separate iodine.

$$4KI + 2H_2SO_4 = 4HI + 2K_2SO_4 \qquad (4)$$

$$2MnMnO_3 \downarrow +4H_2SO_4 + HI = 4MnSO_4 + 2I_2 + 6H_2O \qquad (5)$$
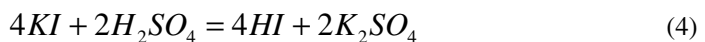
Take starch as indicator, and titrate iodine released with sodium thiosulfate, to calculate the DO content, chemical formulas as following:

$$2Na_2S_2O_3 + I_2 = Na_2S_4O_6 + 4NaI \tag{6}$$
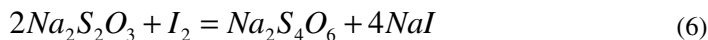
Where V , is the amount used of $Na_2S_2O_3$ solution, M is the concentration of $Na_2S_2O_3$, a is the volume of water samples used in titrating. DO could be calculated as following:

$$DO(mol/L) = \frac{V \cdot M}{a} \tag{7}$$

Iodometric method is applicable to clean water, such as source water and surface water. Iodometric method is a kind of traditional measurement method of DO, it has high measurement accuracy, and the measurement uncertainty is 0.19mg/L. But the method is a kind of pure chemical testing method, time consuming and program trivial, cannot satisfy the online measurement requirements. At the same time, the organic matter easy oxidizing, such as tannic acid, humic acid and xylogen, can lead to disturbance with measurement results.

2.2   Current Determination Method (Clark OXYSENS)

Current determination method measure DO content in water according to diffusion rate of molecular oxygen spread through the film This kind of film can only through gas and generally use the polytetrafluoroethylene film. Oxygen through the gas diffusion to electrolyte, reduction reactions occur immediately in the cathode (positive), oxidation Reaction occur immediately in the anode (negative), the current produced is proportional to oxygen concentration, by means of measuring the size of the current can get dissolved oxygen (DO) concentrations.

Oxygen is consumed at the cathode according to the reaction (Yongxia Zhang, 2003):

$$O_2 + 2H_2O + 4e^- \rightarrow 4OH^- \tag{8}$$

In response to the production of hydroxyl ions (OH^-) at the cathode, and to preserve the charge balance of the electrolyte (saturated KCl) solution, chloride ions react with silver at the anode according to the reaction:

$$Ag + Cl^- \rightarrow AgCl + e^- \tag{9}$$

Therefore, the chloride ions in the electrolyte solution function as a carrier of the electric charges.

Galvanic electrodes use silver or platinum as the cathode and lead, iron, or zinc as the anode:

$$\text{cathode}: O_2 + 2H_2O + 4e^- \rightarrow 4OH^- \tag{10}$$

$$\text{anode}: 2Pb + 4OH^- \rightarrow 2PbO + 2H_2O + 4e^- \tag{11}$$
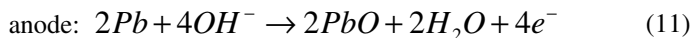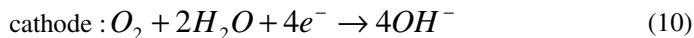
The measurement speed of current determination method is fast than iodometric method, easy operation and less disturbance, and can realize automatic and continuous detection. However, its oxygenater membrane and electrode easy ageing, when water samples containing the algae, sulfide, carbonate, oil and other material, oxygenater membrane will be jam or damage, need to pay attention to protect and change in time. Besides, it determines the concentration of oxygen relies on the electrode occur oxidation reduction reactions under the action of oxygen. Currently, instruments obtainable in the market belong to the type of Clark electrode mostly, activated at set intervals and oxygenater membrane should be changed regularly. Representative products is a series of portable dissolved oxygen meter of YSI company of United States, this device used widely in our country, such as YSI58 Dissolved oxygen meter, it could complete the test work of laboratory and field effectively, easy to carry and operate, measurement range   is 0-20 mg/L, accuracy is ±0.03mg/L.

2.3   Fluorescence Quenching

Fluorescence quenching refers to any process which decreases the fluorescence intensity of a molecule. A variety of processes can cause fluorescence quenching, such as excited state reactions, energy transfer, complex formation and molecular collision (Lakowicz, 1999). This phenomenon may pose challenges for the implementation of fluorescence spectroscopy for water quality monitoring method due to the complex and variable water matrices that exists within water recycling systems and distribution networks. The potential influence on fluorescence measurements from temperature variations, pH, metal ions and oxidation processes need to be considered (R.K. Henderson, 2009).

From the beginning of the 80s, people have begun to explore the work of applied to oxygen probe of fluorescent indicator. Four alkyl amino ethylene has been adopted for chemical fluorescer, it fall into disuse for its attenuate gradually in 12 hours in application of response with oxygen. Pyrene, Pyrene butyrate and Fluorine anthraquinone are good indicator of oxygen.

Wolfbeis had reported a fluorescent sensor response to oxygen fast in 1984, took pyrene butyrate fixed in porous glass as indicator. The advantage of the sensor is response fast (low in 50 ms), and have a good stability. Philip had fixed three fluorescence indicators of coumarin 1, coumarin 103 and coumarin 153 in XAD-4, XAD-8 and silica gel respectively, make comparisons in sensitivity, launch of strength and stability, coumarin 103 fixed in XAD-4 is the best option as interface in fiber optic oxygen sensor. Fiber optic oxygen sensor takes it as indicator has the application scope widely. Later, organic compounds of transition metal (Ru, Os, Re, Rh and Ir) got attention for its special performance. In generally, selecting ruthenium complex as

indicator. Fluorescence intensity of ruthenium complexes exist one-to-one correspondence with the partial pressure of oxygen. Excited state of ruthenium complexes has long life, no need oxygen; its chemical composition is very stable, and insoluble in water. During the process of metal ligand charge transfer (MLCT) from ground state to excited state of Ruthenium complexes of chromium, the nature of excited state is closely related with the ligand structure, usually with the ligand conjugated system increases, the fluorescence intensity and fluorescence life expectancy increases.

The most common method of optically detecting oxygen has been through the measurement of the fluorescent intensity of suitable fluorophores, the intensity decreasing with increasing oxygen concentration (Lee, S., Shin, 2001; Choi, M.M.F., 1999).

The oxygen quenching process is described by the Stern-Volmer equations:

$$\frac{I_0}{I} = 1 + K_{SV}\, pO_2 \tag{12}$$

$$\frac{\tau_0}{\tau} = 1 + K_{SV}\, pO_2 \tag{13}$$

$$K_{SV} = k\tau_0 \tag{14}$$

Where I and τ are, respectively, the fluorescence intensity and excited state lifetime of the fluorophores, the subscript 0 denotes the absence of oxygen, K_{SV} is the Stern-Volmer constant, k is the diffusion-dependent bimolecular quenching constant, and pO_2 is the oxygen partial pressure. According to the actual measurement of fluorescence intensity, I_0, I and K_{SV}, DO concentration can be calculated.

Much has been published on optical oxygen sensors that measure the intensity of phosphorescence. It has now been established that these sensors have many drawbacks. These include susceptibility to light source, detector drift, and changes in the optical path (A.K. McEvoy, 1996). These effects can be minimized by operating the sensor in the time domain instead of in the intensity domain. The lifetime τ is an intrinsic property of the lumophore that, unlike intensity, is virtually independent of external perturbations. The fluorescence is quenched in the presence of oxygen, and this quenching is described in Eq. (13). Using this relationship, the dissolved oxygen concentration can be measured.

Phase modulation fluorometry is the preferred method of conducting fluorescence lifetime measurements particularly with long lifetime fluorophores due to the simplicity and low cost of the equipment required in comparison to direct lifetime measurements. With a sinusoid ally modulated excitation signal, any fluorescence produced will also be modulated at the same frequency but will be time delayed, or phase shifted, relative to the excitation signal. The relationship between the lifetime,

τ , and the phase angle, θ , for a single exponential decay is given by (A. Campbell, 2004):

$$\tan\theta = 2\pi f \tau \qquad (15)$$

Where, f is the modulation frequency. The optimum modulation frequency that maximizes sensitivity is generally found by adjusting the frequency so that the phase change between extremes of measurements (0 and 100% DO) is maximized.

Fig.1 illustrates the principles of phase modulation fluorometry. Many optical oxygen sensors have been tested that utilize phase modulation or other lifetime measurements for oxygen and DO measurements.
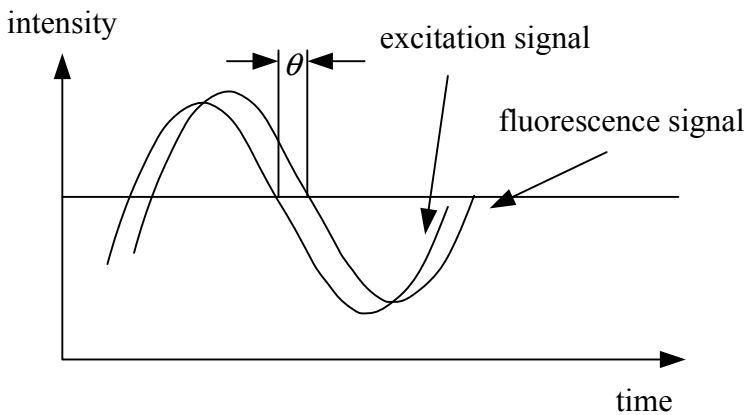
Fig. 1. Illustration of phase shift in fluorescence emission

Experiments show the method overcome the disadvantage of iodometric method and current determination method, photochemical stability and reproducibility is very good, without delay, high precision, long life, Can be real-time monitoring of dissolved oxygen in water, The range is generally 0~20mg/L, the general accuracy is lower to 1%, response time is lower to 60s.

2.4   Other Detection Methods

Conductometry**:** Use conductive metal thallium or other compounds to react with dissolved oxygen to produce the thallium ion. Through the determination of water conductivity delta, can get the concentration of dissolved oxygen (DO).Experiments show that increase by 0.035 S/cm of conductivity is equivalent to 1 mg/L dissolved oxygen (DO). This method is the most sensitive method for monitoring dissolved oxygen (DO).

Anodic stripping voltammetry: Using conductive metal thallium to react with dissolved oxygen to produce the thallous ion. Then use template leaching to determine the concentration of the Tl+ ions, thereby obtained the concentration of dissolved oxygen (DO). This method has the advantages of sampling quantity is little, high sensitivity and temperature effect is not seriously.

3   The Present Situation of the Research on DO in Water at Home and Abroad

Conventional program of water quality tests in China is analysis sampling in laboratory, the middle part of the work complexity, leading to long time detection, unable to get the water quality timely. Now some of the domestic unit and research institutions have developed some minitype dissolved oxygen detector, general base on current determination method, such as JPSJ-605 dissolved oxygen analyzer of Shanghai ray magnetic instrument factory, H-BD5W handheld water quality instrument of Beijing BeiDouXing industrial chemical institute, there is still a gap in measurement speed with foreign similar instrument. Domestic fluorescent DO sensor also has some research (Xingliang Wang, 2009; Zhaoquan Lu, 2007), technology has reached the international average, but less research on commercialization. Foreign research general use of new dissolved oxygen meter based on fluorescence quenching effect (Xiaoli Xiong, 2006; Bharathibai J. Basu, 2005; R.N. Gillanders, 2004), representative products are the MICROXI dissolved oxygen meter of DMP company of Switzerland, OXYMON oxygen measurement system of United States, measurement precision is high and speed is fast, and can be used in remote measurements.

In general, most of the commercial dissolved oxygen meter on the market is based on Clark oxygen electrode, fiber-optic oxygen sensor based on fluorescence quenching method is less.

4   Conclusion

Currently the mainstream of international development is fiber-optic oxygen sensor based on fluorescence quenching, According to fluorescence quenches characteristics, use the following method to realize dissolved oxygen detector: optical signal emitted by the source filter to a area of the fluorescent indicator, dissolved oxygen react with fluorescent indicator, causing change of the light intensity, wavelength, frequency, phase, polarization and other optical characteristics, then arrive in the detectors and signal processing devices, get information of the concentration of dissolved oxygen.

To prevent contamination, corrosion and interference of water biological, anti-interference ability is the key of instrument. We should be research on chemical stability of sensing membrane, anti-corrosion apparatus and stability of the circuit.

View of the optical sensor based on fluorescence quenching method with high accuracy and strong anti-interference ability, and good repeatability and stability, can be used for water quality measurements in aquaculture and water pollution level measurement in agriculture, so research on this sensor has important practical applications and commercial value.

References

1. Cox, B.A.: A review of dissolved oxygen modeling techniques for lowland rivers. The Science of the Total Environment (314-316), 303–334 (2003)
2. Best, M.A., Wither, A.W., Coates, S.: Dissolved oxygen as a physico-chemical supporting element in the Water Framework Directive. Marine Pollution Bulletin 55, 53–64 (2007)
3. Helm, I., Jalukse, L., Vilbaste, M., Leito, I.: Micro-Winkler titration method for dissolved oxygen concentration measurement. Analytica Chimica Acta 648, 167–173 (2009)
4. Zhang, Y., Johnson, D.: Supersaturated dissolved oxygen measured by the phosphorescence decay rate. Sensors and Actuators B 96, 379–384 (2003)
5. Lakowicz, J.R.: Principles of Fluorescence Spectroscopy. Academic/Plenum Publishers, New York (1999)
6. Henderson, R.K., Baker, A., Murphy, K.R., Hambly, A., Stuetz, R.M., Khan, S.J.: Fluorescence as a potential monitoring tool for recycled water systems: A review. Water Research 43, 863–881 (2009)
7. Lee, S., Shin, Y.B., Pyo, H., Park, S.H.: Highly sensitive oxygen sensing material: thin silica xerogel doped with tris(4, 7-diphenyl-1, 10- phenanthroline)ruthenium. Chem. Lett. 4, 310–311 (2001)
8. Choi, M.M.F., Xiao, D.: Linear calibration function of luminescence quenching-based optical sensor for trace oxygen analyses. Analyst. 224, 695–698 (1999)
9. McEvoy, A.K., McDonagh, C., MacCraith, B.D.: Dissolved oxygen sensor based on fluorescence quenching of oxygen: sensitive ruthenium complexes immobilized in sol–gel-derived porous silica coatings. Analyst. 121, 785–788 (1996)
10. Campbell, A., Uttamchandani, D.: Optical dissolved oxygen lifetime sensor based on sol-gel immobilization. IEE Proc.-Sci. Meas. Technol. 151 (2004)
11. Wang, X., Ren, Y., Chen, S.: On-line DO survey meter based on optic fiber sensor. Mechanical & Electrical Engineering Magazin. 26, 89–91 (2009)
12. Lu, Z., Zhao, X.: Development of the fluorescence quenching seawater dissolved oxygen detector based on AT89S52. Journal of Hefei University of Technology 30, 174–177 (2007) (in Chinese)
13. Xiong, X., Xiao, D., Choi, M.M.F.: Dissolved oxygen sensor based on fluorescence quenching of oxygen-sensitive ruthenium complex immobilized on silica–Ni–P composite coating. Sensors and Actuators B 117, 172–176 (2006)
14. Basu, B.J., Thirumurugan, A., Dinesh, A.R., Anandan, C., Rajam, K.S.: Optical oxygen sensor coating based on the fluorescence quenching of a new pyrene derivative. Sensors and Actuators B 104, 15–22 (2005)
15. Gillanders, R.N., Tedford, M.C., Crilly, P.J., Bailey, R.T.: A composite sol–gel/fluoropolymer matrix for dissolved oxygen optical sensing. Journal of Photochemistry and Photobiology A: Chemistry 163, 193–199 (2004)

Study on the Walnut Mechanical Characteristics and Shucking Technology Based on Finite Element Analysis

Hongmei Xu, Shuiping Yan, Yi Wang, and Meiying Liu

College of Engineering, Huazhong Agriculture University, Wuhan 430070, China
xhm790912@163.com

Abstract. In order to investigate the shucking technology and design the efficient Sheller, it's highly significant to study on the mechanical characteristics of walnut. The Pro/E software and finite element method was adopted to build the geometry and finite element model of walnut. Afterwards, the walnut stress distribution with different loads was analyzed to seek the optimum force-applying forms which usually produce the smaller deformation of walnut shell and more local crackles with better expansibility. The results show that the walnut shell can be shelled from the kernel easily when applied with the uniform and linear loads in the direction of shorter axis.

Keywords: Walnut, Shucking, Mechanical Characteristics, Finite Element Analysis.

1   Introduction

Walnut is a kind of dried fruit with rich nutrition. If save with shell, it is easy to mildew. It is of great significance to shuck the walnut effectively. Because of the unreasonable force-applying forms, most peanut shellers used in domestic is of larger breaking rate, and can't meet the requirements of hulling rate, rate of whole and price-performance ratio[1,2].

Crack propagation is the basic condition of walnut shucking. In order to improve the shucking quality of shellers, it is important to apply a reasonable force on the walnut and make it crack easily. So it is necessary to analyze the force-applying forms and responses, and seek the optimum force-applying forms [3, 4].

The finite element method is an effective method to simulate the load conditions and compute the response. Nowadays, the method has been widely used for analyzing the mechanical characteristics of nut[5]. The Pro/E software and finite element method was adopted to build the geometry and finite element model of walnut. Afterwards, the walnut stress distribution with different loads was analyzed to seek the optimum force-applying forms which usually produce the smaller deformation of walnut shell and more local crackles with better expansibility. The results can provide references for the designment of walnut shellers.

2   Geometry Size and Mechanical Characteristics of Walnut

2.1   Geometry Size

Take leatheroid walnut as the research subject, it behaves itself with irregular shape, so the model can be considered as an elliptic model in simulation. Take 20 grains of walnut into consideration, the geometry sizes are computed in the method of statistical analysis. The geometry sizes in the three direction of XYZ are described as follows: 27 mm, 25 mm and 30 mm. Additionally, the average thickness and mean value of the gap between walnut shell and kernel are 1mm and 0.2 mm respectively.

2.2   Mechanical Characteristics

The shell and kernel of walnut belong to two different kinds of material, and the stiffness of shell is greater than that of the kernel. So their elastic modulus is inevitability different from each other. Additionally, the materials of shell and kernel can be regarded as isotropic materials, that is to say the horizontal elastic modulus of materials are equivalent to the longitudinal and tangential elastic modulus. Refer to the wood and other nuts, and the elastic modulus of walnut shell is taken as 10 Mpa, which is ten times of that of walnut and kernel. The Poisson Ratio is 0.3[5, 6].

3   Finite Element Model of Walnut

3.1   Geometry Model

The structure of walnut is symmetric, so the 1/2 symmetric model can be used for simulation. Fig.1 shows the 1/2 solid model of walnut built by the Pro/E software. When the model is converted to the IGES files, it can be imported to the software of ANSYS for finite element analysis.
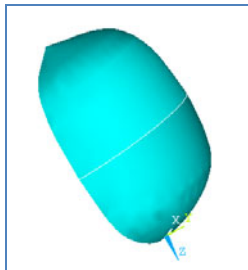
Fig. 1. Half model of walnut

3.2   Element Type Selection and Division

Walnut consists of two parts, shell and kernel. In order to simplify the computation, the materials of shell and kernel can be regarded as linear elastic and isotropic materials. Furthermore, in the process of element type selection, the element type of walnut shell is defined as Shell 93, while the kernel is defined as Solid 95.In addition, the inner surface of walnut shell and the outer surface of walnut kernel makes for a contact. In order to simulate the interaction between the walnut shell and kernel, the contact element was introduced into the finite element model of walnut. The inner surface of walnut shell is defined as the target surface, while the outer surface of walnut kernel is defined as the contact surface. When using the contact-creating guide to create the contact, the software system can automatically generate the target element of Targe170 and contact element of Conta 174.Since the walnut model is relatively simple, the automatic meshing tool was adopted to mesh the walnut shell and kernel.

3.3   Definition of Load and Constraint

In order to seek the optimum force-applying forms for shell shucking, the stress and strain distribution with six different loads was investigated by simulation. The loading forms can be described as follows :(1) applied with the concentrated load in the long axis (Fig. 2 (a)), and the constraint with zero displacement is applied to the bottom of walnut; (2) applied with the uniform and linear load in the long axis (Fig. 2 (b)), and
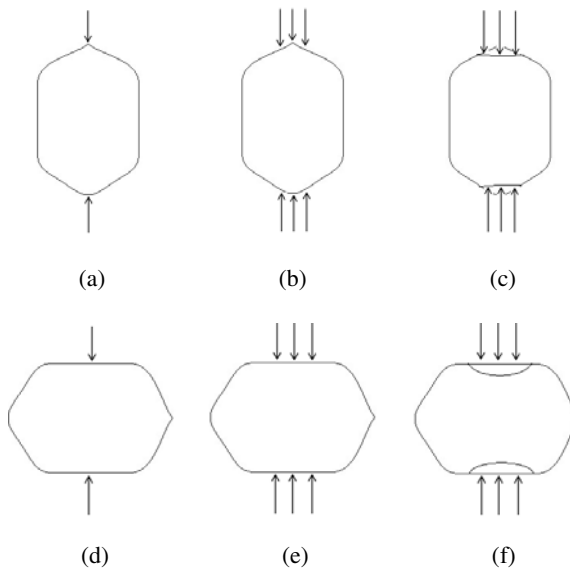
Fig. 2. Loading forms of walnut

the constraint with zero displacement is applied to the bottom of walnut; (3) applied with the uniform surface load in the long axis (Fig. 2 (c)), and the constraint with zero displacement is applied to the bottom of walnut;(4) applied with the concentrated load in the short axis (Fig. 2 (d)), and the constraint with zero displacement is applied to the opposite key nodes of walnut; (5) applied with the uniform and linear load in the short axis (Fig. 2 (e)), and the constraint with zero displacement is applied to the opposite key nodes of walnut; (6) applied with the uniform surface load in the short axis (Fig. 2 (f)), and the constraint with zero displacement is applied to the opposite key nodes of walnut.

4   Finite Element Analysis of Walnut

4.1   Stress and Strain Analysis of Walnut Applied with the Concentrated Load in the Long Axis

Fig.3 shows the finite element model of walnut applied with the concentrated load in the long axis. The load is 700 N. Fig.4 shows the Z-Component of displacement, which is also known as the deformation in the loading direction. As shown in the figure, the largest deformation is located at the loading point, and the deformation area looks like a semicircle about the loading point. Fig.5 and Fig.6 shows the equivalent stress and strain distribution of walnut respectively. It can be seen that the largest stress and strain is located at the loading point. The stress and strain spreads from the point out to the periphery, and decreases gradually.
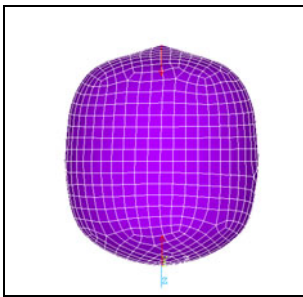
Fig. 3. Finite element model of walnut
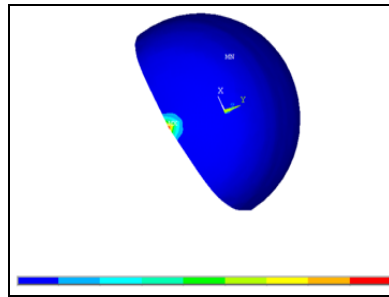
Fig. 4. Z-Component of displacement
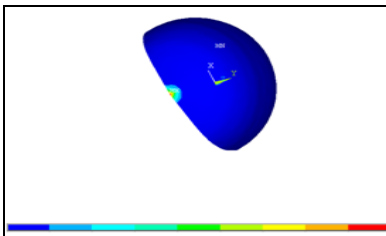
Fig. 5. Von Mises stress graph of walnut
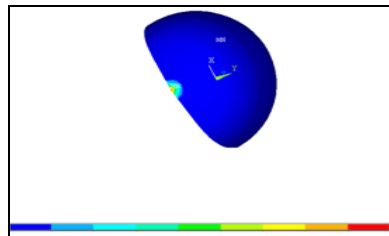
Fig. 6. Von Mises strain graph of walnut

According to the above analysis, we can predict that the stress and strain distribution of walnut behaves itself with no definite direction. The walnut cracks from the loading point out to the periphery, which usually brings about the local fracture, and works against the shell shucking of walnut.

4.2   Stress and Strain Analysis of Walnut Applied with the Uniform and Linear Load in the Long Axis

Select the upper endpoint of the long axis as the central point of the loading curve, and apply the uniform and linear load to the loading curve in the positive direction of Z axis. The total length of loading curve is 2mm, and the uniform load is 350N/mm.

Fig.7 shows the Z-Component of displacement when walnut is applied with the uniform and linear load in the long axis. As shown in the figure, the largest deformation is located in the loading curve, and the neighboring areas are the next. In other words, outward from the loading curve, the deformation decreases gradually.

Fig.8 and Fig.9 shows the equivalent stress and strain distribution of walnut respectively. Obviously, the distribution of equivalent stress is similar to that of walnut strain, and the distribution area looks like a semi ellipse. The direction of longer axis is consistent with the loading direction, and the maximum stress and strain is located in the loading curve, and the neighboring areas are the next.

On the basis of this, it can be concluded that the stress and strain distribution of walnut behaves itself with definite direction. The walnut should crack from the loading curve out to its sideways. However, since the distribution areas of walnut stress and strain are larger, the rupture rates can meets the corresponding requirements, but a great deal of broken kernels are caused meanwhile.
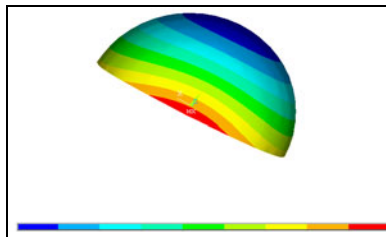
Fig. 7. Z-Component of displacement
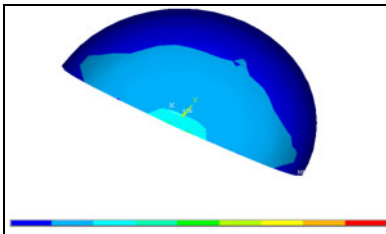
Fig. 8. Von Mises stress graph of walnut
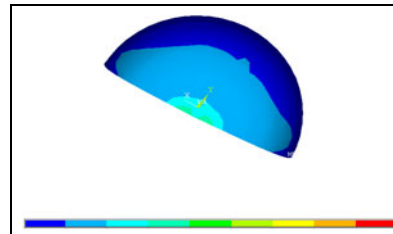
Fig. 9. Von Mises strain graph of walnut

4.3   Stress and Strain Analysis of Walnut Applied with the Uniform Surface Load in the Long Axis

Select the upper endpoint of the long axis as the central loading point, and apply the uniform surface load to the loading surface in the positive direction of Z axis. The total area of loading surface is 10 mm^2, and the uniform load is 70N/ mm^2.Fig.10 shows the Z-Component of displacement when walnut is applied with the uniform surface load in the long axis. As shown in the figure, the maximum and minimum deformations are located on the two sides of walnut respectively.

Fig.11 and Fig.12 shows the equivalent stress and strain distribution of walnut respectively. It is clear that the stress distribution is similar to that of walnut strain, and the greater stress and strain is mainly concentrated in the loading area. The stress and strain almost remains unchanged in the loading area.

According to the above analysis result, it can be speculated that the crack of walnut is lack of directionality and the loading area cracks first. Compared with the first two conditions, the rupture zone is much larger, and meets the requirements of the rupture rate. Being lack of definite directionality, the stress of the two sides usually produces lots of broken kernels.
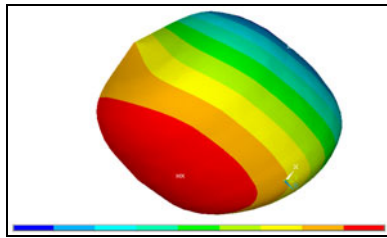
Fig. 10. Z-Component of displacement
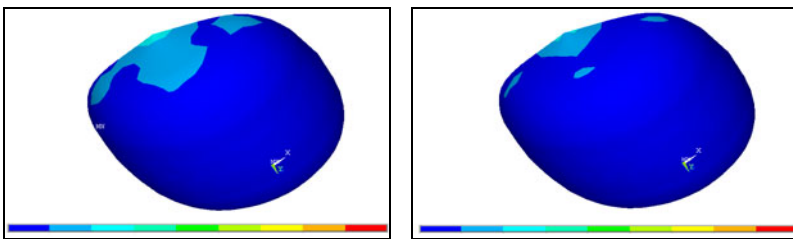
Fig. 11. Von Mises stress graph of walnut   **Fig. 12.** Von Mises strain graph of walnut

4.4   Stress and Strain Analysis of Walnut Applied with the Concentrated Load in the Short axis

Select the endpoint of the short axis as the central loading point, and apply the concentrated load of 300N to the loading point in the negative direction of Y axis.

Fig.13 shows the Y-Component of displacement, which is also known as the deformation in the loading direction. As shown in the figure, the largest deformation is located at the loading point, and the deformation area looks like a circle about the loading point.

Fig.14 and Fig.15 shows the equivalent stress and strain distribution of walnut respectively. It can be seen that the stress and strain distribution is similar to that of Y-Component displacement, and lack of definite directionality. The distribution area just looks like a circle about the loading point and the maximum stress and strain is just located at the loading point.

Since the stress and strain distribution of walnut behaves itself with no definite direction, it can be concluded that the walnut cracks from the loading point out to the periphery, which usually brings about the local fracture, and is not useful for the shell shucking of walnut.
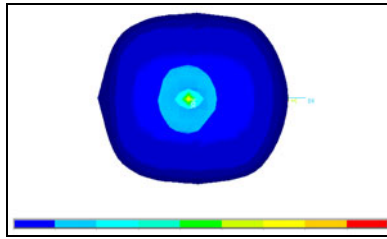
Fig. 13. Y-Component of displacement
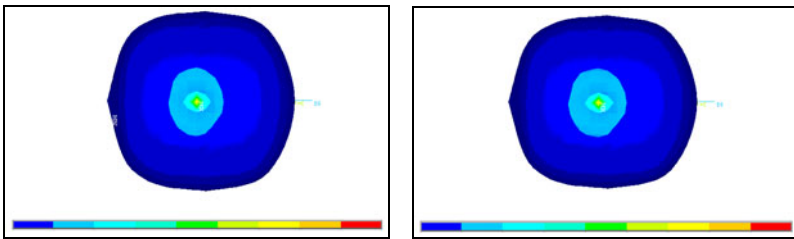
Fig. 14. Von Mises stress graph of walnut   **Fig. 15.** Von Mises strain graph of walnut

4.5   Stress and Strain Analysis of Walnut Applied with the Uniform and Linear Load in the Short Axis

Select the endpoint of the short axis as the central loading point and apply the uniform and linear load to the loading curve in the negative direction of Y axis. The total length of loading curve is 2mm, and the uniform load is 150N/mm.

Fig.16 shows the Y-Component of displacement when walnut is applied with the uniform and linear load in the short axis. As shown in the figure, the largest deformation is located in the loading curve and outward from the loading curve, the deformation decreases gradually.

Fig.17 and Fig.18 shows the equivalent stress and strain distribution of walnut respectively. Obviously, the distribution of equivalent stress is similar to that of walnut strain, which shows definite directionality. Furthermore, the distribution area can be regarded as ellipse, whose direction in the long axis is in accordance with the direction of load distribution. The maximum stress and strain is located in the loading curve, and the neighboring areas are the next. That is to say, the stress and strain decreases gradually from the loading curve to the sideways.

Since the stress and strain distribution of walnut behaves itself with definite direction, the walnut should crack from the loading curve out to its sideways, and the crack is highly expandable. Therefore, this kind of loading form is suitable for shell shucking.
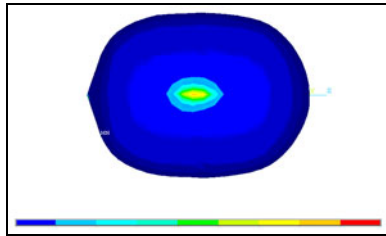
Fig. 16. Y-Component of displacement
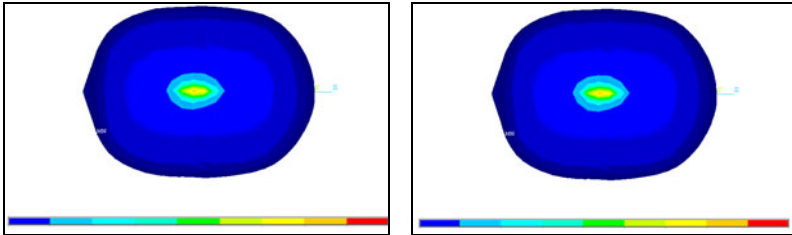
Fig. 17. Von Mises stress graph of walnut   **Fig. 18.** Von Mises strain graph of walnut

4.6   Stress and Strain Analysis of Walnut Applied with the Uniform Surface Load in the Short Axis

Select the endpoint of the short axis as the central loading point, and apply the uniform surface load to the loading surface in the negative direction of Y axis. The total area of loading surface is 10 mm^2, and the uniform load is 30N/ mm^2.

Fig.19 shows the Y-Component of displacement when walnut is applied with the uniform surface load in the short axis. As shown in the figure, the deformation area is similar to the loading area, but much larger that the latter.

Fig.20 shows the equivalent stress of walnut. It can be seen that the greater stress is mainly concentrated in the loading area, much smaller stress is also scattered in the other areas.

Fig.21 shows the equivalent strain distribution of walnut. Compared with the equivalent stress, the distribution area of greater strain is relatively small. The greater strain is mainly concentrated in the loading area.

On the basis of this, it can be speculated that the crack forms of walnut under this condition is similar to that of the condition with concentrated load, and lack of definite directionality. Because the rupture zone is much larger, this loading form meets the rupture requirements. However, it produces lots of broken kernels at the same time, and works against the shell shucking of walnut.
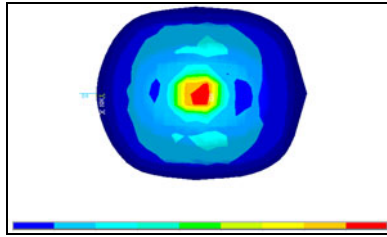
Fig. 19. Y-Component of displacement
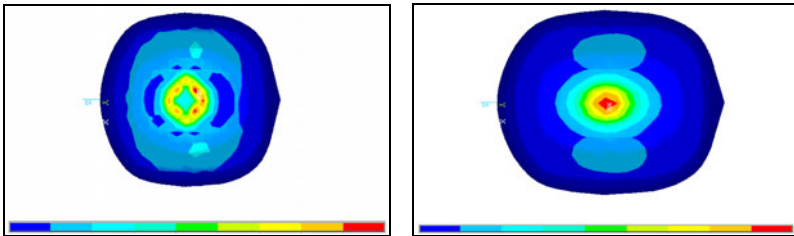
Fig. 20. Von Mises stress graph of walnut   **Fig. 21.** Von Mises strain graph of walnut

5   Conclusions

The Pro/E software and finite element method was adopted to build the geometry and finite element model of walnut. On the base of this, the stress and strain distribution with six different loads was investigated by simulation. The conclusions can be described as follows:

(1)When applied with the concentrated and uniform surface load in the long axis, the stress and strain distribution of walnut behaves itself with no definite direction, the walnut cracks from the loading point out to the periphery, which usually brings about the local fracture, and works against the shell shucking of walnut; when applied with the uniform and linear load in the long axis, the stress and strain distribution of walnut shows definite directionality, and the walnut cracks from the loading curve out to its sideways. The larger distribution areas usually bring about lots of broken kernels.

(2) When applied with the concentrated and uniform surface in the short axis, the stress and strain distribution of walnut is lack of definite directionality, the walnut cracks from the loading point out to the periphery, which usually brings about the local fracture, and works against the shell shucking of walnut; when applied with the uniform and linear load in the short axis, the walnut cracks from the loading curve out to its sideways, and the crack is highly expandable. This kind of loading form is very suitable for shell shucking.

(3) It is the most suitable loading mode for walnut shell shucking to impose the uniform and linear load on the short axis.

References

1. Yuhua, C., Changyou, L., Yanmei, Q., et al.: Finite element analysis on mechanical properties of castor-capsule. Journal of Jiangsu University 31(4), 383–387 (2010)
2. Zhengyao, D., Dequan, Z., Liang, Q.: Finite element modeling and mechanical analysis of hickory kernel. Agriculture Equipment and Vehicle Engineering 6, 6–9 (2010)
3. Jianxin, S., Haijun, Z., Dongjun, X.: Technology for breaking walnut shell based on finite element analysis. Transactions of the CSAE 21(3), 185–188 (2005)
4. Rong, W., Qunying, J., Dejiang, W., et al.: Analysis of finite element method for threshing gingko fruit shell. Transactions of the CSAE 19(4), 58–61 (2003)
5. LingJun, W., YanMing, Q., WenJun, D.: Mechanical characteristics and the finite element analysis of grapes. Transactions of the CSAE 21(3), 7–10 (2005)
6. Lijuan, X., Li, Z.: Research on Optimal Crack Position of Lotus Seed based on FEA AND experiment. Huazhong Agriculture University, Wuhan (2005)

Experimental Study on Forced-Air Precooling of Dutch Cucumbers

Jingying Tan, Shi Li, and Qing Wang

School of Mechanical Engineering, Hunan Institute of Science and Technology, Yueyang, Hunan Province, China 41400
{jingyingtan,wqwyy}@163.com, 7750496@qq.com

Abstract. Experiments about the forced-air precooling of cucumbers were carried out under different operation parameters. The results show that the relation between cooling rate and air velocity can be given by power function. The linear relationship between the initial temperature of cold air and cucumbers and the final temperature of cucumbers is gained. Temperature gradient exists in different radius position in the same cross section. But the temperature gradient is not significant for the beginning and the final during the procooling process.

Keywords: Forced-air precooling, air velocity, initial temperature, temperature gradient.

1   Introduction

The precooling means that the tempreature is quickly decreased from the room temperature to the storage temperature. The precooling process is an important technological part for the food storage. The shelf life of the fruit-vegetable can be delayed significantly and the freshness and the nutritional ingredient can be kept better after precooling. There are several types for precooling, such as vacuum precooling, forced-air precooling, water precooling, water-air precooling. Several factors, such as precooling cost, precooling time, and the conditions of the cooling process, must be considered in the choice of the precooling types [1-2]. The parameters of the precooling process can be predicted accurately according to the experimental data of precooling, experimental formula, and some charts. Ibrahim Dincer researched the precooling process of grape monocase by experiments [3]. Vigneault researched the ventilation resistance of the precooling [4]. Some researches did some experiments about the differential pressure precooling for the opening packing container [5-7]. The experiements were done in the long and narrow duct.

The previous researches focused on the experiments and there were some differences comparing with the practical precooling production. The forced-air precooling, the best suitable types of the precooling for my country, is studied for the Dutch cucumbers in this paper. The effect of the precooling parameters during the precooling process is given.

2   Materials and Methods

2.1   Materials

The column Dutch cucumbers gained from the demonstration base of Country Farming Production in Tianjin were used in the experiment. The sizes of the cucumbers are uniform, diameter 25mm, and there are no plant diseases and insect pests and mechanical damnification.

2.2   Experimental Conditions

The experiments are conducted in the fruit-vegetable pre-refrigeratory and the conditions are follows:

    (1) air velocity: 0.5m/s, 1.0m/s, 1.25m/s, 1.5m/s, 1.75m/s, and 2m/s.
    (2) initial temperature for cold air: 2℃, 3℃, 4℃, 5℃, and 6℃.
    (3) initial temperature for cucumber: 18℃, 19℃, 20℃, 21℃, 22℃, 23℃, 24℃, 25℃, 26℃, 27℃, and 28℃.
    4) measuring position of the cucumber: center section, 1/2 radius, 3/4 radius, and surface.

2.3   Data Acquisition System

(1) temperature acquisition system
The armoured thermocouple, WRNK-111, φ1×1000ms, leading wire 5m, is used to measure the temperature. The standard mercury filled thermometer is used to calibrate, ±0.1℃ for the measurement accuracy.
    (2) humidity acquisition system
EE21-FT6A23CA0170 system is used to defect the humidity real-time.
    (3) air velocity acquisition system
The sensor of EE65-VB5 ELEKTRONIK is chosen.
    The data of the temperature, humidity, and the air velocity is collected by computer, 30s for the minimum collection interval. The interval of 1min is chosen for the data acquisition in the experiment.
    The air velocity sensor, the temperature sensor, and the humidity sensor of SwemaAir300 are used for the experiment.

3   Results and Discussions

3.1   The Effect of the Air Velocity on the Falling Temperature of the Center

The falling temperature curves, the initial temperature of cold air 4℃ and cucumber 25℃, are shown in Fig. 1.
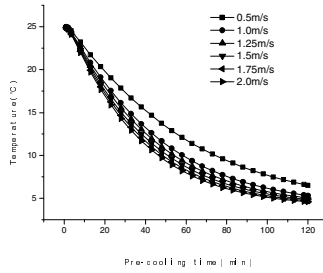
Fig. 1. The effect of the air velocity during the precooling process

It can be seen that the falling temperature curves are similar under air velocities. But the falling temperature effect is not proportional to the air velocity. The better precooling effect can not be gained by increasing air velocity, but the cost will quickly increase, especially more than 1.5m/s. So it can be concluded that the air velocity 1.5m/s is suitable.

The falling temperature equation for different air velocities is given by fitting the falling temperature curves.

$$T = T_0 + A_0 e^{-R_0 t} \tag{1}$$

The parameter T0, R0, and A0 are shown in Table 1.

Equation 1 can be rewritten

$$\frac{T - T_0}{T_i - T_0} = e^{-R_0 t} \tag{2}$$

Where T, center temperature at any time; T0, initial temperaure of center temperaure; Ti, cold air temperature; A0=Ti-T0, fitting value; R0, cooling rate in center.

The cooling rate will increase with the air velocity increasing and the relationship between them can be discribed by the power function according to table 1 and figure. 2. Equation 3 is given.

$$R_0 = 0.2207 u^{0.35352} \tag{3}$$

Table 1. Fitting results for different air velocities

		A0	R0	T0	Correlation coerricent R2
Air velocity(m/s)	0.5	Fitting value	21.8373	0.01717	4	0.99853
		Standard deviation	± 0.11115	± 0.00017	± 0	
	1.0	Fitting value	21.96015	0.02215	4	0.99868
		Standard deviation	± 0.11982	± 0.00022	± 0	
	1.25	Fitting value	21.99931	0.02396	4	0.99874
		Standard deviation	± 0.1211	± 0.00024	± 0	
	1.5	Fitting value	22.03124	0.02551	4	0.99879
		Standard deviation	± 0.12169	± 0.00026	± 0	
	1.75	Fitting value	22.03496	0.02557	4	0.99879
		Standard deviation	± 0.12207	± 0.00026	± 0	
	2.0	Fitting value	22.08216	0.02809	4	0.99887
		Standard deviation	± 0.12196	± 0.00029	± 0	
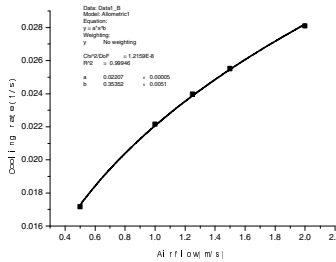
Fig. 2. Fitting curves between the cooling rate and the air velocity in the center

3.2 The Effect of the Initial Temperature of Cold Air during the Precooling Process

The results for the effect of the initial temperature of cold air during the precooling process is shown in Fig. 3, air velocity 1.5m/s and the initial temperature of cucumber 25℃.

It can be seen that the temperature decreases quickly when the lower initial temperature and the lower final temperature will be gained after precooling for two hours. The temperature curves between the initial temperature of cold air and the final temperature of cucumber are given in Fig. 4 and the linear relationship is given.
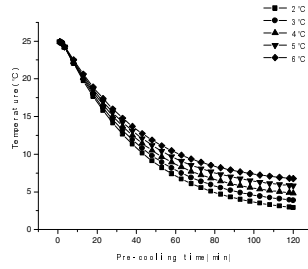
Fig. 3. The effect of the initial temperature of cold air during the precooling process in center
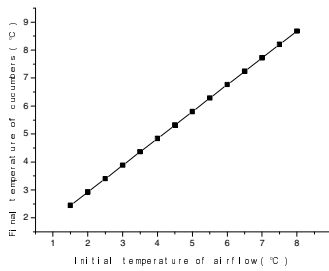
Fig. 4. The relationship between the initial temperature of cold air and the final temperature of cucumber in center

3.3   The Effect of the Initial Temperature on the Final Temperature for the Cucumber

The precooling curves of different initial temperature are given in Fig. 5, the initial temperature of the cold air 4℃ and air velocity 1.5m/s.
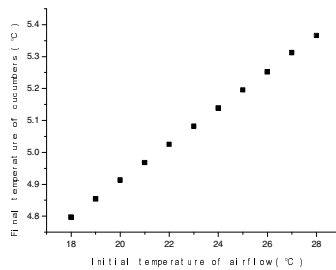
Fig. 5. The relationship between the initial temperatures and the final temperature

The linear relationship between the initial temperatures of airflow and final temperature of cucumbers is gained and it is similar to the previous results.

3.4   The Temperature Changes in Different Section for the Whole Precooling Process

The temperature change curves for different sections, center, 1/2 radius, 1/4 radius, and surface, are given, air velocity 1.5m/s, initial temperature for cucumber 25℃, initial temperature for cold air 4℃.
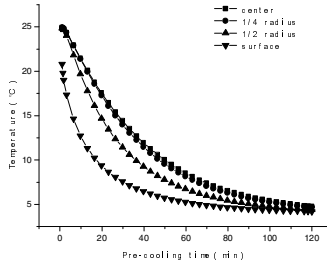
Fig. 6. The temperature changes in different section during the precooling process

It can be concluded that there is temperature gradient in different radius of the same cross section during the precooling process. The high cooling rate near the outside of the cucumber is gained due to the small resistance. The final temperature of the cucumber is close to the initial temperature of cold air 4℃ after the long precooling process.

4   Conclusions

It can be concluded from the forced-air precooling experiments for the cucumbers.

(1) Cooling rate increases with the increase in air velocity and the power function relationship between them is gained. air velocity 1.5m/s is best suitable for the precooling.

(2) There is the linear relationship between the initial temperature of the cold air and the cucumber and the final temperature of cucumber. The better cooling effect can be gained if the lower temperature of cold air is used, but the damage should be prevented when lower temperaure. The operation efficiency can be impoved when the fruit-vegetables are picked not at noon but in dusk or in morning.

(3) There is temperature gradient in different radius of the same cross section. Sometimes the temperature gradient is significant during the precooling process due to the inside resistance of the cucumbers.

References

1. Jin, D., Tan, J.Y., Wang, Q.: Study on Refrigeratory Compressor with Frequency Conversion and its Economical Efficiency. In: International Conference on Computer and Computing Technologies in Agriculture, Nanchang, pp. 445–451 (2010)
2. Tan, J.Y., Jin, D., Wang, Q.: Study on the forced-air precooling process of torispherecal fruits and vegetables in cooling bed. Intillegent Automation & Soft Cumputing 16(6), 1187–1195 (2010)
3. Ibrahim, D.: Air flow precooling of individual grapes. Journal of Food Engineering 26, 243–249 (1995)
4. Vigneault, C., Markarian, N.R., Silva, A., et al.: Pressure drop during forced-air ventilation of various shorticultural produce in containers with different opening configurations. Transactions of the ASAE 47(3), 807–814 (2004)
5. Yang, Z., Zhao, C.E., Wang, L.Y., et al.: Pressure Drop Characteristics in Forced-air Pre-cooling of Longan Fruits. Transactions of the Chinese Society for Agricultural Machinery 38(1), 105–108 (2007)
6. Huang, J., Liu, J., Fang, Z.: Experimental study on pressure drop in pressure-difference precooling of cylinder fruits. Refrigeration and Air Conditioning 5(3), 61–65 (2005)
7. Wang, Q., Liu, F.Z., Lian, T.D.: Experimental study on parameters affecting grape pressure pre-cooling. Transactions of the Chinese Society of Agricultural Engineering 22(4), 212–215 (2006)

Experimental Study on the Effects of Mechanical and Physical Characteristics on Walnut Shucking

Hongmei Xu, Shuiping Yan, and Yapeng Bai

College of Engineering, Huazhong Agriculture University, Wuhan 430070, China
xhm790912@163.com

Abstract. The mechanical and physical characteristics have great impact on the shell shucking of walnut. Taking the shucking force as the evaluation indicator, the effects of mechanical and physical characteristics, which include the loading orientation, loading rate, moisture content and walnut partical size, on walnut shucking were investigated by means of single factor test. The results show that: 1) the loading orientation and loading rate affect the shucking force significantly, and the needed force is the minimum when the walnut is horizontally placed and loaded at the rate of 50mm/min; 2) the moisture content has significant effects on the shucking force, and with the increase of drying time, the moisture content of walnut decrease gradually, which results in the increases of the needed shucking force;3) the particle size has no great effect on the shucking force, but great difference in particle size usually corresponds to different walnut shucking force. In order to investigate the shucking technology and design the efficient sheller, it's highly significant to study on the mechanical and physical characteristics of walnut.

Keywords: Walnut, Shucking, Mechanical Characteristic, Physical Property.

1 Introduction

Walnut is one of the four nuts in the world. In recent years, with the great increase of walnut yield, the walnut foods have developed in the direction of deep processing. However, a critical step for the deep processing of walnut is to shuck the shell and take the meat. Because of the irregular shape and smaller clearance between shell and kernel, it's very difficult for walnut to shuck the shell and take the meat [1].Nowadays, because of the unreasonable force-applying forms, most walnut shellers used in domestic are of larger breaking rate and behave themselves with poor shucking quality [2, 3].

Mechanical and physical characteristics can provide fundamental data for developing the sheller and realizing the effective separation of walnut shell and meat. In order to design the highly effective and low damage sheller, mechanical and physical characteristics of walnut must be first investigated.

Taking the dried leatheroid walnut as the research object, the compression test was conducted on the universal material testing machine. The effects of mechanical and physical characteristics, which include the loading orientation, loading rate, moisture content and walnut partical size, on the shucking force of walnut were investigated by

means of single factor test. The research results are of great significance for improving the working performance of walnut sheller and designing the new and high-efficiency shucking equipments.

2 Materials and Methods

2.1 Materials

The sold leatheroid walnut from Yunnan was selected as the specimen, and the experiment has been finished in the comprehensive laboratory of College of Engineering, Huazhong Agriculture University, in May 2011.

1. Upper platen   2.Specimen   3. Bottom platen

Fig. 1. Sketch of the loading device

2.2 Devices

The experiment was conducted on the universal material testing machine of RGT2000-10, and the loading rate varies from 0.01mm/min to 500mm/min.Fi.1 shows the sketch of the loading device. The pressure imposed on the walnut specimen by the plate-type pressure head is measured by the force transducer.

2.3 Methods

The experiment was conducted on the universal material testing machine, and the specimen was applied with the static load. In the beginning of the experiment, put the specimen on the bottom platen of the testing machine, and adjust the location of the upper platen to make sure it contacts with the specimen and imposes a compression force on the specimen. In order to determine the influence of the mechanical and physical characteristics on walnut shucking, the single factor test was

conducted to investigate the effects of the loading orientation, loading rate, moisture content and particle size on the walnut shucking force. Additionally, for each experimental condition, the experiment was repeated ten times to reduce the test tolerance, and the result is the mean value of 10 tests.

3 Results and Discussion

3.1 Analysis of the Cracking Process of Walnut Shell

3.1.1 Cracking Forms of Walnut Shell

Fig.2 shows the cracking forms of walnut shell when loaded in different orientations. The cracking forms of walnut shell are of two main types, namely, cracking longitudinally and cracking transversely. Because of the structural inhomogeneity and discontinuity in the sewed interface of walnut, it features low compression strength. When the walnut is horizontally placed and applied with a load, the stress concentration and longitudinal crack usually occur. The longitudinal crack propagates in the direction of sewed interface, and plays an important role in the walnut shucking. When the walnut is vertically placed and applied with a load, the cracking form is similar to that of walnut shell when it is horizontally placed. The crack propagates in the direction of sewed interface and behaves itself with definite direction. However, when the walnuts are placed on their sides, as shown in Fig.2 (b), and applied with a load, the transverse crack occurs, and the walnut cracks from the loading point out to its sideways. Furthermore, the extension direction of crack is vertical to the sewed interface, and the crack usually ceases to expand when it reaches the interface [4].

| (a) Cracking longitudinally | （b） Cracking transversely |
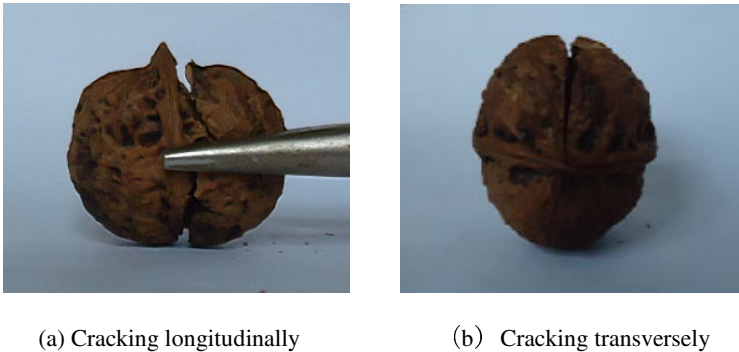
Fig. 2. Crack propagation of walnut shell

3.1.2 Cracking Process of Walnut Shell

Fig.3 shows the relation curve between the loading force and deformation of walnut when it is horizontally placed and applied with a load. As shown in the figure, the relation curve deviates from the origin of coordinate system. The reason is that there is a certain clearance between the upper platen and the walnut specimen. Thus, the

pressure maintains 0N until the upper platen contacts with the walnut specimen. The initial deformation is actually the clearance between the upper platen and the walnut specimen.

The cracking process of walnut shell mainly includes three stages, namely, elastic deformation, plastic deformation and sudden rupture. The elastic deformation mainly refers to the early loading stage, in which the pressure linearly increases with the deformation of walnut. Afterwards, with the increase of loading force, the elastic deformation goes hand in hand with plastic deformation. The plastic deformation mainly focuses in the later loading stage. However, when the loading force achieves the peak value for the first time, the walnut begins to crack and the force declines sharply. The peak value mentioned above mainly depends on the compressive strength of walnut, so it's also known as the walnut shucking force. Then, the walnut shell and meat is compressed together, and the required loading force increases again [5]. In the course of the experiment, in order to protect the testing machine, the experiment is called for a halt once the walnut begins to crack and the loading force achieves the maximum.
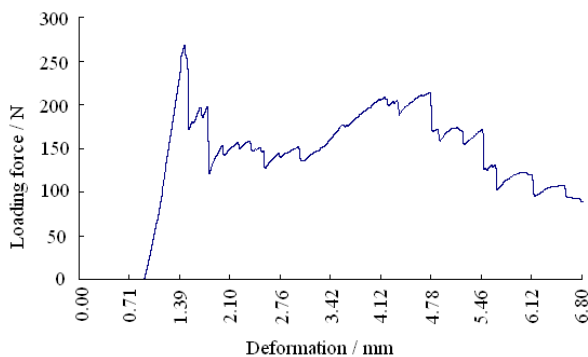
Fig. 3. Relation curve between the loading force and deformation of walnut

3.2   Analysis of the Influence Factors of Walnut Cracking

3.2.1   Effects of Loading Orientation on Walnut Cracking

Select thirty medium-size walnuts as the specimen, which is coincided with each other in the size and moisture content. Divide the walnuts into three groups, and each group contains ten walnuts. In the course of the experiment, the ten walnuts in group 1 are all horizontally placed as shown in Fig. 4(a).The walnuts in group 2 and group 3 are placed on their sides and vertically placed respectively. Compress the three groups of specimens at the loading rate of 10mm/min, and don't cease the experiment until the walnut shell begins to crack, which results in the great reduction of pressure. Record the peak value of the pressure, and regard it as the walnut shucking force.

(a) Walnut is horizontally placed (b) Walnut is placed on its side (c) Walnut is vertically placed

Fig. 4. Loading orientation of walnut

Fig.5 shows the testing results of the walnut shucking force when the walnuts are applied with loads in different directions. As shown in the figure, the needed loads are clearly different when the walnuts are applied with loads in different directions. When the walnut is horizontally placed, the needed load is the minimum, which is equal to 195.69N, and then is the needed load when the walnut is placed on its sides. The needed load is the maximum when the walnut is vertically placed. Variance analysis was conducted for the testing results of walnut shucking force, and the corresponding F-statistic is equal to 12.73, which is greater than the critical value $F_\alpha(2, 27)$ ($\alpha=0.05$).The variance analysis results are consistent with the testing results. It is, therefore, believed that the loading orientation has great impact on the walnut shucking force.
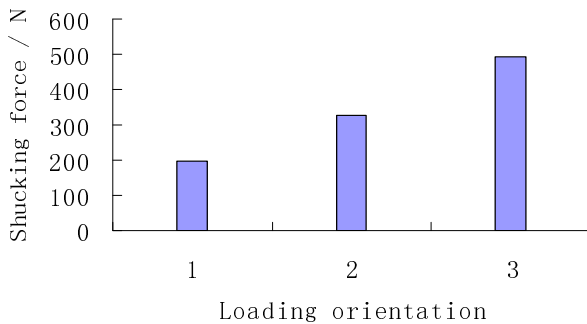
1: Walnut is horizontally placed
2: Walnut is placed on its side
3: Walnut is vertically placed

Fig. 5. Shucking force of walnut with different loading orientation

3.2.2   Effects of Loading Rate on Walnut Cracking

Select thirty medium-size walnuts as the specimen, which is coincided with each other in the size and moisture content. Divide the walnuts into three groups, and each group contains ten walnuts. In the course of the experiment, the thirty walnuts are all

horizontally placed. Compress the three groups of specimens at the loading rate of 10mm/min, 30mm/min and 50mm/min respectively.

Fig.6 shows the testing results of the walnut shucking force when the walnuts are compressed at different loading rates. As shown in the figure, the needed shucking forces are different when the walnuts are compressed at different loading rates. When the walnut is compressed at the loading rate of 50mm/min, the needed load is the minimum, which is equal to 114.04N, and then is the loading rate of 10mm/min. The needed load is the maximum when the walnut is compressed at the loading rate of 30mm/min. Variance analysis was conducted for the testing results of walnut shucking force when the walnuts are compressed at different loading rates, and the corresponding F-statistic is equal to 8.46, which is greater than the critical value $F_\alpha(2, 27)$ ($\alpha=0.05$).So it can be concluded that the loading rate affects the walnut shucking force significantly.
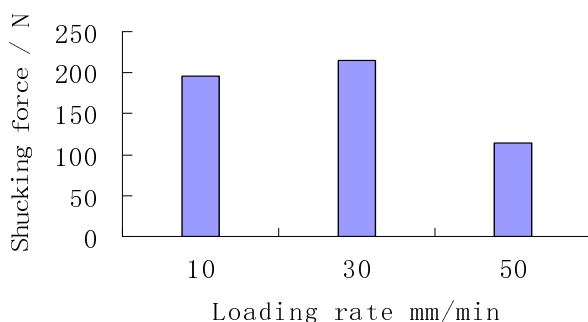
Fig. 6. Shucking force of walnut with different loading rate

3.2.3   Effects of Particle Size on Walnut Cracking

Select thirty walnuts as the specimen, which is coincided with each other in the moisture content. Divide the walnuts into three groups, and each group contains ten walnuts. The ten walnuts in group 1 with smaller size ranges from 23mm to 27mm in height when they are horizontally placed. The ten walnuts in group 2 are of medium size, and range from 28mm to 32mm in height when they are horizontally placed. The ten walnuts in group 3 are the biggest ones, and range from 33mm to 37mm in height when they are horizontally placed. In the course of the experiment, the thirty walnuts are all horizontally placed, and compressed at the loading rate of 10mm/min.

Fig.7 shows the testing results of the shucking force of walnut with different particle size. For the walnuts with different particle size, the needed loads are different, but the difference is not large. The needed shucking force of the medium-size walnuts is the minimum, which is equal to 166.47N, and then is the smaller ones. The needed shucking force of the large-size walnuts is the maximum. Variance analysis was conducted for the testing results of the shucking force of walnut with different particle size, and the corresponding F-statistic is equal to 1.43, which is less than the critical value $F_\alpha(2, 27)$ ($\alpha=0.05$).It can be concluded that particle size has no great effect on the walnut shucking force.
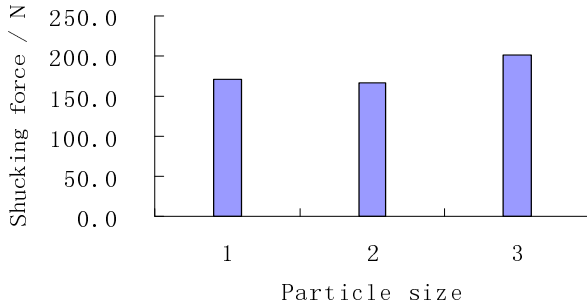
1: Walnut is of larger size
2: Walnut is of medium size
3: Walnut is of smaller size

Fig. 7. Shucking force of walnut with different particle size

3.2.4   Effects of Moisture Content on Walnut Cracking

Select thirty medium-size walnuts as the specimen, which is coincided with each other in the particle size .Divide the walnuts into three groups, and each group contains ten walnuts. In order to investigate the effect of the moisture content on the walnut shucking force, the walnuts must be pretreated before the compression test. The drying experiment is one of the most important preprocessing methods, which is usually used to control the moisture content of the samples. In this research, the drying experiments were conducted for three groups of walnut specimens, and the corresponding drying time is 0h, 1h and 2h respectively. Afterwards, the walnuts are all horizontally placed, and compressed at the loading rate of 10mm/min.

Fig.8 shows the testing results of the shucking force of walnut with different drying time and moisture content. As shown in the figure, different drying time corresponds to different walnut shucking force. When the drying time is 0h, the needed shucking force is the minimum, which is equal to 114.49N, and then is the walnuts with the drying time of 1h. The needed shucking force of the walnuts with the drying time of
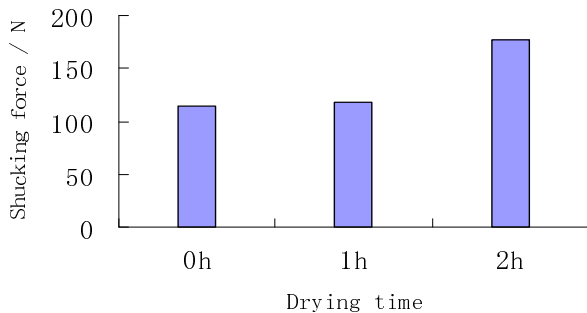
Fig. 8. Shucking force of walnut with different moisture content

2h is the maximum. It can be seen that the moisture content of walnuts decreases with the increase of drying time. As a result, the needed shucking force increases which makes walnut shucking more difficult. Variance analysis was conducted for the testing results of the shucking force of walnuts with different moisture content, and the corresponding F-statistic is equal to 49.23, which is greater than the critical value $F_\alpha(2, 27)$ (α=0.05). Therefore, it can be concluded that the drying time and moisture content has great effects on the walnut shucking force.

4   Conclusions

Take the shucking force as the evaluation indicator, and the effects of the loading orientation, loading rate, moisture content and walnut particle size on walnut shucking were investigated by means of single factor test. The conclusions can be described as follows:

(1) Loading orientation has great impact on the walnut shucking force. When the walnuts are horizontally placed, the needed shucking force is the minimum, and then is the walnuts placed on their side. The needed shucking force of the walnuts vertically placed is the maximum.

(2) Loading rate affects the walnut shucking force significantly. When the walnut is compressed at the loading rate of 50mm/min, the needed shucking force is the minimum, and then is the loading rate of 10mm/min. The needed force is the maximum when the walnuts are compressed at the loading rate of 30mm/min.

(3) Moisture content has significant effects on the shucking force, and with the extension of drying time, the moisture content of walnut decrease gradually, which results in the increases of the needed shucking force.

(4) Particle size has no great effect on the shucking force, but great difference in particle size usually corresponds to different walnut shucking force.

(5) The cracking forms of walnut shell are of two main types, namely, cracking longitudinally and cracking transversely. When the walnuts are horizontally and vertically placed and applied with a load, the longitudinal crack usually occurs. The transverse crack occurs only when the walnuts are placed on their sides.

References

1. Jianxin, S., Haijun, Z., Dongjun, X.: Technology for breaking walnut shell based on finite element analysis. Transactions of the CSAE 21(3), 185–188 (2005)
2. Yuande, D., Jianxin, S., Yuanyuan, Q.: Effect of the shell-breaking and kernel-taking about walnut for various shell-breaking methods. Academic Periodical of Farm Products Processing 9, 4–9 (2007)

3. Jianxin, S., Dongjun, X.: Discussions on the status and problems of the walnut shellers at home and abroad. Agriculture machinery of Xinjiang 6, 29–32 (2001)
4. Yichuan, H., Jianxin, S.: Analysis and Experiment on Mechanical Characteristic of Walnut Shell. Journal of Xinjiang Agricultural University 32(6), 70–75 (2009)
5. Lihua, Z., Wen, Z., Wen, Q., Mingxia, L.: Mechanical Characteristics of Peanut Cracking. Food Science 31(13), 52–55 (2010)

Research on the Effects of Mechanical and Physical Characteristics on Peanut Shucking

Hongmei Xu, Shuiping Yan, and Qi Huang

College of Engineering, Huazhong Agriculture University, Wuhan 430070, China
xhm790912@163.com

Abstract. Shucking is one of the necessary parts for the deep process of peanut, and has great impact on the quality of peanut kernel.Taking the shucking force as the evaluation indicator, the effects of mechanical and physical characteristics, which include the loading orientation, loading rate, moisture content, partical size and the number of peanut kernel, on peanut shucking were investigated by means of single factor test. The results show that: 1) the loading orientation and moisture content affect the shucking force significantly, and the needed force is the minimum when the peanut is vertically placed and loaded. With the decrease of moisture content, the needed force decreases gradually.2) the loading rate, particle size and the number of peanut kernel have no great effect on the shucking force, but the deformation and maximum shucking force vary with the loading rate. Additionally, the needed force increases with the number of peanut kernel. The research results can provide some reference for the designment of peanut sheller and selection of the technique parameters.

Keywords: Peanut, Shucking, Mechanical Characteristic, Physical Property.

1 Introduction

In recent years, with the great increase of peanut yield, the peanut foods have developed in the direction of deep processing. Shucking is one of the necessary steps for the deep process of peanut, which usually results in the damage and crack, and determines the quality of peanut kernel [1]. Nowadays, because of the unreasonable force-applying forms, most peanut shellers used in domestic are of larger breaking rate and behave themselves with poor shucking quality [2].

Mechanical and physical characteristics can provide fundamental data for developing the sheller and realizing the effective separation of peanut shell and meat. In order to design the highly effective and low damage sheller, mechanical and physical characteristics of peanut must be first investigated.

Take the peanut of E-hua NO.6 as the research object, and the compression test was conducted on the universal material testing machine. The effects of the loading orientation, loading rate, moisture content, partical size and the number of peanut kernel on the peanut shucking force were investigated by means of single factor test. The research results are of great significance for improving the working performance of peanut sheller and designing the new and high-efficiency shucking equipments.

2   Materials and Methods

2.1   Materials

The sold E-hua NO.6 peanut was selected as the specimen, and the experiment has been finished in the comprehensive laboratory of College of Engineering, Huazhong Agriculture University, in May 2011.
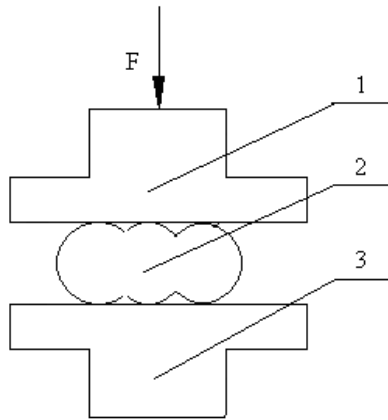
1. Upper platen   2.Specimen   3. Bottom platen

Fig. 1. Sketch of the loading device

2.2   Devices

The experiment was conducted on the universal material testing machine of RGT2000-10, and the loading rate varies from 0.01mm/min to 500mm/min. Fig.1 shows the sketch of the loading device. The pressure imposed on the peanut specimen by the plate-type pressure head is measured by the force transducer.

2.3   Methods

The experiment was conducted on the universal material testing machine, and the specimen was applied with the static load. In the beginning of the experiment, put the specimen on the bottom platen of the testing machine, and adjust the location of the upper platen to make sure it contacts with the specimen and imposes a compression force on the specimen. In order to determine the influence of the mechanical and physical characteristics on peanut shucking, the single factor test was conducted to investigate the effects of the loading orientation, loading rate, moisture content, particle size and the number of peanut kernel on the peanut shucking force. Additionally, for each experimental condition, the experiment was repeated ten times to reduce the test tolerance, and the result is the mean value of 10 tests.

3   Results and Discussion

3.1   Analysis of the Cracking Process of Peanut Shell

3.1.1   Cracking Forms of Peanut Shell

Fig.2 shows the cracking forms of peanut shell when the peanuts are applied with loads in different directions. The cracking forms of peanut shell are of two main types, namely, cracking longitudinally and cracking transversely. Because of the structural inhomogeneity and discontinuity of the peanut ridge, it features low compression strength. When the peanut is horizontally placed and applied with a load, the stress concentration and longitudinal crack usually occur. The longitudinal crack propagates in the direction of the peanut ridge, and plays an important role in the peanut shucking. When the peanut is vertically placed and applied with a load, the cracking form is similar to that of peanut shell when it is horizontally placed. The crack mainly concentrates on the end and ridge of peanut. However, when the peanut is placed on its side, as shown in Fig.2 (b), and applied with a load, the transverse crack occurs, and the peanut cracks from the loading point out to the outward, which usually brings about a depression in the loading point [3].

(a) Cracking longitudinally   (b) Cracking transversely
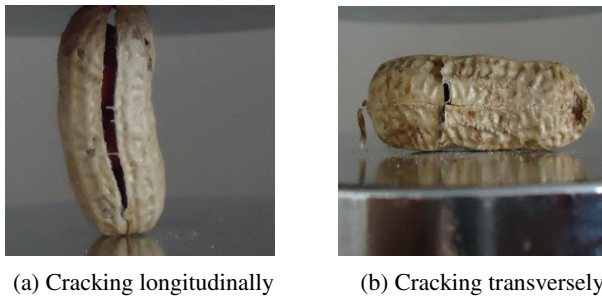
Fig. 2. Crack propagation of peanut shell

3.1.2   Cracking Process of Peanut Shell

Fig.3 shows the relation curve between the loading force and deformation of peanut when it is horizontally placed and applied with a load. As shown in the figure, the cracking process of peanut shell mainly includes three stages, namely, elastic deformation, plastic deformation and sudden rupture. The elastic deformation mainly refers to the early loading stage, in which the pressure linearly increases with the deformation of peanut. Afterwards, with the increase of loading force, the elastic deformation goes hand in hand with plastic deformation. The plastic deformation mainly focuses in the later loading stage. However, when the loading force achieves the peak value for the first time, the peanut begins to crack and the force declines sharply. The peak value mentioned above mainly depends on the compressive strength of peanut, so it's also known as the peanut shucking force. Then, the peanut shell and meat is compressed together, and the required loading force increases again [3,4]. In the course of the experiment, in order to protect the testing machine, the experiment is called for a halt once the peanut begins to crack and the loading force achieves the maximum.
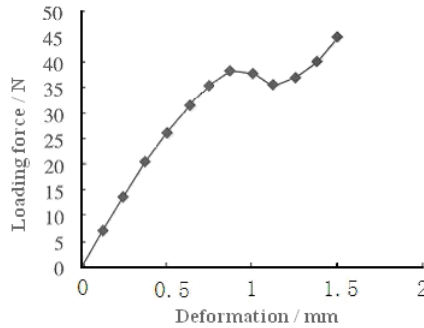
Fig. 3. Relation curve between the loading force and deformation of peanut

3.2   Analysis of the Influence Factors of Peanut Cracking

3.2.1   Effects of Loading Orientation on Peanut Cracking

Select thirty medium-size dried peanuts as the specimen, which is coincided with each other in the size and moisture content, and each peanut contains three kernels. Divide the peanuts into three groups, and each group contains ten peanuts. In the course of the experiment, the ten peanuts in group 1 are all horizontally placed as shown in Fig. 4(a).The peanuts in group 2 and group 3 are placed on their sides and vertically placed respectively. Compress the three groups of specimens at the loading rate of 10mm/min, and don't cease the experiment until the peanut shell begins to crack, which results in the great reduction of pressure. Record the peak value of the pressure, and regard it as the peanut shucking force.

(a) Peanut is horizontally placed (b) Peanut is placed on its side (c) Peanut is vertically placed

Fig. 4. Loading orientation of Peanut

Fig.5 shows the testing results of the peanut shucking force when the peanuts are applied with loads in different directions. As shown in the figure, the needed shucking forces are clearly different when the peanuts are applied with loads in different directions. When the peanut is vertically placed, the needed force is the minimum, which is equal to 19.3N, and then is the needed force when the peanut is horizontally placed. The needed force is the maximum when the peanut is placed on its side. The reason is that the load is mainly concentrated on the end and ridge of peanut when the

peanut is vertically placed and applied with a load. Because of the smaller stressed area, the stress is larger, which makes the peanut crack easily. When the peanut is horizontally placed, the load mainly acts on the ridge of peanut. Because of the structural inhomogeneity and discontinuity of the peanut ridge, the stress concentration and longitudinal crack with definite direction usually occurs, which contributes much to shell shucking. However, when the peanut is placed on its sides, the load is mainly concentrated on the middle part of peanut, and the larger stressed area results in smaller stress, which works against the shell shucking of peanut [4]. Variance analysis was conducted for the testing results of peanut shucking force, and the corresponding F-statistic is equal to 77.26, which is greater than the critical value $F_\alpha(2, 27)$ (α=0.05).The variance analysis results are consistent with the testing results. It is, therefore, believed that the loading orientation has great impact on the peanut shucking force.
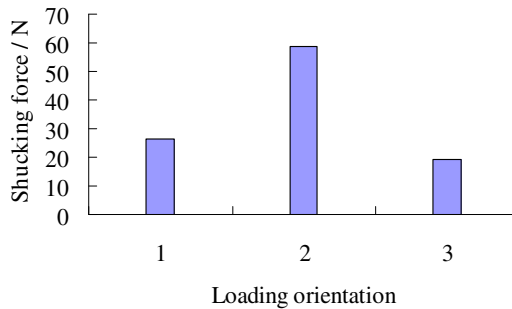
1: Peanut is horizontally placed
2: Peanut is placed on its side
3: Peanut is vertically placed

Fig. 5. Shucking force of peanut with different loading orientation

3.2.2   Effects of Loading Rate on Peanut Cracking

Select forty medium-size dried peanuts as the specimen, which is coincided with each other in the size and moisture content, and each peanut contains three kernels. Divide the peanuts into four groups, and each group contains ten peanuts. In the course of the experiment, the forty peanuts are all horizontally placed. Compress the four groups of specimens at the loading rate of 10mm/min, 20mm/min, 30mm/min and 40mm/min respectively.

Fig. 6 shows the testing results of the peanut shucking force when the peanuts are compressed at different loading rates. As shown in the figure, when the peanuts are compressed at different loading rates, the needed shucking forces are different, but the difference is not large. When the peanut is compressed at the loading rate of 10mm/min, the needed force is the minimum, which is equal to 26.1N, and then is the loading rate of 20mm/min. The needed force is the maximum when the peanut is compressed at the loading rate of 30mm/min and 40mm/min. Variance analysis was conducted for the testing results of peanut shucking force when the peanuts are

compressed at different loading rates, and the corresponding F-statistic is equal to 0.44, which is less than the critical value $F_{\alpha}(4, 35)$ (α=0.05).So it can be concluded that the loading rate has no great effect on the peanut shucking force.
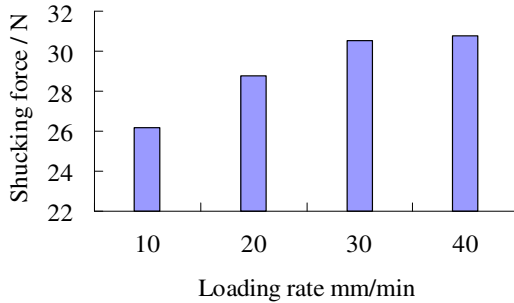
Fig. 6. Shucking force of peanut with different loading rate

3.2.3   Effects of Particle Size on Peanut Cracking

Select thirty dried peanuts as the specimen, which is coincided with each other in the moisture content, and each peanut contains three kernels. Divide the peanuts into three groups, and each group contains ten peanuts. The ten peanuts in group 1 with larger size range from 19mm to 21mm in diameter .The ten peanuts in group 2 are of medium size, and range from 14mm to 16mm in diameter. The ten peanuts in group 3 are the smallest ones, and range from 9mm to 11mm in diameter. In the course of the experiment, the thirty peanuts are all horizontally placed, and compressed at the loading rate of 10mm/min.

Fig.7 shows the testing results of the shucking force of peanut with different particle size. For the peanuts with different particle size, the needed forces are different, but the difference is not large. The needed shucking force of the large-size peanuts is the minimum, which is equal to 24.6N, and then is the smaller ones.
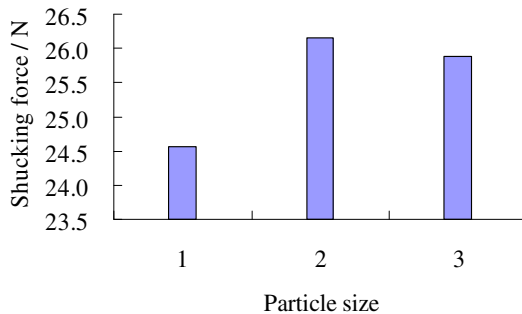
1: Peanut is of larger size
2: Peanut is of medium size
3: Peanut is of smaller size

Fig. 7. Shucking force of peanut with different particle size

The needed shucking force of the medium-size peanuts is the maximum. Variance analysis was conducted for the testing results of the shucking force of peanut with different particle size, and the corresponding F-statistic is equal to 0.15, which is less than the critical value $F_\alpha(2, 27)$ (α=0.05). It can be concluded that particle size has no great effect on the peanut shucking force.

3.2.4   Effects of Peanut Kernel Number on Peanut Cracking

Select forty dried peanuts as the specimen, which is coincided with each other in the moisture content. Divide the peanuts into four groups, and each group contains ten peanuts. The peanuts in group 1 and group 2 contain one and two kernels respectively. The peanuts in group 3 and group 4 contain three and four kernels respectively. In the course of the experiment, the four groups of peanuts are all horizontally placed, and compressed at the loading rate of 10mm/min. Fig. 8 shows the testing results of the shucking force of peanut with different kernel number. As shown in the figure, different kernel number corresponds to different shucking force, but the difference is not large. The needed shucking force of the one-kernel peanuts is the minimum, which is equal to 21.3N.Furthermore, the needed force increase with the increase of the peanut kernel number. The needed shucking force of the four-kernel peanuts is the maximum. Variance analysis was conducted for the testing results of the shucking force of peanut with different kernel number, and the corresponding F-statistic is equal to 2.19, which is less than the critical value $F_\alpha(3, 36)$ (α=0.05). It can be concluded that kernel number has no great effect on the peanut shucking force.
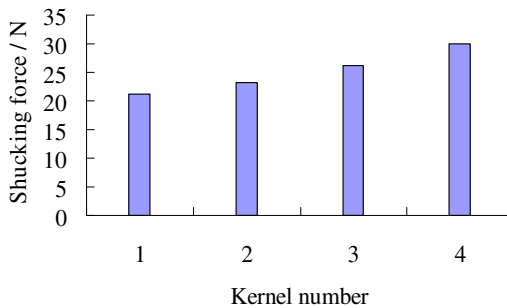
Fig. 8. Shucking force of peanut with different kernel number

3.2.5   Effects of Moisture Content on Peanut Cracking

Select fifty medium-size dried peanuts as the specimen, and each peanut contains three kernels. Divide the peanuts into five groups, and each group contains ten peanuts. In order to investigate the effect of the moisture content on the peanut shucking force, the peanuts must be pretreated before the compression test. The drying experiment is one of the most important preprocessing methods, which is usually used to control the moisture content of the samples. In this research, the drying experiments were conducted for five groups of walnut specimens, and the

corresponding drying time is 0h, 0.5h, 1h, 1.5h and 2h respectively. Afterwards, the peanuts are all horizontally placed, and compressed at the loading rate of 10mm/min.

Fig.9 shows the testing results of the shucking force of peanut with different drying time and moisture content. As shown in the figure, different drying time corresponds to different peanut shucking force. When the drying time is 2h, the needed shucking force is the minimum, which is equal to 21.9N, and then is the peanuts with the drying time of 1h and 1.5h. The needed shucking force of the peanuts with the drying time of 0h is the maximum. It can be seen that the moisture content of peanuts decreases with the increase of drying time. As a result, the needed shucking force decreases which makes peanut shucking easier. The reason is that the lower moisture content corresponds to greater brittleness, which brings about the decrease of the peanut anti-force capability [5].Variance analysis was conducted for the testing results of the shucking force of peanuts with different moisture content, and the corresponding F-statistic is equal to 10.72, which is greater than the critical value $F_{\alpha}(4, 45)$ (α=0.05). Therefore, it can be concluded that the drying time and moisture content affect the peanut shucking force significantly.
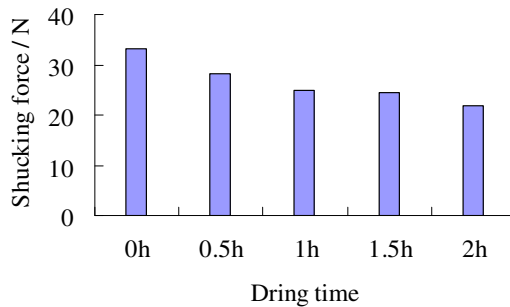
Fig. 9. Shucking force of peanut with different moisture content

4   Conclusions

Take the shucking force as the evaluation indicator, and the effects of the loading orientation, loading rate, moisture content, particle size and peanut kernel number on peanut shucking were investigated by means of single factor test. The conclusions can be described as follows:

(1) Loading orientation has great impact on the peanut shucking force. When the peanuts are vertically placed, the needed shucking force is the minimum, and then is the peanuts horizontally placed. The needed shucking force of the peanuts placed on their sides is the maximum.

(2) Moisture content affects the peanut shucking force significantly, and with the increase of drying time, the moisture content of peanut decrease gradually, which results in the decreases of the needed shucking force.

(3) Loading rate, particle size and number of peanut kernel have no great effect on the shucking force, but different loading rate and particle size correspond to different shucking force. Additionally, the needed force increases with the number of peanut kernel.

(4) The cracking forms of peanut shell are of two main types, namely, cracking longitudinally and cracking transversely. When the peanuts are horizontally and vertically placed and applied with a load, the longitudinal crack usually occurs, which makes peanut shucking easier. The transverse crack occurs only when the peanuts are placed on their sides.

References

1. Shuqi, S., Shuguang, L., Fangyan, W.: Current Situation and Development of Peanut Production Machinery. Transactions of the Chinese society for Agricultural Machinery 36(3), 143–147 (2005)
2. Jiandong, L., Shuqi, S., Xizhen, L., et al.: Current Situation and Development of Peanut Shelling Machine and Equipment in China. Journal of Peanut Science 35(4), 23–27 (2006)
3. Hongli, L., Yongli, Z., Lianxing, G., et al.: Experimental Research of the Shelling Force Performance of Groundnuts. Journal of Shenyang Agricultural University 37(6), 900–902 (2006)
4. Lihua, Z., Wen, Z., Wen, Q., et al.: Mechanical Characteristics of Peanut Cracking. Food Science 31(13), 52–55 (2010)
5. Xianli, C., Lianxing, G., Mingguo, L., et al.: Experimental Study on Peanuts Mechanical Performance under Impacting Test. Journal of Shenyang Agricultural University 40(1), 111–113 (2009)

Preliminary Study on Soil Moisture Forecast of Winter Wheat[*]

Xiumei Zhang[1,2], Yuchun Pan[2], Bingbo Gao[2], Chunshan Su, and Jihua Wang[2,**]

[1] School of Geography and Remote Sensing Science, Beijing Normal University, 100875, Beijing, P.R. China
Xiumei_zhang@163.com
[2] Beijing Research Center for Information Technology in Agriculture. 100097, Beijing, P.R. China
wangjh@nercita.org.cn
3 Beijing MAG Agricultural Science & Technology Development Co, LTD. 100089, Beijing, P.R. China
Mofais1987@163.com

Abstract. Using four groups of survey data of soil moisture of winter wheat from returning green stage to harvest stage in Daxing region, the variation laws of soil moisture exponential depletion relations in 0-60cm depths during different growth stage are analyzed ,and based on soil moisture exponential depletion relations the forecast model of soil moisture depletion is built. The results indicate that the soil moisture exponential depletion relations in the same depth layer increase at first and then decrease. Moreover, for different depth layers they decrease with the increase of depth, and the dispersion degree also changes with the depth. Using the exponential depletion relations method, the soil water storage at the 20, 40 and 60cm depth layers are forecasted, and the minimum forecast error is 0.09%, the forecast accuracy in the layer of 60cm is highest, and the absolute error is between 0-1.5%. This method does not need much data, and its results are reliable and precise. So it is a practical method for soil moisture forecast in winter wheat field of Beijing and others whose soil condition is same as Beijing.

Keywords: Winter wheat, Exponential depletion relations, Soil moisture, Forecast method.

1   Introduction

Winter wheat is a main grain in China which is planted in Northern part of China where is lack of water. The soil moisture forecast which is the main contents of the field moisture forecast is the best method to understand the dynamic variation rule and to adjusting ways of the soil moisture of the farmland soil [1]. Also it is meaningful for the reasonable distribution of the irrigation system and improving the utilization rate of

---

water resources. Besides the soil type, there are lots of factors which affect the soil moisture, such as precipitation, temperature, air moisture, wind velocity, plant type, different growing periods and so on. By now many research methods for soil moisture forecast have been put forward, such as Soil-Moisture-Balance Model [2,3], soil hydrodynamics model[4], remote sensing monitor method[5,6], nerural network model[7,8], empirical formula method[9,10], and exponential depletion model[11]. Soil-Moisture-Balance Model, soil moisture dynamics model which is simple are based on soil moisture analysis equation and the principle of soil moisture dynamics to forecast future soil moisture variation. They are simple and are often used to calculate the change of the soil moisture when the needed factors and boundary conditions are satisfied they can give out satisfactory result. However, the fact that they need lots of factors and boundary conditions which are hard to get prevents the popularization of these models. Remote sensing monitor method is to build statistical model between statistics of soil moisture and factor (Thermal inertia, normalized difference vegetation index) affecting the soil moisture and soil moisture using remote sensing images to forecast the soil moisture. It can give out the change of the soil moisture in time, but the stability of model need to be researched further. The parameters of soil and planting in neural network model always have no physical significance and can't reflect the physical relationship between input parameters and soil moisture. Exponential depletion model can analyze the soil moisture depletion law by directly using the observed soil moisture data. It is simple and need less factors, but can produce exact forecast result. To give out an easy, reliable, exact and feasible method for winter wheat soil moisture forecast, this paper studied the soil moisture variation rule along with time of different depth layers based on exponential depletion model, tested the results with observed data, and discussed the applicability of exponential depletion model. The monitoring data from monitoring site of winter wheat soil moisture and meteorological site of Daxing in Beijing in 2009 was used in the study.

2   The Monitor and Frequency for Soil Moisture

2.1   The Method of Soil Moisture Monitor

The volumetric water content of soil is measured using ECH2O which is a ruler like soil moisture probe produced in America. ECH2O sends voltage signal and feedback voltage signal from soil to calculate the dielectric permittivity from which the soil moisture is calculated. It has the features of lower power consumption, higher resolution, less effected by temperature and salinity, simpler installation and use, solider design, longer buried.

2.2   Soil Moisture, the Content and Frequency of Meteorological

The monitoring data from monitoring site of winter wheat soil moisture and meteorological site of Daxing in Beijing in 2009 include soil moisture 0~20cm, 20~40cm, 40~60cm, 80~100cm and meteorological data such as daily precipitation, evaporation, temperature and so on.

3   Method and Model for Soil Moisture Forecast

Soil moisture is decided by lots of factors like Climate, soil feature, plant suction and so on. During wheat growing periods in winter, it rarely rains and it hard to form runoffs. Soil water balance can be describe as below:

$$W_E = W_S + P + I - ET - Q \tag{1}$$

W_E and W_S in the equation represent water storage in one soil layer at the very ending and very beginning respectively; P represents the effective amount of precipitation during this period, mm; I represents the amount of irrigation, mm; ET represents the amount of plant evaporate and transpiration, mm; Q represents the water flux of lower boundary.

W_E, W_S can be calculated from observed soil moisture; P and I can be gotten from meteorological site; Q is much less than ET. Under soil water stress, ET and W has linear relationship, so in period there is no precipitation and irrigation, the relationship between ET and W can be described as:

$$ET = dW / dt = -kW \tag{2}$$

Where K is soil moisture exponential depletion coefficient, it is affected by meteorological, underlying surface, and growth phase of crop.

The soil moisture exponential depletion during no precipitation and irrigation period can be gotten by integrating the formula above within $t_E - t_S$.

$$W_E = W_S e^{-k(t_E - t_S)} \tag{3}$$

If there is precipitation and irrigation during (tS ~tE), soil moisture exponential depletion can be separate to two stages.

$$W_E = (W_S e^{-k\Delta t'} + P + I) e^{-k(t_E - t_S - \Delta t)} \tag{4}$$

Where $\Delta t'$ represents soil moisture exponential depletion time before raining or irrigating. The key to use depletion index forecasting soil moisture is to identify the value of depletion index. For some special area the soil feature is constant, and the change of meteorological elements have annual periods, depletion index is mainly is mainly affected by the species and growth season of plant. If there is no precipitation and irrigation, depletion index is:

$$k = (\ln W_S - \ln W_E)/(t_E - t_S) \tag{5}$$

4   Experimentation Design and Results Analysis

4.1   Experimentation Design

Experimentation site which belongs to Weishan village of Daxing region locates at E116°25, N39°37′. It is part of alluvial plain formed by ancient canal. Its altitude is between 20m and 40m and its annual precipitation is approximately 489.9mm.The annual average temperature is 11.4°C. Its prevailing wind directions are south and north. The underlying surface is comparatively flat and the depth of its groundwater is 3.3m. The soil texture is siltloam and the soil dry bulk density is 1.4g/cm3. This paper use the soil moisture data of the Experimentation site from winter wheat turning green to winter wheat harvesting (March 5th to June 6th, counting 93 days).

The experimental area was separated into six small areas and was planted with winter wheat. They are numbered as 1A, 2A, 1C, 1B, 2B and 2C from east to west (shown in tab.1). 1A, 2A,1B and 2B are used to calculate the water storage of 20,40 and 60cm which are used to build models. 1C and 2C are used to test of the accuracy of the model.

Table 1. Small area charts of the experimentation site

1A	2A	1C	1B	2B	2C
For prediction models		For test	For prediction models		For test

4.2   Result of Analysis

4.2.1   The Law of Exponential Depletion According to Depth

According to the observation data and the formula of exponential depletion, we can get exponential depletion on 20cm, 40cm, 60cm and 100cm depths. The law is below:
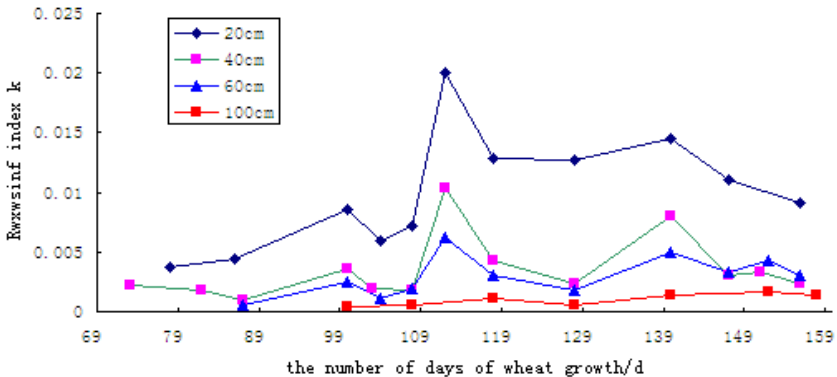
Fig. 1. The law of exponential depletion according to depth

(1) The results indicate that the soil moisture exponential depletion relations in the same depth layer show the trend of increasing at first and then decreasing. That is because when the winter wheat turn green from winter to spring, it needs least water, so soil moisture depletion is slow, and the exponential depletion gets small. When it goes into jointing stage, exponential depletion begins to get bigger. Exponential depletion gets biggest when it goes into heading stage in late April. When it goes into filling stage in middle May, the exponential depletion gets biggest again, and the soil moisture depletion gets fast. After that the consumption of water decreases, the exponential depletion gets small until the winter wheat gets ripe.

(2) With depth increasing, the exponential depletion gets small gradually. The water moisture is easy to be lost from surface layer to the layer of 20cm depth. It is also affected relatively large by evaporation, irrigation and rainfall. So the exponential depletion fluctuation is large. With the depth increasing, the influence of external conditions gets small gradually, so 40cm and 60cm change smoothly and there is nearly no measurable change in 100cm. That's because the crop evaporation and water absorption has little influence in deep soil layer.

4.2.2   The Analysis of Result for Soil Moisture
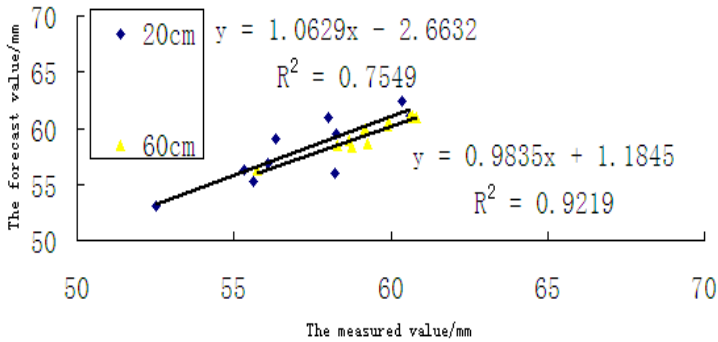
Fig. 2. The linear regression result of measured value and forecasting value in different soil layers for winter wheat

The roots of winter wheat are mostly located in 0-60cm soil layers, so we research 20cm, 40cm and 60cm soil depth layers. The initial soil moisture value was the average value when the winter wheat turnes green on March 5th for 1C and 2C. Having deleted outliers caused by rainfall and irrigation. or measuring errors .We forecasted the soil moisture of 20cm,40cm and 60cm soil layers using formula 3 for data without rainfall and irrigation ,otherwise using formula 4. And then used the measured data and forecasted data to do linear regression for the measured soil moisture and the forecasting value of the exponential depletion for 20, 40 and 60cm layers. The result is shown in Figure 2. The correlation index of the forecast model of exponential depletion in 20cm layer is 0.7549, the maximum forecast error on this layer is 5.02%, and the minimum error is 0.9%. By counting the relative error, the proportion of forecast exponential

depletion whose relative error is less than 20% is higher than 80% on 20cm .The forecasting value in 40cm is slightly bigger than that of 20cm and 60cm, maybe it is caused by the soil circumstance and the law root system pattern. The biggest forecasting error on this layer is 3.91%, and the smallest is 0.09%, the proportion of forecast exponential depletion whose relative error is less than 20% is higher than 90% in 40cm layer, The correlation index on 60cm layer is 0.9219, The biggest forecasting error on this layer is 1.14%, and the smallest is 0.31%, by counting relative error, the proportion of forecast exponential depletion whose relative error which is less than 20% is higher than 90% on 60cm. With the depth gets deeper, the effect of temperature and humidity gets smaller, so the forecasting error gets smaller, causing the forecasting result became exact. Several studies show that the result of the exponential depletion forecasting model is more accurate than empirical formula method whose error is between 15%-30% [12,13]. The mainly reason is the exponential depletion method of this paper not only taking the meteorological factors into account but also taking into account the difference of the exponential depletion caused by the difference of different growth period of plant, calculating several exponential depletion k of different growth periods of plant, using K in different period to forecast. However empirical models mainly consider those effects caused by rainfall, temperature and other meteorological factors for soil moisture to generate a multiple linear equations to forecast. So exponential depletion method is more suitable for soil moisture forecast of different periods and the forecast is more accurate. The exponential depletion can reflect water receding process of winter wheat field. Using exponential depletion to forecast soil moisture, the needed the factors only include time, rainfall and irrigation data, and the result is stable, reliable and accurate.

5   Conclusion and Epilogue

This article build the soil moisture forecasting exponential depletion model using the measured data in Daxing monitor site, through this article, we can get several conclusion:

(1) The changing trade of the soil moisture exponential depletion relations of winter wheat in each depth layer is same, that's increasing first then decreasing during the period, returning green stage to harvest stage. They get maximum when the winter wheat gets into ear emergence and filling stage, and after that they become smaller and smaller until the winter wheat is ripe. For different depth layers they decrease with the increment of the depth. They fluctuates frequently in 20cm depth layer, mainly caused by external factors. As the depth increasing, the influence of external factors gets less, and the exponential depletion fluctuates more gently.

(2) As the depth of soil increasing, the water storage gets more stable, the influence of external meteorological factor get smaller, and the forecast became more exact. The absolute error of exponential depletion of 60cm depth layer is between 0-1.5%, and the forecast accuracy is much higher. The forecast accuracy of other depth layers can also meet the practical requirements. It shows that the model is suit to the forecast of soil moisture of winter wheat.

(3) Due to different soil feature, meteorological factors, and plant water absorption in different area, the variation laws and area features of exponential depletion relations in other planting area and different planting conditions should be studied further.

References

1. Qiao, G.: Prediction Model of Soil Moisture for Arid areas in North. J. South-to-North Water Transfers and Water Science &Technology, 39–41 (2009)
2. Shen, H., Yan, C.: Progress and application of soil moisture monitoring and forecasting models. J. Ecological Science, 366–370 (2003)
3. Mahmood, R., Hubbard, K.G.: An analysis of simulated long-term soil moisture data for three land uses under contrasting hudroclimatic conditions in the northern gread plains. J. Hydrometerorology, 160–179 (2004)
4. Jiang, T., Liu, H.: Analysis of vertical variation characteristic of soil moisture in yellow soil sloping field. J. Agriculture Engineering Transaction, 6–11 (2005)
5. Li, Y., Xu, M., Tang, Y.: The Status and Advances of Thermal Inertia Models in Measuring Soil Moisture by Remote Sensing in China. J. China Agriculture Weather, 40–43 (2000)
6. Weimann, A.: Soil moisture estimation with ERS-1 SAR data in the East-German loess soil area. J. Remote Sensing, 237–243 (1998)
7. Li, F., Shi, Y., Sun, K., Fu, Z.-T., Liang, A.: Sturdy on Forecasting Soil Moisture Content Based on BP web. Journal of Laiyang Agricultural College, 136–140 (2006)
8. Lippmann, R.P.: An introduction to computing with neural network. J. IEEE ASSP Magazine 22 (1987)
9. Kong, F., Liu, J., Zhang, C., et al.: Analysis of Soil Moisture Changes in Southwestern Shandong Province. Chinese Journal of Agrometeorology, 162–165 (2008)
10. Su, R., Kang, S.-Z., Jia, Y.-M., Zhang, B.-Z., Wei, Y.-B.: Preliminary Study on Soil Moisture Forecast in Fen River Irrigation District, 92–95 (2005)
11. Pu, S., He, X., He, C., Zhang, Z.: Study on Soil Moisture Forecast Method Based on Exponential Depletion Relations for Cotton Drip Irrigation under Mulch in Xinjiang Autonomous Region. Water Saving Irrigation, 5–8 (2008)
12. Ren, G.: The Empirical Model for the Forecast of Soil Moisture in Fenhe Irrigated Area. J. Sci-Tech Information Development &Economy, 88–90 (2008)
13. Shang, S., Lei, Z., Yang, S.: Empirical Model for Soil Moisture Forecast in Winter Wheat Field. Transactions of the Chinese Society of Agriculture Engineering, 31–33 (2000)

Qualitative Analysis of Age and Brand of Unblended Brandy by Electronic Nose

Yang Yang[*], Yu Zhao[*], Shuming Zhang, Yuanying Ni, and Jicheng Zhan[**]

College of Food Science and Nutrient Engineering, National Engineering Research Center for Fruit and Vegetable Processing, China Agricultural University, Beijing 100083, China

Abstract. This paper reports the capability of electronic nose based on gas chromatograph (GC-Flash) in age identification and brand classification of brandy. Three different kinds of brandies by the static headspace sampling for the injection of the volatile compounds were analyzed. The data were disposed by multivariate data processing based on principal component analysis (PCA) and cluster analysis (CA). The results show that PCA could identify the age as PC1 represented the raw information of age well; PCA on small difference ages of the samples could classify the brands of the samples. The CA would cluster samples by age through its analysis on the three different kinds of brandies respectively; The CA on all samples indicated that significant differences existed among different brands with age under 20 years and clustering process performed firstly in each group by age information and then clustering progressively by age among all groups. GC-Flash electronic nose can be applied to identify the brand and age of different brandies.

Keywords: electronic nose, brandy, principal component analysis, cluster analysis.

1 Introduction

Grape brandy is a kind of alcoholic beverage, taking grapes as raw material and whose formation follows fermentation, distillation and storing in oak barrels [1]. The unblended liquor after wine aging is namely original wine, which in the storing undergoes complicated physical, chemical and biological changes. The assessment by chemical methods and sensory evaluation is highly time-, labour- and capital-consumed; therefore the age of wine is regarded as a significant index to determine the quality and class of brandy [2]. With the development of detecting instruments, the rapid analysis technology (electronic nose, electronic tongue, electronic eyes and near-infrared spectroscopy, etc.), based on rapid analysis of the total sample information ("fingerprint"), is gradually applied to rapid identification of wine age, quality, origin and brand.

Electronic nose comprises an array of electronic sensors with partial specificity and an appropriate pattern-recognition system, capable of recognizing simple or complex

---

[*] Yang Yang and Yu Zhao contributed equally to this work.
[**] Corresponding author.

odors [2]. Electronic nose technology is mainly applied in the food industry in quality control, process monitoring, freshness evaluation, shelf life prediction, authenticity assessment, etc.[2].Researches in fermentation monitoring[3-7], identification of brand[8-10] and origin[11-13], aroma evaluation [13-16] and false distinguishing[17] by electronic nose have been carried out at home and abroad with excellent performance. Electronic nose, as a rapid qualitative analysis instruments at first, after decades of development, has also made certain progress in quantitative analysis. The mass spectrometry-based electronic nose invented in the nineties [5, 18], stepped forward a great deal in quantitative analysis.

The former researches on wine analysis by electronic nose have focused on the sensor-based or mass spectrometry-based electronic nose, but seldom on the ones analyzing brandy, and even the research on testing original brandy has not been reported yet. This study, using gas chromatography-based electronic nose, carried out studies on the wine age and brand identification of 3 domestic and overseas brands for the basis of wider application of electronic nose.

2   Experimental

2.1   Samples

The studies were performed on 29 brandies from three different brands (Group A, n = 12; Group B, n =13; Group C, n = 4). Group A, France Brandy, of which the ageing time is from 4 to 46 years old. Group B and Group C are made in China, whose ageing time are 2 to 15 and 1 to 5, respectively. Those samples (brandies) came from the same variety of grape (Ugni Blanc) and the same traditional brewing process applied in Cognac region in France. They are all raw brandies without blending after maturing in oak barrels. The samples were taken directly from the winery and then stored at 4℃. All samples were kept at 20℃for 2 h before testing.

2.2   Electronic Nose

Experiments were performed with the HERACLES Flash Electronic nose (Alpha M.O.S Company, France) consisted of a Heracles host machine and a headspace auto-sampler unit. The HERACLES is a highly selective and sensitive specialty gas chromatograph, capable of performing very fast, low level hydrocarbon measurements in laboratory or field environments. Key features include versatility and speed of analysis, giving the productivity advantages over the more traditional gas chromatographs. The HERACLES is a programmed temperature gas chromatograph using syringe or valve inlets to a flash evaporator. The sample is delivered to an adsorbent trap to concentrate the sample for delivery to twin capillary columns and flame ionization detectors (FID) simultaneously [19]. Trace level gas phase hydrocarbon samples down to a few parts per thousand million can currently be measured. Instrument control and the raw data process were performed by Alpha Soft (V11, Alpha M.O.S). A general scheme of the HERACLES Flash Electronic nose is presented in Fig. 1.
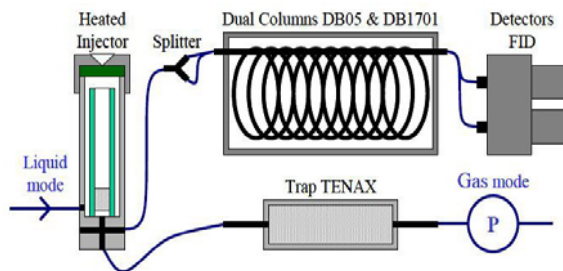
Fig. 1. Set-up applied in the measurement

In order to obtain maximum sensor response, operating parameters for the electronic nose were optimized for the brandy samples. The analysis parameters are as follows: (1) Injector: injection time 1500 ms, injector temperature 200 ℃, injection volume 2000 μL; (2) Temperature program: initial temperature 40 ℃, final temperature 200 ℃, heating rate 2.0 ℃/s; (3) Detector temperature: 220 ℃; (4) Trap temperature: 250 ℃, purge time 15 s. (5) Acquisition time: 84 s.

For the measurements, 1000 μL of the sample was placed in a 10mL glass vial provided with a pierceable disk in the cap. Before gas sample injection, sample was agitated by the auto-sample unit at 40 (±0.2) ℃, 500 rpm. All samples were analyzed in 7 times and the average of the chromatograms was used for subsequent statistical analysis.

2.3   Data Analysis

Two pattern recognition methods, principal component analysis (PCA) and Cluster analysis (CA) have been used for data analysis. Principal component analysis (PCA) is a chemometric linear, unsupervised and pattern recognition technique used for analyzing and reducing the dimensionality of numerical datasets in a multivariate problem. It can observe the classification results of sample principal components through the PCA map. PCA was performed in Alpha Soft (V11, Alpha M.O.S, France). Cluster analysis (CA) is a kind of unsupervised pattern recognition. It is a method for grouping objects of similar kind into respective categories. CA of Samples was performed using SAS 9.2 software.

3   Results and Discussion

Detect and analyze Group A, B, C by electronic nose, eliminate abnormal data [20], obtain the peak figure (retention time and peak area) for each sample. After deducting the blank peak 24 peaks were obtained, including 12 peaks from gas spectrum column 1 and 12 peaks from column 2. Retention time and peak area of those 24 peaks are the "fingerprint" of static headspace volatile components which acquired by electronic nose, and also the raw data of principal component analysis and clustering analysis. For better understanding, the 24 retention time obtained by electronic nose can be defined as sensors while the corresponding peak area can be defined as the response value of sensors.

3.1   Principal Component Analysis Results

In this study, the PCA was firstly applied to group A. Using the raw data directly, three principal components were extracted: PC1-68.36%, PC2-9.90%, PC3-7.59%, The ellipses, which correspond to the 85.85% confidence intervals, are basically reflect the raw information [21]. Fig. 2 demonstrates the score plot of the first two principal components of Group A. It is easy to distinguish brandy by their different ageing time and furthermore, with the extension of ageing time, the trends appear obviously in PC1: on the PC1 axis from right to left. Meanwhile, samples of close age which are distributed concentrated in the Fig. 2 and can be divided into 4 groups. Information from PC2 is inefficient in distinguishing samples of 30, 36 and 39 years and the supplement from PC2 turns to be necessary.

The PCA analysis results displays the similar scene in Fig. 3, four principal components were extracted: PC1-45.32%, PC2-18.34%, PC3-15.19%, PC4-6.29%, and 85.04% of the total variance of the data is displayed in the plot. The PCA analysis on Group A and B show that electronic nose is able to distinguish the ageing time of two brandies respectively clearly and the collected first principle components represent ageing time well. The PCA analysis on Group C represent an even better results, however, because of the limited volume of samples, the results are not listed in this paper.
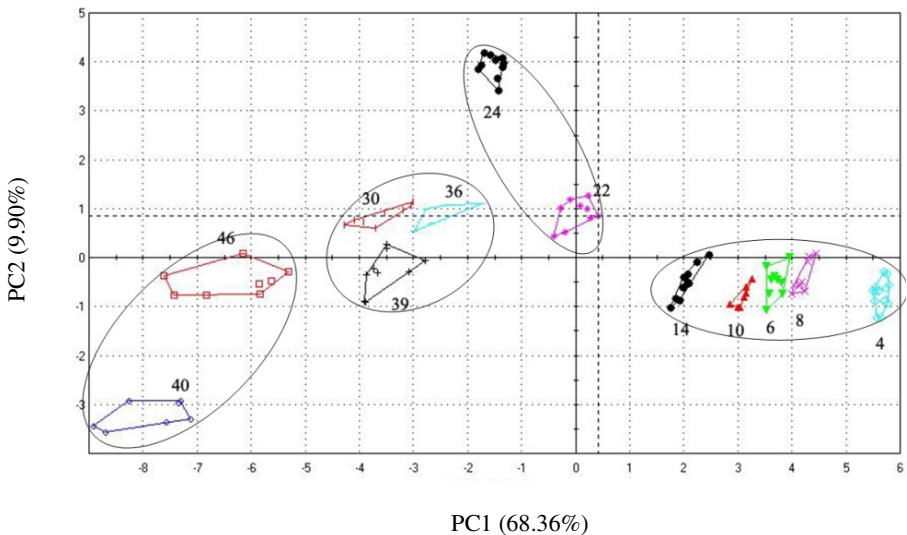
Fig. 2. Score plot of the first two principal components of Group A. The number 4, 6, 8,…,46 represents for the Brandies' ageing time.
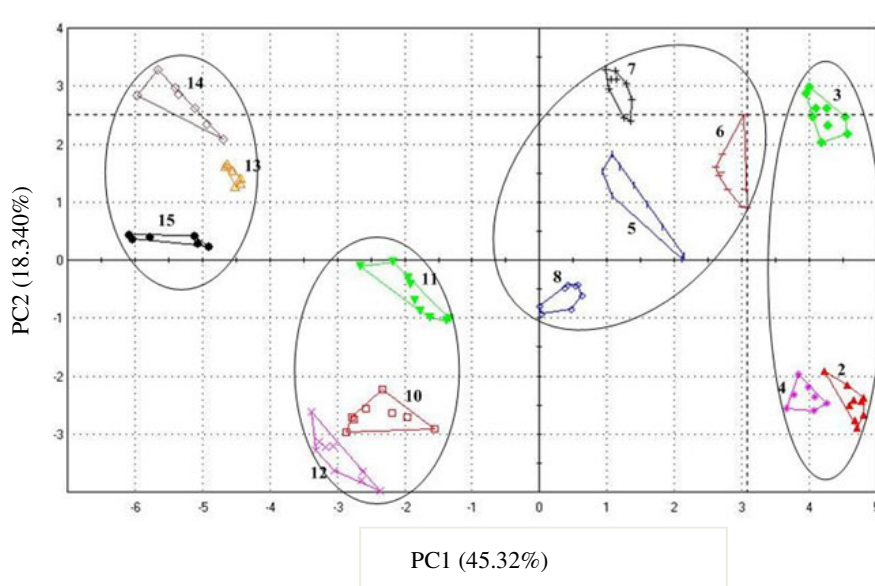
Fig. 3. Score plot of the first two principal components of Group B. The number 2, 3, 4,…,15 represents for the Brandies' ageing time.

In the analysis process, we optimized sensors (eliminate the small peak with poor reproducibility and strongly linear-correlated sensor) and screened 12 sensors out of 24 sensors. Finally extracted two principal components: PC1 -70.32%, PC2-19.20% and the cumulative contribution rate is 89.52%. Fig. 4 shows the score and loading map of PCA analysis, in which we can see that electronic nose can distinguish the age and brand of different design groups. PC1 axis indicates information of wine ages and PC2 axis indicates information of brands. But PC2 is insufficient to differentiate the aged wine in Group B and the wine in Group A, which, is possibly relevant to the difference of wine ages among samples. At the same time, the loading map shows the information of ageing time and brand of sensors (retention time and peak area of electronic nose). In order to study the brand-recognition capacity of electronic nose for original brandies, we designed analysis process including samples of the three brands. Due to the large difference among the 3 groups of wine, Group A (aged for 4 and 6 years) and Group B (aged for 2,3,4,5,6 years) and Group C(aged for 1,2,3,5 years) were chosen to conduct PCA . The peak value at 4.17s, 10.35s, 14.17s, 16.62s, 21.79s of Column gas spectrum 1 and peak value at 4.53s, 8.30s of Column gas spectrum 2 represent more brand information while the left 5 peak value represent ageing time information.

Fig. 5 is the score plot of PCA after decreasing the samples difference in wine age. The afterwards results illustrated a better effect in distinguishing , which means that ageing time difference will effect analysis results in brands distinguishing, overlarge difference in sample ages brings a great deal of difficulties to distinguishing brand information among different groups by electronic nose.
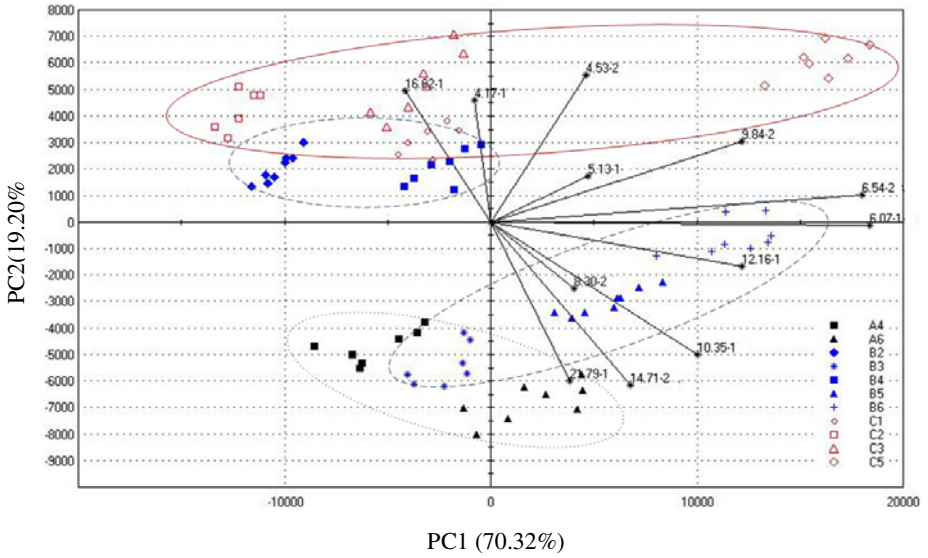
Fig. 4. Score and loading plots of PCA for Group A, B and C. The figure represents for the retention time of peak and the number of column. (e.g. 4.35-2 represents the peak value at 4.53s of Column gas spectrum 2).
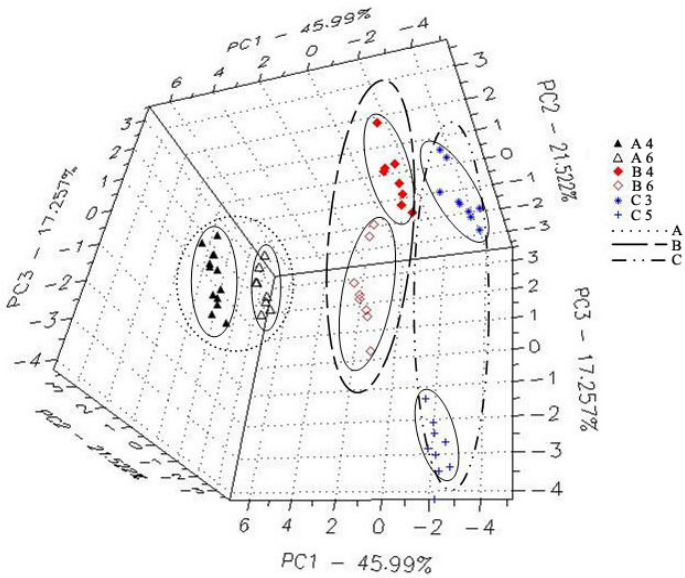
Fig. 5. Score plots of PCA for Group A, B and C

3.2   Clustering Analysis Result

In order to excavate more test results of the original brandy from the electronic nose, cluster analysis method was applied in Group A and equally in Group A, B and C respectively. Cluster analysis belongs to unsupervised pattern recognition, through which man can observe the effects of automatic classification and recognition of different ages and brands by electronic nose. Cluster analysis method is Group Average Method, and standardizes the data before the cluster analysis [22]. (each data point minus its average point in the sensor column then divided by the standard deviation in its sensor column)

Fig. 6 is the cluster result of Group A, the different age of the wine under the same brand, clustering of brandies of different brands was conducted according to the age of wine, firstly brandy of similar age can be divided into the same class, and then samples of different aging stage can be classified. The similar test result can be seen compared with the principal component analysis (Fig. 2) when the average distance of the classification is 5.0, three different types of the wine are less than 20 years, 20-30 year and 30-46 years.
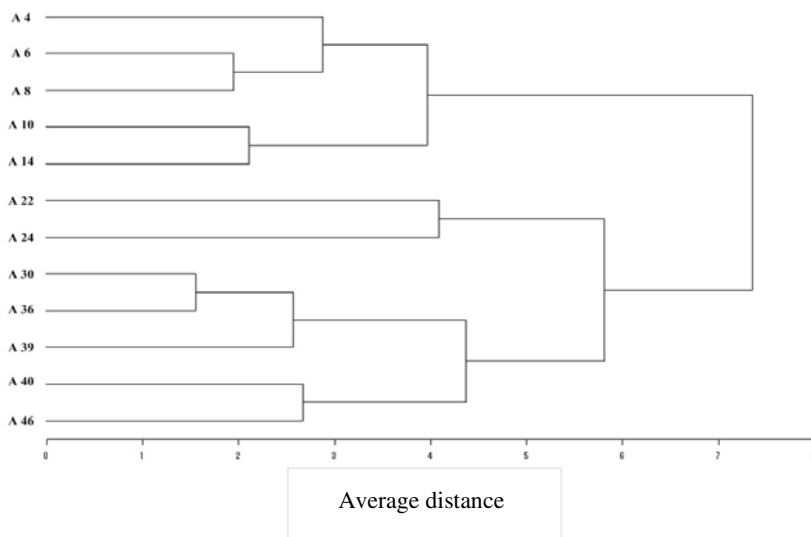
Fig. 6. Tree diagram from clustering analysis in Group Average method for Group A

Cluster analysis results also showed in a regular pattern after cluster analyzing from the Group A, B and C. As we can see in the Fig. 7, three different brands and ages of the brandies were mixed together, clustering firstly process in the same brand of similar age. When the average distance is 3.89, the cluster can be divided into four groups: Group A, 4-14 years, Group B, 2-6 years, Group B, 7-15 year and Group C, 1-5 years, the brand element is now in the dominant place which illustrated that aroma

components have significant differences when the age of the wine is under 20 years, electronic nose can distinguish the three brands. When the average distance is 5.43, the cluster can be assorted into three categories: Group A, B and C wine age less than 20 years, Group A, 22-24 years, Group A, 30-46 year, the component of the wine age placed the most important role during this time, the information reflected from electronic nose to different brand of the wine has been reduced.
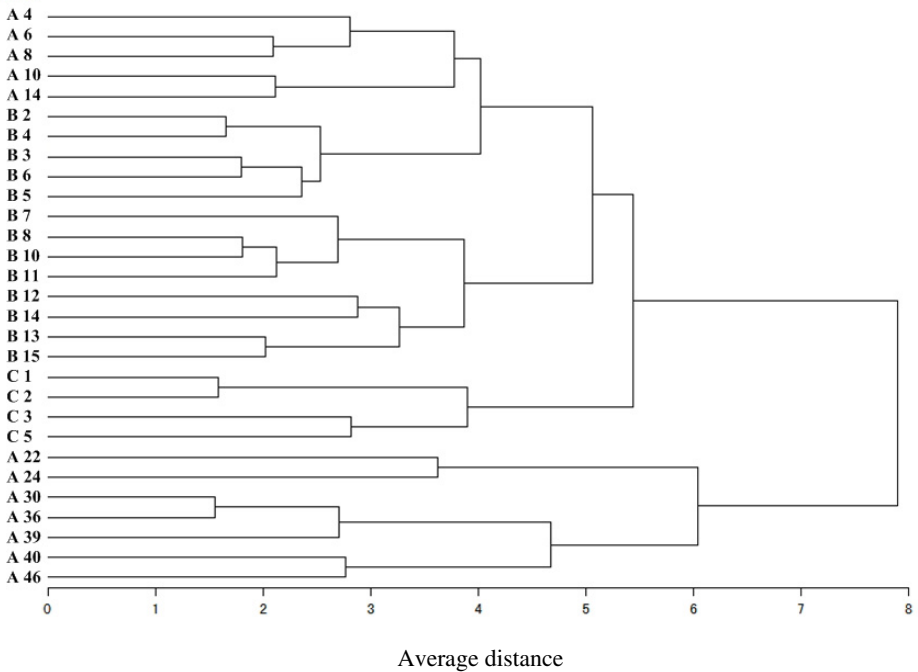
Fig. 7. Tree diagram from clustering analysis in Group Average method for Group A, B and C

4   Conclusions

GC-flash Electronic nose was used in present study to analyze 29 kinds of raw brandy with different ageing years. It illustrated the identification effect of electronic nose on raw brandy ages and brands, using principal component analysis and clustering analysis.

The principal component of wine sample A, B, C was analyzed. The principal component analysis represents the basic information of original data, and it also presents a good separation of brandy with different ages. The first principal component can reflect the wine age information. Principal component analysis also indicated that stage differences exist among the different brandy ages. Even though the sample age

difference is big, distance between the brandy samples with small age difference is close; thereby several "ages' stage" was obtained.

Due to the large age difference among three samples, brandy samples with close age differences were chosen to study the identification function of electronic nose among different brands. Although the difference of age exists in different brands, principal component could still be useful in the identification among different brands. In addition, principal component analysis can reflect certain age information. The results indicated that when electronic nose distinguishes the brands of raw brandy, they can accept some wine age differences.

The clustering analysis was conducted on Group A, B and C. It was a process of clustering wine according they ages, which aggregated the brandy with close ages. Clustering analysis was performed on three sample groups together. Clustering process presented same brand wine with close ages firstly. As for the samples with age less than 20 years, clustering according to different brands, brands differences were obvious. When the age gap was big (more than 20 years), brands identification was not good, but age difference became obvious. This result could be caused by the characteristics of brandy themselves. This phenomenon will make contribution to the use of electronic noses on brandy brands and ages identification. When a model was established, brandy ages should be taken into account. When electronic noses can identify wine age and wine brands correctly, the age difference between different brands could be allowed. However, model establishment should minimize the age differences, building models according to different brands wine with different ages respectively.

References

1. Jiang, Z., Li, J.: Study on relation between components and ages of brandy. J. Food and Fermentation Ind. 34(3), 137–140 (2008) (in Chinese)
2. Schaller, E., Bosset, J.O., Escher, F.: 'Electronic noses' and their application to food. J. LWT - Food Sci. Technol. Int. 31(4), 305–316 (1998)
3. Pinheiro, C., Rodrigues, C.M., Schfer, T., Crespoet, J.: Monitoring the aroma production during wine-must fermentation with an electronic nose. J. Biotechnol. Bioeng. 77(6), 632–640 (2002)
4. Pearce, T.C., Gardner, J.W., Friel, S., Bartlett, P.N., Blair, N.: Electronic nose for monitoring the flavor of beers. J. Analyst. 4(118), 371–377 (1993)
5. Kojima, H., Araki, S., Kaneda, H.: Application of a new electronic nose with fingerprint mass spectrometry to brewing. J. J. Am. Soc. Brew. Chem. 63(4), 151–157 (2005)
6. Lozano, J., Arroyo, T., Santos, J.P., Cabellos, J.M., Horrillo, M.C.: Electronic nose for wine ageing detection. J. Sens. Actuators, B 133(1), 180–186 (2008)
7. Marti, M.P., Busto, O., Guasch, J.: Application of a headspace mass spectrometry system to the differentiation and classification of wines according to their origin, variety and ageing. J. Chromatogr. A 1057(1-2), 211–217 (2004)
8. Garc, A.M., Aleixandre, M., Gutierrez, J., Horrillo, M.C.: Electronic nose for wine discrimination. J. Sens. Actuators, B 113(2), 911–916 (2006)
9. Pornpanomchai, C., Suthamsmai, N.: Beer classification by electronic nose. In: International Conference on Wavelet Analysis and Pattern Recognition, pp. 333–338. IEEE Press, Hong Kong (2008)

10. Lozano, J., Santos, J.P., Sayago, I., Gutierrez, J., Horrillo, M.C.: Identification of typical wine aromas by means of an electronic nose. J. IEEE Sens. 6(1), 173–178 (2006)
11. Cynkar, W., Dambergs, R., Smith, P., Cozzolino, D.: Classification of Tempranillo wines according to geographic origin: Combination of mass spectrometry based electronic nose and chemometrics. J. Anal. Chim. Acta 660(1-2), 227–231 (2010)
12. Aleixandre, M., Gonzalez, J.A., Sayago, I., Ferrnandez, M.J., Gutierrez, J., Horrillo, M.C.: Analysis of grape variety and denomination of origin of several wines with an artificial nose. In: Electron Devices, pp. 309–311. IEEE Press, Santiago de Compostela (2009)
13. Berna, A.Z., Trowel, l.S., Clifford, D., Cynkar, W., Cozzolino, D.: Geographical origin of Sauvignon Blanc wines predicted by mass spectrometry and metal oxide based electronic nose. J. Anal. Chim. Acta 648(2), 146–152 (2009)
14. Arroyo, T., Lozano, J., Cabellos, J.M.: Evaluation of wine aromatic compounds by a sensory human panel and an electronic nose. J. Agric. Food. Chem. 57(24), 11543–11549 (2009)
15. Ragazzo-Sanchez, J.A., Chalier, P., Chevalier, D., Calderon-Santoyo, M., Chommidh, C.: Identification of different alcoholic beverages by electronic nose coupled to GC. J. Sens. Actuators, B 134(1), 43–48 (2008)
16. Ragazzo-Sanchez, J.A., Chalier, P., Chevalier, D., Calderon-Santoyo, M., Chommidh, C.: Off-flavors detection in alcoholic beverages by electronic nose coupled to GC. J. Sens. Actuators, B 140(1), 29–34 (2009)
17. Guadarrama, A., Iniguez, M., Souto, J., de Saja, J.A., Feranadez, J.A.: Array of conducting polymer sensors for the characterisation of wines. J. Anal. Chim. Acta 411(1-2), 193–200 (2000)
18. Berna, A.Z., Trowell, S., Cynkar, W., Cozzolino, D.: Comparison of metal oxide-based electronic nose and mass spectrometry-based electronic nose for the prediction of red wine spoilage. J. Agric. Food. Chem. 56(9), 3238–3244 (2008)
19. Gao, Y., Liu, Y., Li, Y., Li, J., Shi, B., Ni, Y.: Electronic nose fingerprint of liquor of main flavor types. J. Liquor-Making Sci. Tech. 5, 38–41 (2008) (in Chinese)
20. Liu, H., Niu, Z.: Comparison of total volatile basic nitrogen detection models in fishmeal based in electronic nose. J. Chin. Soc. Agric. Eng. 26(4), 322–326 (2010) (in Chinese)
21. Liu, M., Oan, L., Tu, G., Liu, P.: Determination of egg freshness during shelf life with electronic nose. J. Chin. Soc. Agric. Eng. 26(4), 317–321 (2010) (in Chinese)
22. Aleixandre, M., Lozano, J., Gutierrez, J., Sayago, I., Fernandez, M.J., Horrillo, M.C.: Portable e-nose to classify different kinds of wine. J. Sens. Actuators, B 131(1), 71–76 (2008)

The Measurement of the Agricultural Informatization Contribution Rate Based on the Analysis of the Composite Index and the C-D Production Function

Liyong Liu and Daoliang Li

College of Information and Electrical Engineering, China Agricultural University, PRC.100083
P.O. Box 121, College of Information and Electrical Engineering, China Agricultural University,
17 Tsinghua East Road, Beijing, 100083, P.R. China

Abstract. The agricultural informatization contribution rate is the proportion of the contribution of agricultural informatization to the algebraic summation of agricultural economy growth accounting for the agricultural economy growth. It is a comprehensive index which reflects the function of agricultural informatization in enhancing agricultural productivity and agricultural economy development. This paper combines the analysis of the composite index with C-D production function to measure the contribution rate of agricultural informatization. Firstly this paper is going to establish the agricultural information evaluation index system including the infrastructure, resources, application, personnel, and policy environment. Then it is going to calculate each index and the composite index of agricultural informatization by means of synthetic rating analysis. Based on the above, this paper is going to build up a C-D production function model to measure the agricultural informatization contribution rate by regression function.

Keywords: Agricultural Informationalization, Contribution rate, C-D production function.

1 Introduction

The agricultural informatization contribution rate is the proportion of the contribution of agricultural informatization to the algebraic summation of agricultural economy growth accounting for the agricultural economy growth. It is a comprehensive index which reflects the function of agricultural informatization in enhancing agricultural productivity and agricultural economy development. (Liu, 2011). Its calculation method is shown as Formula (1):

$$\text{Agricultural informatization conbrituion rate(\%)}$$

$$= \frac{\text{the contribution of agricultural informatization to the algebraic summation of agricultural economy growth}}{\text{the overall growth of agricultural economy}}$$

$$\times 100\% \quad \text{Formula(1)}$$

However agricultural informatization is a complex system which has too many influencing factors to be calculated directly. Therefore it is essential to establish a feasible way to measure the agricultural informatization indirectly in order to meet

practical needs. Hence this paper is going to identify agricultural informatization indices, discuss the relation of agricultural informatization and agricultural economic growth, and finally set up the measure method for agricultural informatization.

2   Establishing the Agricultural Informatization Evaluation Index System

An index system scientific enough to truly reveal the object should be able to cover indices all-sidedly. Based on the comprehensiveness, scientificalness, objectivity, practicability, and systematicness principles of index, combining the practice of China's agricultural informatization and related statistical yearbooks, China's agricultural informatization evaluation index system should consist of five indices: the infrastructure (F), resources (R), application (A), personnel (H), and policy environment (E) and 19 sub-indices.

2.1   The Infrastructure of Agricultural Informatization (F)

The infrastructure of agricultural informatization is the hardware basis and neural system of agricultural informatization; it is also the precondition for the functioning of agricultural information resources and technologies; and the infrastructure of agricultural informatization is composed of traditional network (postal network, television and radio network, and telephone and mobile network), and modern information network (internet services and computer equipment). The infrastructure of agricultural informatization enables people to transmit information beyond geographical and temporal constraints and facilitates the collaboration through resource sharing. In this paper the infrastructure of agricultural informatization includes five sub-indices, which are the population coverage of radio, the population coverage of television, telephone density, mobile phone penetration rate, and family computers with internet access (Zhao, 2007).

2.2   Resources of Agricultural Informatization (R)

Resources of agricultural informatization are the content carried by the infrastructure of agricultural informatization. It involves document, information, diagrams and data, and demonstrates connections of various subjects and inherent laws. This index includes two aspects of document information content and digital information resources covering four sub-indices of agricultural book types, publishing newspaper and magazines, libraries and websites.

2.3   The Application of Agricultural Informatization (A)

Information technologies are essential to the development of informatization. Not only fundamentally change the form and content of human information activities, information technologies but also strong impact on people's lifestyle, behaviour

pattern, and thinking and production mode (TKularahte, 1999). Moreover the application of information technologies into the agriculture serves the purpose of agricultural informatization which is accelerating the development of agriculture, rural area and farmers. Therefore in this paper three sub-indices, the population of netizens in rural area, number of key agricultural scientific and technological achievements, and the aggregation of tele-service in rural area (Gao, 2009).

2.4   Personnel in Agricultural Informatization (H)

Personnel in informatization have a decisive role in informatization and is a significant condition securing the success of informatization construction. This paper studies this issue by looking at personnel in scientific activities, communication and transportation, education and culture, radio and television, computer service and software development. As a result this paper selects the population taking up scientific activities out of ten thousand people, the population in information transmission by computer service and software industry, the population in transportation, logistics and telecommunication industry, and the population in education, cultural, art, television and radio sector, as four sub-indices to evaluation the personnel in agricultural informatization (Yu, 2007).

2.5   The Policy Environment of Agricultural Informatization (E)

A favourable policy environment facilitates the informatization development as the informatization is a complex system which is influenced by the society, the state and the government (Li, 2009). Policies in papers refer to those laws, regulations, and policies made by legislatures, administrative authorities, and other related authorities in order to coordinate and accelerate the informatization, and those fund and supplies allocated by the government to support the informatization. Therefore this paper uses the proportion of S&T expenditure in GDP, local government fund for S&T, and the proportion of local government fund for S&T in the total expenditure of local government as three sub-indices to assess the government support to agricultural informatization.

3   Testing Indices of Agricultural Informatization

This paper measures the development of agricultural informatization by synthetic rating analysis. First of all, this paper uses Delphi method to determine the weight of each index and sub-index in the index system which is demonstrated by Table 1. All the corresponding data for indices in Table 1 can be traced in China Statistical Yearbook, China's Major S&T Indicators Database, and China's Internet Development Survey. However this paper, as a method study, is not going to put any data into the model for calculation.

Table 1. The evaluation index system of agricultural informatization development

Index	Sub-index	No.
1. The infrastructure of agricultural informatization	1.1 Population coverage rate of radio (%)	1
	1.2 Population coverage rate of television (%)	2
	1.3 Telephone density (%)	3
	1.4 Mobile phone penetration rate (%)	4
	1.5 Family computers with internet access (per hundred households)	5
2. Resources of agricultural informatization	2.1 Agricultural book types	6
	2.2 Publishing newspaper and magazines	7
	2.3 Libraries	8
	2.4 Websites	9
3. The application of agricultural informatization	3.1 The population of netizens in rural area (ten thousand)	10
	3.2 Number of key agricultural scientific and technological achievements	11
	3.3 The aggregation of tele-service in rural area (hundred million RMB)	12
4. Personnel in agricultural informatization	4.1 Population taking up scientific activities out of ten thousand people (per ten thousand people)	13
	4.2 Population in information transmission by computer service and software industry (per ten thousand people)	14
	4.3 Population in transportation, logistics and telecommunication industry (per ten thousand people)	15
	4.4 Population in education, cultural, art, television and radio sector (per ten thousand people)	16
5. The policy environment of agricultural informatization	5.1 The proportion of S&T expenditure in GDP (%)	17
	5.2 Local government fund for S&T(hundred million RMB)	18
	5.3 The proportion of local government fund for S&T in the total expenditure of local government (%)	19

3.1   Measuring the Indices

For the sake of simplicity, this paper uses the measurement of the infrastructure of agricultural informatization (F) to demonstrate the measurement method for the indices and sub-indices and obtain the final index of agricultural informatization in order to avoid repetition. The corresponding data for sub-indices under the infrastructure of agricultural informatization from 2001 to 2010 is illustrated in Table 2.

Table 2. Sub-indices under the infrastructure of agricultural informatization (2001-2010)

Year	Population coverage rate of radio	Population coverage rate of television	Telephone density	Mobile phone penetration rate	Family computers with internet access
2001	F_{11}	F_{21}	F_{31}	F_{41}	F_{51}
2002	F_{12}	F_{22}	F_{32}	F_{42}	F_{52}
2003	F_{13}	F_{23}	F_{33}	F_{43}	F_{53}
2004	F_{14}	F_{24}	F_{34}	F_{44}	F_{54}
2005	F_{15}	F_{25}	F_{35}	F_{45}	F_{55}
2006	F_{16}	F_{26}	F_{36}	F_{46}	F_{56}
2007	F_{17}	F_{27}	F_{37}	F_{47}	F_{57}
2008	F_{18}	F_{28}	F_{38}	F_{48}	F_{58}
2009	F_{19}	F_{29}	F_{39}	F_{49}	F_{59}
2010	F_{10}	F_{20}	K_{30}	K_{40}	K_{50}

All the data need to be processed with dimensionless method prior to calculation in order to unify units. As Table 3 shows, this paper sets the year 2001 as the benchmark and all the index data in this year as 1, then divides the absolute value of each index data in year t by the corresponding index data in the benchmark year, and obtains the dimensionless value of each index in year t.

Table 3. The dimensionless processing of sub-indices under the infrastructure of agricultural informatization (2001-2010)

Year	Population coverage rate of radio	Population coverage rate of television	Telephone density	Mobile phone penetration rate	Family computers with internet access
2001	1	1	1	1	1
2002	F_{12}/F_{11}	F_{22}/F_{21}	F_{32}/F_{31}	F_{42}/F_{41}	F_{52}/F_{51}
2003	F_{13}/F_{11}	F_{23}/F_{21}	F_{33}/F_{31}	F_{43}/F_{41}	F_{53}/F_{51}
2004	F_{14}/F_{11}	F_{24}/F_{21}	F_{34}/F_{31}	F_{44}/F_{41}	F_{54}/F_{51}
2005	F_{15}/F_{11}	F_{25}/F_{21}	F_{35}/F_{31}	F_{45}/F_{41}	F_{55}/F_{51}
2006	F_{16}/F_{11}	F_{26}/F_{21}	F_{36}/F_{31}	F_{46}/F_{41}	F_{56}/F_{51}
2007	F_{17}/F_{11}	F_{27}/F_{21}	F_{37}/F_{31}	F_{47}/F_{41}	F_{57}/F_{51}
2008	F_{18}/F_{11}	F_{28}/F_{21}	F_{38}/F_{31}	F_{48}/F_{41}	F_{58}/F_{51}
2009	F_{19}/F_{11}	F_{29}/F_{21}	F_{39}/F_{31}	F_{49}/F_{41}	F_{59}/F_{51}
2010	F_{10}/F_{11}	F_{20}/F_{21}	K_{30}/F_{31}	K_{40}/F_{41}	K_{50}/F_{51}

For simplicity, we set the weight of each index to equal. Since there are five sub-indices under the infrastructure of agricultural informatization, the weight of each index is Pi=0.2. Then we can use Formula (2) to calculate IIF which is the annual index value for infrastructure of agricultural informatization:

$$II_F = \sum\nolimits^n P_i W_i \qquad \qquad Formula\ (2)$$

In Formula (2), Pi is the weight of each index, and the mean of each third-class index under the infrastructure of agricultural informatization is given 0.2. Wi is the dimensionless value of index i, and n is the amount of indices in the index system, which is five in this formula.

3.2   Measuring the Overall Index

Likewise, four index value, IIR, IIA, IIH, and IIE can be calculated for the other four indices which are resources, application, personnel and policy environment. Then we adopt Formula (3) which is listed below to find out the index II, the overall index of agricultural informatization.

$$II = \sum_{i=1}^{n} \left(\sum_{j=1}^{m_i} P_{ij} W_{ij} \right) \cdot W_i \qquad \qquad Formula (3)$$

Hereby, II is the overall index value of regional agricultural informatization; n is the amount of first-class indices in the evaluation system; mi is the amount of sub-indices under the index i; Wij is the standardized value of sub-index j under the first-class

index i; and Pij is the weight of sub-index j accounting in the first-class index which it belongs to.

4   Measuring the Agricultural Informatization Contribution Rate by C-D Production Function

4.1   Model Establishment

Charles Cobb and Paul Dauglas break the hypothesis of fixed capital-output ratio and capital-saving ratio raised by Harrod-Domar model and establish C-D production function is established by to analyse the relation of input and output in a simple way. In general C-D production function divides the input into capital and labour (Wang, 2010). But in order to calculate the contribution made by agricultural informatization to agricultural economic growth, this paper revises the C-D production function based on the new growth theory proposed by Paul M. Romer. Romer argues that it is not only capital and labour, but also S&T development that should be included in C-D production function as the third element (Romer, 1994). Therefore this paper argues that information is the foremost income of S&T development and consequently information can take the place of S&T development to be factored in the C-D production function as the third element in the input. Hence C-D production function can be revised to Formula (4):

$$Y = AK^{\alpha} L^{\beta} I^{\gamma} \qquad\qquad Formula(4)$$

Its log-linear model is:

$$lo\,g(Y) = log(A) + \alpha log(K) + \beta lo\,g(L) + \gamma log\,(I) \qquad Formula(5)$$

In Formula (4) and Formula (5), Y, K, L, and I represent the total yield of agricultural products, the quantity of capital input, the quantity of labour input, and the index of agricultural informatization respectively; α, β, and γ refer to the capital elasticity of the output, the elasticity of labour, and the elasticity of informatization respectively; A, which is set to be a constant, refers to the influence of other factors on agricultural economic growth except capital, labour and informatization.

4.2   Model Regression

All the index data adopted in the model, including the overall index I, can be calculated through the index system established in this paper; the total yield of agricultural production (Y), the fixed asset investment (K), and the population working in agriculture (L) can be looked up in China Statistical Yearbook. Then a regression model can be established as shown in Table 4.

Table 4. The regression model matrix

Year	The total yield of agricultural production (Y)	The fixed asset investment in agriculture (K)	The population working in agriculture (L)	The overall index of agricultural informatization (I)	The infrastructure of agricultural informatization (F)	Resources of agricultural informatization (R)	The application of agricultural informatization (A)	Personnel of agricultural informatization (H)	The policy environment of agricultural informatization (E)
2001	Y_1	K_1	L_1	I_1	F_1	R_1	A_1	H_1	E_1
2002	Y_2	K_2	L_2	I_2	F_2	R_2	A_2	H_2	E_2
2003	Y_3	K_3	L_3	I_3	F_3	R_3	A_3	H_3	E_3
2004	Y_4	K_4	L_4	I_4	F_4	R_4	A_4	H_4	E_4
2005	Y_5	K_5	L_5	I_5	F_5	R_5	A_5	H_5	E_5
2006	Y_6	K_6	L_6	I_6	F_6	R_6	A_6	H_6	E_6
2007	Y_7	K_7	L_7	I_7	F_7	R_7	A_7	H_7	E_7
2008	Y_8	K_8	L_8	I_8	F_8	R_8	A_8	H_8	E_8
2009	Y_9	K_9	L_9	I_9	F_9	R_9	A_9	H_9	E_9
2010	Y_{10}	K_{10}	L_{10}	I_{10}	F_{10}	R_{10}	A_{10}	H_{10}	E_{10}

A linear regression analysis of Y', K', L', I', F', R', A', and H' corresponding to indices in Table 4 is conducted based on the Enter method by SPSS 13.0. The regression result is shown in Table 5. Based on the regression result, a regression model can be set up which can be sorted into a simulation equation consisting three variables, the fixed asset investment, working population in agriculture, and the overall index of agricultural informatization, another five variables regarding the five sub-indices, and the total yield of agricultural production.

Table 5. The regression result of the model (the overall index of agricultural informatization)

Item	Constant term	(K')	(L')	(I')
Standardized coefficient		a	j	g
B	b	h	w	m
(T-statistic)	t	c	f	i
R-squared		r^2		
F-statistic		f		
Durbin-Watson statistic		d		

According to the regression model above, we determine the fitting degree of selected samples and the entirety, and test the linearity degree of the model by F-statistic and the explanation degree of each variable to the economic growth by T-statistic. If the fitting degree, the linearity degree and the explanation degree is high, and variable *K, L, I* and the constant term is able to meet the test at a high probability, it means that the model can meet the test as a whole and have a good simulate effect on the reality. Thus the model of agricultural economic growth is shown as Formula (6):

$$Y^{'} = b + h\,K^{'} + wL^{'} + mI^{'} \qquad\qquad Formula(6)$$

Formula (6) can be converted into Formula (7):

$$Y_t = e^b K_t^h L_t^w I_t^m \qquad\qquad Formula\ (7)$$

Similarly, we can also set up regression models for the five first-class indices including the infrastructure of agricultural informatization (F), resources of agricultural informatization (R), the application of agricultural informatization (A), and personnel of agricultural informatization (H), and the policy environment of agricultural informatization (E), which this paper won't repeat.

5   Conclusion

This paper finds out the standardized coefficient and the output elasticity of the three factors in agricultural economy which are capital, labour and information. From the perspective of standardized coefficient, in the regression result of the model of the overall index of agricultural informatization, the standardized coefficients of the fixed asset investment, population working in agriculture and information are a, j, and g respectively, which can assess the influence of the three factors on agricultural economy development.

From the point of view of the output elasticity, in the regression result of the model of the overall index of agricultural informatization, the output elasticity of K', L', and I' are: h, w, and m. According to the definition of output elasticity, the agricultural economy increases by $m\%$ for one further percentage point of growth of agricultural informatization; whereas one further percentage point of growing of capital and labour can increase the agricultural economy by $h\%$ and $w\%$. Thus the contribution rate of agricultural informatization to the agricultural economic growth in any year can be calculated after the index of agricultural informatization is acquired.

Likewise, the contribution rate of each index in the model to agricultural economic can be measure`ed.

References

1. TKularahte, E.D.: Information Needs and information Provision in Developing Countries. Inofation Delment 13, 9 (1997)
2. Gao, Y., Gan, G.H.: A preliminary research on the measurement index system for agricultural informatization. Agriculture Network Information 8, 49–53 (2009)
3. Li, B., Liu, L.: An empirical study on the agricultural informatization's contribution to the economic growth in Hunan Province. Economic Geography 10, 63–66 (2009)
4. Liu, F.F.: An empirical analysis on the agricultural informatization contribution based on the data envelopment analysis (DEA). Journal of Library and Information Sciences in Agriculture 2, 21–25 (2011)
5. Lu, L.N.: Constructing the measurement index system for agricultural informatization. Journal of Library and Information Sciences in Agriculture 4, 39–41 (2007)
6. Romer, P.M.: The origins of endogenous growth. Journal of Economic Perspectives 8(1-winter), 3–22 (1994)
7. Tseng, C.C., Gmytrasiewicz, P.J.: Real-time decision support and information gathering system for financial domain. Physical A: Statistical Mechanics and its Applications 363(5), 417–436 (2008)
8. Wang, L.: An empirical analysis on the informatization's contribution to the economic growth in rural area in Shandong Province. Northwest University of Farming and Forestry Technology (2010)
9. Yu, L., Geng, H.J.: A study on the measurement index system of Chinese agricultural informatization development and the measurement method. New Science 1, 33–36 (2007)
10. Zhao, H.Y., Dai, Y.M., Li, Z.H., Ding, J.B.: An empirical analysis on agricultural informatization contribution rate based on Cobb-Douglas production function. Journal of Jiangxi Agriculture University (Social Science Edition) 3, 44–49 (2009)

Estimation of Soil Total Nitrogen and Soil Moisture Based on NIRS Technology

Xiaofei An, Minzan Li, and Lihua Zheng

Key Laboratory of Modern Precision Agriculture System Integration Research,
China Agricultural University, Ministry of Education, Beijing, China
`limz@cau.edu.cn`

Abstract. Estimation model between soil moisture content and the near infrared reflectance was established by the linear regression method and the models between soil total nitrogen content and the near infrared reflectance were also established by the BP neural network method and Support Vector Machine (SVM) method. Forty-eight soil samples were collected from China Agricultural University Experimental Farm. After the soil samples were taken into the laboratory, NIR absorbance spectra were rapidly measured under the original conditions by the FT-NIR (Fourier Transform Near Infrared Spectrum) analyzer. At the same time the soil moisture (SM) and soil total nitrogen (TN) were measured by the laboratory analysis methods. The results of the study showed that a linear regression method achieved an excellent regress effect for soil moisture. The correlation coefficient of the calibration (R_C) was 0.88, and the correlation coefficient of the validation (R_V) was 0.85. The model was passed F test and t test. For soil total nitrogen, the model effect of BP neural network was better than that of SVM method, and the correlation coefficient of the calibration (R_C) coefficient and the validation (R_V) was 0.92 and 0.88, respectively. Both RMSE and PMSE were low. The results provided an important reference for the development of a portable detector.

Keywords: Soil total nitrogen, Soil moisture, BP neural network, Support vector machine.

1   Introduction

Precision agriculture requires a clear understanding of the spatial variability of soil characteristics and the nutrient status in real time. Spectroscopy, as a rapid, non-destructive, simple, and green measure and analysis technique, plays an important role in the determination of soil nutrients [1]. Research based on real-time spectral analysis for determination of soil nutrients in precision agriculture can provide accurate scientific data and theoretical basis for the development of a low-cost and intelligent measure instrument.

   Many scholars have researched on the relation between the soil parameters and the spectral characteristics in the 20th century. Alabhas et al (1972) found that soil

organic matter in the near infrared area had characteristics related with several functional groups in the organic compound[2]; Li et al (1999) analyzed soil parameters, such as NO3-N, SOM, EC and pH based on NIRS, and then determined sensitive wavebands of the parameter[3]; Sun Jianying (2006) measured soil data of NIRS by a FT-NIR analyzer at the wavelength range of 830-2500 nm and several forecasting models for soil moisture, soil organic matter and soil total nitrogen were established[4];Zheng Lihua (2010) established the forecasting models for soil total nitrogen and soil organic matter by WT (Wavelet Transform) and SVM method[5]; Gao Hongzhi (2011) selected nine wavelengths to establish the forecasting models of soil total nitrogen and soil organic matter based on actual experience [6].

Remarkable progress has been made in the forecasting of soil moisture, soil organic matter and soil total nitrogen based on NIRS[7-14].However, above studies provided the dried and crushed soil samples in the experiment so that the results obtained could not be directly used to real-time evaluate soil parameters. Hence, it was important to reveal the feasibility and possibility of analyzing soil parameters with NIR spectroscopic techniques using raw soil samples and provide a reference for the development of a portable detector.

2   Materials and Methods

2.1   Experiments and Collection

Forty-eight soil samples were collected from the experimental field of China Agricultural University located in the suburb of Beijing. Soil samples were collected from different experimental areas and depth for the different levels. A double-layer sealing plastic bag and kraft paper bag were adopted to avoid soil moisture loss. The NIRS of the forty-eight soil samples were measured in the laboratory by a MATRIX-I type of FT-NIR analyzer with a rotating sample pool produced by the Germany Bruker Optical Company. About 15 grams of soil sample could be directly placed into the rotating sample pool by a forty-eight mm diameter of quartz glass. The spectral measure range was 12798~3599 cm-1(781~2779nm), resolution 4 cm-1, wavenumber accuracy 0.1 cm-1, scan times 32. A spectral curve could be obtained after taking a small amount of soil sample into the rotating sample pool and setting the relevant parameters. Fig.1 shows several spectral curves of soil samples in the range of 850~2450 nm.
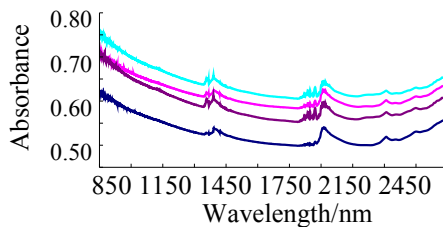
Fig. 1. Spectral curves of soil samples

Soil moisture contents and soil total nitrogen contents were measured by physical method and chemical method respectively in the laboratory. Soil moisture contents were measured by the drying and weighing method with an electric heater, and soil total nitrogen contents were measured by the semi-micro kjeldahl determination with a FOSS2300 Kjeldahl nitrogen instrument.

2.2   Modeling

Single linear regression analysis is a mathematical statistical method to deal with the correlation between variables [14]. A linear regression method can be used for the obvious correlation between soil moisture contents and the absorbance of the soil samples. But the correlation between the soil total nitrogen content and the absorbance was lower than 0.3 in full spectral range. In addition, several papers demonstrated that there was weak correlation and nonlinear relation between soil total nitrogen contents and the near infrared spectra [4-9]. So that it was a better attempt to estimate the soil total nitrogen content by a nonlinear prediction model. In our study, soil total nitrogen contents were predicted by BP neural network method and support vector machine method respectively.

BP neural network is back-propagation network using nonlinear differentiable function to serve as weight value train [13,14]. According to the theorem of BP neural network, single hidden layer BP network can be appropriated for any continuous function in a closed interval. A three layers BP network can complete the mapping from uncertain dimension to determined dimension. This network has strong nonlinear mapping ability, which is very suitable for nonlinear function approximation. In the training period, the error is inverted propagation.

Support vector machine was proposed in 1992 by Vapnik et al [15].One of the most important characters in SVM method is to generate hypotheses for the sparse matrix of dual representation, according to the general theory of reinforcement learning. This makes it become a very effective algorithm. Another important characteristic is that the corresponding optimization problem becomes convex problem and no local minimum, because of the introduction of kernel function. SVM has the advantages of simple structure and the solution of the small sample, nonlinear, high dimension and local minima, compared with the traditional BP neural network.

3   Results and Discussion

3.1   Data Preprocessing

Statistic characteristics of soil moisture contents and soil total nitrogen contents are shown in Fig.2. In the box graph, the first column is the soil moisture content distribution and the second column is the soil total nitrogen content distribution. It was observed from Table 1 that the standard deviation of the soil moisture content was larger than that of the soil total nitrogen content.
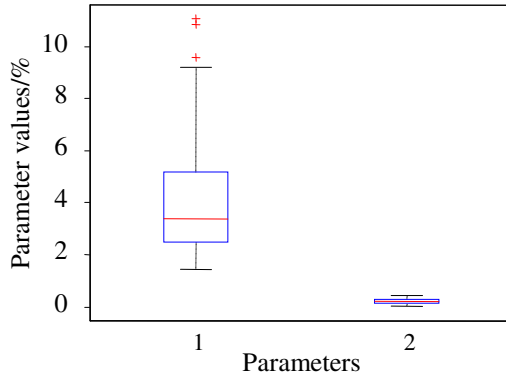
Fig. 2. Statistic characteristics of soil samples

Table 1. The measured values of two soil parameters in the laboratory

Parameters	The number of samples	Min /%	Max /%	Average /%	Mean/ %	Standard deviation
SM	48	1.44	11.05	4.18	3.34	2.54
TN	48	0.04	0.45	0.22	0.21	0.10

The soil spectral reflectance needed to be smoothed since the white noise interference could be weakened after smoothing processing in a certain degree. The first deviation could remove the baseline wander or flat background interference effects and improve the original spectral resolution and spectral profile transformation. The forecasting model of soil moisture content was established by the original absorbance spectral data, while the forecasting model of soil total nitrogen content was established by using the data processed by the first deviation of the spectral absorbance. The first deviation of the spectral absorbance was calculated using the Equation (1) [15,16]:

$$y_i = \frac{1}{6h}(-11\,x_i + 18\,x_{i-1} - 9\,x_{i-2} + 2\,x_{i-3}) \tag{1}$$

Where: y_i is the first derivative, x_i is the NIR absorbance, h is the sampling interval.

3.2   Estimation of Soil Moisture

The forecasting model of soil moisture content was established by the linear regression method with the original absorbance spectral data. All the forty-eight soil samples were divided into two parts: calibration set (including 36 soil samples) and

validation set (including 12 soil samples). Fig.3 shows the comparisons between measured data and predicted data of soil moisture. The model formula is shown in Equation (2). The correlation coefficient of the calibration (R_C) was 0.88, and the correlation coefficient of the validation (Rv) was 0.85. The model was passed the F test and t test, which illustrated that the model was significant and the parameters were significant. It was proved that the wavelength of 1450 nm could be taken as an initiative light source selection for a portable detector.

$$y_i = 43.72 \times x_i - 24.61 \tag{2}$$

Where: y_i is prediction of soil moisture content; x_i is absorbance of soil at the wavelength of 1450 nm.
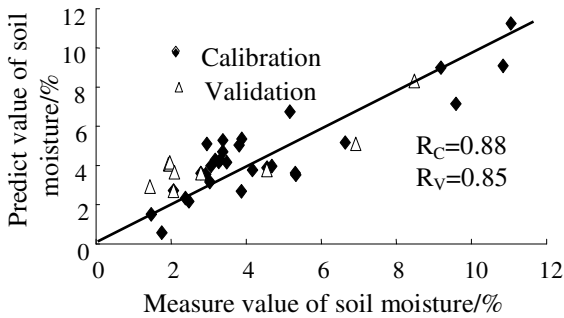
Fig. 3. Comparisons between measured data and predicted data of soil moisture

3.3   Estimation of Soil Total Nitrogen by SVM

The correlation between soil moisture and absorbance at 1450 nm was observed, while the correlation coefficients between absorbance and soil total nitrogen were less than 0.3 in full spectral range. It showed difficult for use linear regression method to complete the prediction of soil total nitrogen. SVM and BP neural network were respectively chosen to establish the soil total nitrogen forecasting model.

There was no unified model to determine parameters of SVM. The parameters could be chosen by virtue of experience, experiment or a large range of search methods. After a lot of attempts, the parameters were determined: the kernel function was RBF, penalty parameter C was 11.314 and kernel parameter g was 2.0. Table 2 shows all the parameters. All the first deviation data was divided into two parts: training set and validation set. The data were used for calibration and validation by the SVM method after normalization.

Table 2. Parameters of SVM

Parameters	Parameter values
kernel function	RBF
penalty parameter C	11.314
kernel parameter g	2.0
MSE	0.0124
input variants	6
output variant	1

Fig.4 shows the comparisons between measured value and predicted value of soil total nitrogen by SVM. The correlation coefficient of the calibration (R_C) was 0.89; the correlation coefficient of the validation (R_V) was 0.78. The result of calibration and validation demonstrated that the model to evaluate TN is feasible. The results showed could be taken as theoretic basis for the analysis of real-time TN.
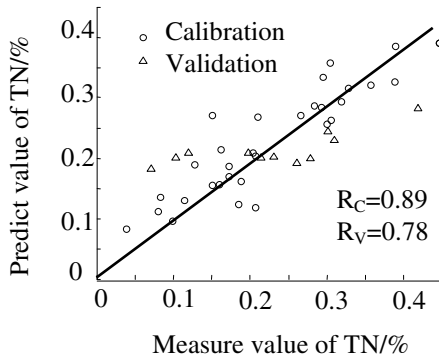
Fig. 4. Comparisons between measure value and predict value of TN by SVM

3.4   Estimation of TN by BP Neural Network

BP neural network can avoid adverse effects of abnormal data caused by equipment, human factor and etc. It makes the results meet practical prediction accuracy. A three layers BP neural network was established, and S type functions were selected as transfer function of each layer. The input neurons were 6 and output neuron was 1. The neurons in hidden layer were 27, error value of 0.005. Other parameters such as momentum factor, step length and initial weights value were set in a certain range of random by the program. The input neurons were the first deviation of absorbance of soil samples at six wavelengths (940, 1050, 1100, 1200, 1300, 1550 nm), and the output neuron stood for the soil total nitrogen.
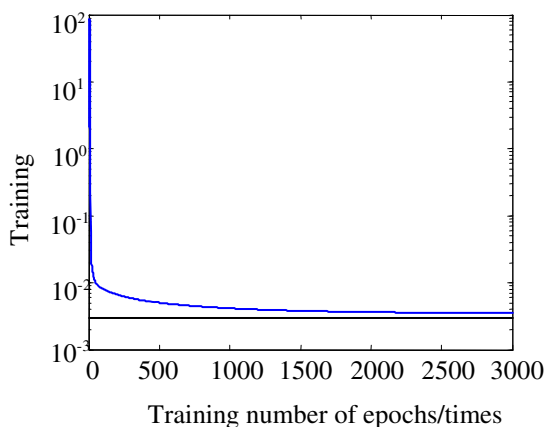
Fig. 5. Bp NN training curve

Table 3. Bp neural network parameters

Parameters of BP NN	Parameter values
input neurons	6
output neurons	1
neurons in hidden layer	27
network structure	single hidden layer
transfer function in input layer	tansig
transfer function in output layer	purelin
training function	trainlm
error	0.00343

All the forty-eight soil samples were divided into two parts: calibration set (including 36 soil samples) and validation set (including 12 soil samples). The correlation coefficient of the calibration (R_C) was 0.92, and the correlation coefficient of the validation (Rv) was 0.88.

The forecasting effect of BP neural network was better than that of SVM method. It could provide a model reference for the development of a portable detector. Fig.6 shows the comparisons between measured dataset and predicted dataset of soil total nitrogen by BP neural network.
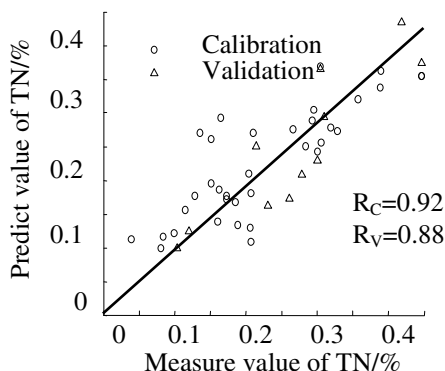
Fig. 6. Comparisons between measured dataset and predicted dataset of TN by BP neural network

4   Conclusions

The forecasting models of soil moisture content and soil total nitrogen were established by the linear regression method and nonlinear regression methods respectively. The following conclusions were obtained.

1. The soil moisture content could be predicted by a linear regression method, and the correlation coefficient of the calibration (R_C) and the validation (Rv) were 0.88 and 0.85, respectively.

2. The first deviation of absorbance data could effectively reduce the system error of the network model. The forecasting model of soil total nitrogen content established by the BP neural network was better than that of SVM method, and the correlation coefficient of the calibration (R_C) and the validation (R_V) were 0.92 and 0.88, respectively.

3. The models were robust. The accuracy of the model would not be affected when the number of the samples was many enough even no abnormal data were eliminated. All the seven wavelengths chosen in the model could be found the corresponding single wavelength LED in the market. Thus, it provided an excellent model reference for the development of a portable detector.

References

1. Li, M.Z.: Evaluating Soil Parameters with Visible Spectroscopy. Transactions of the CSAE 19(5), 36–42 (2003)
2. Al-Abbas, A.H., Swain, P.H., Baumgardner, M.F.: Relating Organic Matter and Clay Content to the Multispectral Radiance of soils. Soil Science 114(6), 477–485 (1973)
3. Li, M.Z., Shibusawa, S., Sasao, A.: Spectroscopic Approach to Soil Parameters Sensing. In: Nir 1999: International Conference on Agricultural Engineering, Beijing, pp. 98–103 (1999)
4. Sun, J.Y., Li, M.Z., Zheng, L.H.: Real-time Analysis of Soil Moisture, Soil Organic Matter, and Soil Total Nitrogen with NIR Spectra. Spectroscopy and Spectral Analysis 26(3), 426–429 (2006)
5. Zheng, L.H., Li, M.Z., An, X.F.: Forecasting Soil Parameters Based on NIR and SVM. Journal of Agricultural Engineering 26(2), 81–87 (2010)
6. Gao, H.Z., Lu, Q.P.: Near Infrared Spectral Analysis and Measuring System for Primary Nutrient of Soil. Spectroscopy and Spectral Analysis 31(5), 1245–1249 (2011)
7. Yu, F.J., Min, S.G., Huang, X.T.: Near Infrared Spectrum Analysis of Soil Organic Matter and Nitrogen. Analysis Laboratory 11(3), 47–51 (2002)
8. Zheng, L.H., Li, M.Z., Sun, J.Y.: Evaluation of Soil Fertility with Spectrophotometer and Spectroradiometer. In: Infrared and Photoelectronic Imagers and Detector Devices. Proceedings of SPIE, vol. 5881, pp. 138–146 (2001)
9. Chen, P.F., Liu, L.Y., Wang, J.H.: Near Infrared Spectroscopy Real-time Measuring Soil Total Nitrogen and Phosphorus Content. Spectroscopy and Spectral Analysis 57(2), 295–298 (2008)
10. Yuan, S.L., Ma, T.Y., Song, T.: Total Nitrogen in Soil and the Total Phosphorus Content in the Near Infrared Spectrum of Real-time Detection Methods. Journal of Agricultural Machinery 9(S1), 150–153 (2009)
11. Barthes, B.G., Didier, B., Edmond, H.: Determining the Distributions of Soil Carbon and Nitrogen in Particle Size Fractions using Near-infrared Reflectance Spectrum of Bulk Soil Samples. Soil Biology & Biochemistry 40(6), 1533–1537 (2008)
12. Song, H.Y.: Based on the OSC and PLS Soil Organic Matter of Near Infrared Spectrometry. Journal of Agricultural Machinery 38(12), 113–115 (2007)
13. Sun, J.Y., Li, M.Z., Zheng, L.H.: Based on the North of the Near Infrared Spectrum Chao Soil Parameter Real-time Analysis. Spectroscopy and Spectral Analysis 26(5), 426–429 (2006)
14. Zheng, L.H., Li, M.Z., Pan, L.: Based on the Soil of Near Infrared Spectral Technology Parameters of the BP Neural Network Forecast. Spectroscopy and Spectral Analysis 28(5), 1160–1164 (2008)
15. Li, M.Z.: Spectral Analysis Technology and Application. China Science Press, Beijing (2008)
16. Lu, W.Z.: Modem Near-Infrared Spectroscopy Analytical Technology. China Petrochemical Press, Beijing (2006)

The Improved DBSCAN Algorithm Study
on Maize Purity Identification

Pan Wang[1], Shuangxi Liu[1], Mingming Liu[1], Qinxiang Wang[1],
Jinxing Wang[1], and Chunqing Zhang[2,*]

[1] College of Mechanical and Electronic Engineering, Shandong Agricultural University,
Tai'an 271018, China
[2] College of Agricultural, Shandong Agricultural University, Tai'an 271018, China
{sdwangpan,lentree,lmmjx,wqx,jinxingw}@163.com,
cqzhang@sdau.edu.cn

Abstract. In order to identify maize purity rapidly and efficiently, the image processing technology and clustering algorithm were studied and explored in depth focused on the maize seed and characteristics of the seed images. An improved DBSCAN on the basis of farthest first traversal algorithm (FFT) adapting to maize seeds purity identification was proposed in the paper. The color features parameters of the RGB, HIS and Lab color models of maize crown core area were extracted, while H, S and B as to be the effective characteristic vector after data analysis. The abnormal points of different density characteristic vector points were separated by FFT. Then clustering results were combined after local density cluster by DBSCAN. According to the result of test, the method plays a great role in improving the accuracy of maize purity identification.

Keywords: Maize, Purity identification, Density, Color, Core area.

Introductions

Maize is one of the major crops in China, which has been used extensively in food and feed field. The purity of corn seed directly is related to the crop yield and quality. In recent years, machine vision technology with its characteristics of objective, non-destructive, rapid speed was widely used for crop detection. It is very significant to research machine vision technology instead of manual testing on maize purity identification. At present, at home and abroad, application of machine vision technology to corn seed appearance characteristics were studied. Foreign scholars used shape parameter to analysis contour features of maize for quality testing and damage identification and identified corn color by chromaticity[2-5]. Domestic scholars used the image processing technology to study the features of flank shape [6-9], HIS[8-11], RGB[8-12] and area of corn white part (embryo) or yellow part (crown) [9-10] to distinguish the quality, varieties[13] and purity.

---

* Corresponding author.

The research on morphological characteristics of corn seed had got good identification results while much less study of corn purity identification and much more application about corn side features. This paper proposed that the color features of crown core area had significant function on corn purity after the study on three commonly used maize varieties. The DBSCAN was optimized by farthest first traversal algorithm for the purity identification. The experimental result indicated this method had high classification rate and the higher precision and offered a reference for building accurate purity identification system.

1 Feature Extraction and Analysis of Image

1.1 Image Acquisition and Processing

The Original image of corn crown captured with CCD camera is processed according to the color character. The image was pre-processed that the background was segmented by R from RGB color model and single colored image was taken by contour labeling. A single crown core area was obtained by taking the single corn crown area centroid as the center, 10 pixels as the radius and signed convenient for purity identification[11].

Took Zhengdan958 for example, the maize original crown image was shown in Fig.1.
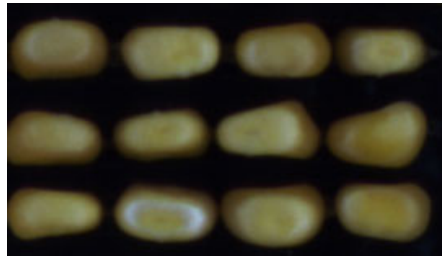
Fig. 1. Original image of Zhengdan958

The pre-processed image was shown in Fig.2, The number mark image and Maize crown core area image were shown in Fig.3 and Fig.4.
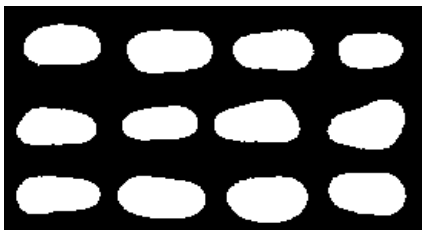
(a) Background segmentation image   (b) Contour markers image
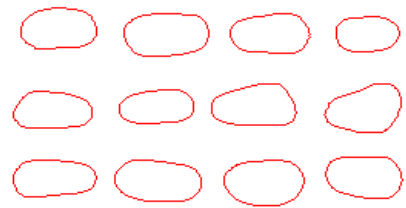
Fig. 2. Preprocessed image
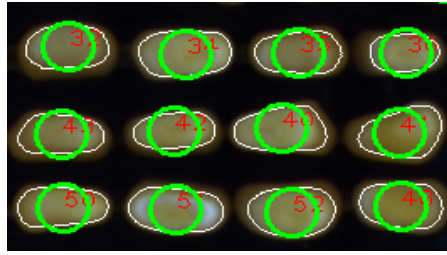
Fig. 3. Number mark image
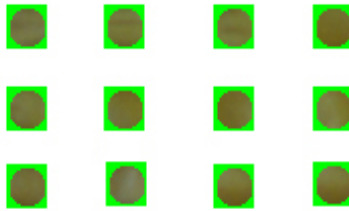
Fig. 4. Maize crown core area image

1.2   Color Feature Extraction and Characteristic Vector Selection

The main information of corn image is the shape and color features. In a real production process, it is different about the shape features . The study selected three models RGB, HSI and CIE Lab color features to extract nine color R (red), G (green), B (blue), H (hue), S (saturation) , I (brightness), L (luminance), a (red-green), b (yellow-blue) of maize crown core area image. The Eigenvalues were average of all pixels of each corn.

In order to reduce the feature vector dimensions and improve recognition rate, the nine characteristics was studied (combined I and L because of their same meaning). Took Zhengdan958 and Ludan981 for example, the means and standard deviation and coefficient of variation were shown in Table 1.

Coefficient of variation showed the difference among sample means. For any data set, if the data near the boundary density was relatively discrete, it would have a larger coefficient of variation. Standard deviation declared the between means and eigenvalue, the degree of dispersion between hybrid seeds and normal ones[14]. As can be seen from Table 1, H, S, B were contributed to identified , and selected the H, S, B, as identification eigenvectors.

Table 1. Color feature parameters statistical data of maize seeds

Varieties	Zhengdan958			Ludan981		
Color features	Means	Standard deviation	Coefficient of variation	Means	Standard deviation	Coefficient of variation
H	64.509	6.106	0.095	70.544	15.660	0.222
S	0.354	0.062	0.176	0.220	0.102	0.464
I	109.820	8.631	0.079	139.477	15.388	0.110
B	71.247	11.167	0.157	110.607	24.0188	0.217
G	130.668	10.742	0.082	154.911	15.3248	0.099
R	126.247	7.013	0.055	151.197	10.1308	0.0679
a	89.043	0.876	0.010	90.662	0.897	0.010
b	174.583	2.208	0.013	169.780	3.805	0.022

The three-dimensional characteristics of scatter points chat of the three varieties all showed strip fluid-like in figure5(S range of 0 to1, magnified it 100-fold in order to reflect its changing patterns [11])There were individual abnormal points deviating from the sample, and the distribution of feature vectors point had density differences.
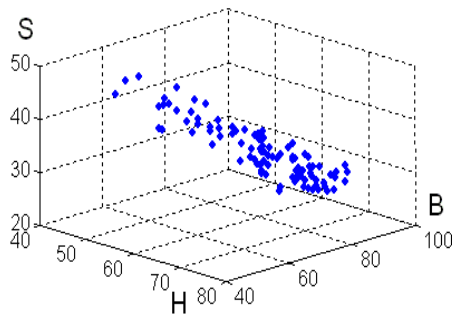
Fig. 5. Three-dimensional diagram of feature vector

2   The Corn Purity Identification Based on the Farthest First Traversal (FFT) Optimization DBSCAN

According to the three-dimensional density distribution characteristics of the color characteristic vector points in the sample, DBSCAN was used for purity identification. In the light of uneven density of the data points, the data was optimized by FFT[15]. The feature vector density difference points in the whole data space were divided into two regions of edge abnormal points that excluded and high density points clustered by DBSCAN, and finally merged cluster results.

This algorithm can roughly be divided into 4 basic steps;

(1) Input the cluster data set S which containing n sample points;

(2) Use the FFT to optimize data. According to the initial clustering centers Z_1 and Z_2, the corresponding dataset is divided into two subsets $Z_1 \in G_1^{(t')}$ and $Z_2 \in G_2^{(t)}$, and then excluded the abnormal points set G_2.

(3) DBSCAN clustering to high density local data set.

(4) Merge the cluster results and get the final cluster result.

2.1   Data Optimization

In the process of recognition, it is poor to take a global variable to cluster no uniform density points (H, S, B). Therefore, obviously abnormal data will be excluded before DBSCAN. FFT (Farthest First Traversal Algorithm) can centralize similar samples fleetly and exclude abnormal samples in dataset. The major parameters are the number of clusters (num) and the initial center point (seed). This is two-groups problem, namely, only normal hybrid seed and miscellaneous ones. In order to get accurately results, the mean value was to as the initial center point.

The main algorithm are summarized as following:

(1) Made the num=2, and calculated the mean value point(H, S,B) which was the initial center Z1. G_1^t is the dataset whose clustering center was $z_1^{(t)}$.

(2) (Represented the points (H, S, B) as (S1, S2…Sn), and calculated the distances between the points to the Z1. Euclidean distance was used in this paper.

$$d(S_j, Z_l) = \sqrt{\sum_{j=l}^{n}(S_{ij} - Z_{li})^2} \qquad (1)$$

Where Sij was the i-D coordinate of the jth points, Z_{1i} was the i-D coordinate of the initial center and n is the number of the example. i=1,2; j=1, 2…n

(3) Took the farthest point $\max d(S_j, Z_1)$ far from Z1 as the other center. $G_2^{(t)}$ was the dataset whose clustering center was $Z_2^{(t)}$.

(4) Calculated the distances from the S_j which were remainder points in S to the Z_1 and Z_2, and define the larger one as max(d_{j1}, d_{j2}). All the points in set S were distributed to the most similar group until the end of the cluster. The results was shown in Fig.6.
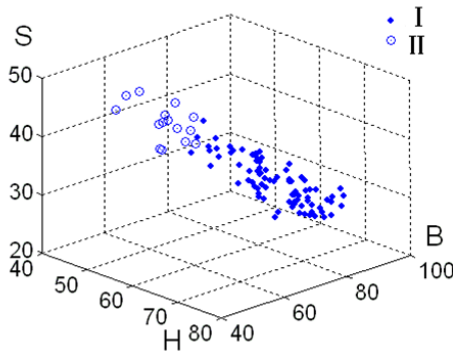
Fig. 6. Data optimization of FFT

2.2   DBSCAN Algorithm

2.2.1   About DBSCAN Algorithm

DBSCAN is a clustering algorithm which relies on a density-based notion of clusters. It is designed to discover clusters of arbitrary shape[16]. The key idea in DBSCAN is that for each object of a cluster, the neighborhood of a given radius has to contain at least a minimum number of objects. The procedure for finding a cluster is based on the fact that a cluster is uniquely determined by any of its core objects.

(1) Given an arbitrary object S for which the core object condition holds, the set $\{o \mid o > G_{1s}\}$ of all objects o density-reachable from S in D forms a complete cluster C and S∈C.

(2). Given a cluster C and an arbitrary core object S∈C, C in turn equals the set $\{o \mid o > G_{1s}\}$.

2.2.2   Determination of Parameters

DBSCAN requires the user to specify the global parameter Eps (The parameter MinPts is fixed to 4 to reduce the computational complexity [17]). In order to determine the Eps, DBSCAN algorithm calculates the distances between all the data objects and found the most adjacent Minpts object. The k-dist diagram was made by the distances sort to confirm the Eps.

The horizontal coordinate of the k-dist diagram represented the number of data objects corresponding to some distance value of the k-dist; the vertical coordinate represented the distance between data objects (H, S and B) and its fourth adjacent objects. The function of k-dist diagram was to determine the most appropriate Eps, namely lower limit density. The distances between the data objects and its kth closest object is ordered by decreasing, so k-dist diagram was also called as sorting k-dist diagram. K-dist diagram was shown in figure 7, while the value of position A in the flat part was set for Eps.
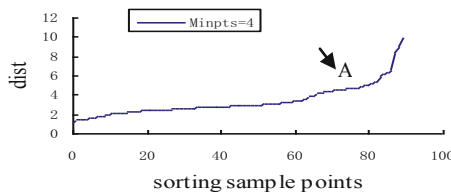
Fig. 7. k-dist diagram

Set(4, 5.2) for (Minpts, Eps), the description of DBSCAN algorithm is as follows:

(1) Choose the object S that belongs to data set G_1 but not any clustering to meeting kernel conditions and create a new clustering.

(2) According to the kernel object of the clustering, collect accessible density of kernel object in a circular until there were no new kernel objects.

(3)The circle will not end until any kernel object does not exist in any clustering, otherwise continue step 1.

In this process, noise points were recorded who's distance from the division boundary were less than Eps while they could be boundary points or the points of some divided cluster.
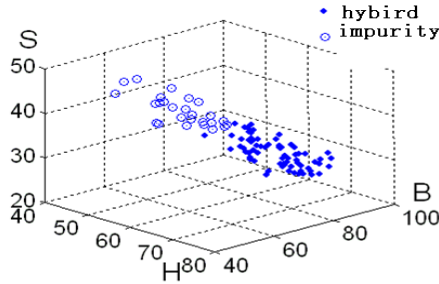
Fig. 8. Clustering results

2.3   Local Clustering Merger

According to the clustering process, the abnormal scattered points removed by FFT data optimization and noise points or small class points by local clustering had been as false seeds. Take Zhengdan958 for example, there just were small class points. The combining clustering results was as shown in Figure 8.

2.4   Experimental Results and Analysis

Three common maize varieties Zhengdan958, Nongda108 and Ludan981 were tested to prove the proposed algorithm. The results with the purity identification rate above 93.3% were as shown in Table2. The basic flow field to meet the. It had met the corn seed purity identified requirements in transaction filed basically.

The features of crown core area were affected to some extent in result of that the crown core area was set by experiences. Artificial placing had some effect on the height of corn crown. The balance of the image acquisition system lighting was also important to the identification results.

Table 2. Maize purity identification result

Varieties	Number	Impurity number in fact	Impurity number by test	identification rate /%
Zhengdan958	100	31	30	96.8%
Nongda108	100	30	29	96.7%
Ludan981	100	30	28	93.3%

3   Conclusions

(1) The core area of crown was chosen as a research object, in which the color characteristics were studied and H, S, B, was determined as eigenvectors for identified corn seed purity.

(2) Three-dimensional space model of eigenvectors were established. Aiming at the existence of density distribution differences the DBSCAN corn purity recognition algorithm which based on the FFTA was proposed. The algorithm optimized data by FFT, excluded the edge abnormal scatter, local clustered high density areas through DBSCAN, merged cluster region, and finally obtained the results of purity recognition.

(3) The test for purity Identification of three selected maize varieties Zhengdan958, Nongda108 and Ludan981 proved that the method can achieve an average recognition rate of 93.3% or more.

References

1. Liu, Y., Ying, Y., Cheng, F.: Research of Machine Vision in Purity Inspection of Seed. Transactions of the CSAM 34(5), 161–163 (2003)
2. Liao, K., Paulsen, M.R., Reid, J.F., Ni, B.C., Bonifacio-Magirang, E.P.: Corn kernel breakage classification by machine vision using a neural network classifier. Transactions of the ASAE 36(6), 1949–1953 (1993)
3. Ni, B., Paulsen, M.R., Reid, J.F.: Corn kernel crown shape identification using image processing. Transactions of the ASAE 40(3), 833–838 (1996)
4. Paliwal, J., Shashidhar, N.S., Jayas, D.S.: Grain kernel identification using kernel signature. Transactions of the ASAE 42(6), 1921–1924 (1999)
5. Liu, J., Paulsen, M.R.: Corn whiteness measurement and classification using machine vision. Transactions of the ASAE 43(3), 757–763 (2000)
6. Ning, J., He, D., Yang, S.: Identification of tip cap and germ surface of corn kernel using computer vision. Transactions of the CSAE 20(3), 117–119 (2004)
7. Hao, J., Yang, J., Du, T., Cui, F., Sang, S.: A Study on Basic Morphologic Information and Classification of Maize Cultivars Based on Seed Image Process. Scientia Agricultura Sinica 41(4), 994–1002 (2008)
8. Cheng, H., Shi, Z., Me, W., Wang, L., Pang, L.: Corn Breed Recognition Based on Support Vector Machine. Transactions of the CSAM 40(3), 180–183 (2009)
9. Wang, Y., Liu, X., Su, Q., Wang, Z.: Maize seeds varieties identification based on multi-object feature extraction and optimized neural network. Transactions of the CSAE 26(6), 199–204 (2010)
10. Shi, Z., Cheng, H., Li, J., Feng, J.: Characteristic parameters to identify varieties of corn seeds by image processing. Transactions of the CSAE 24(6), 193–195 (2008)
11. Yan, X., Liu, S., Zhang, C., Wang, J.: Purity identification of maize seed based on color characteristics. Transactions of the CSAE 26(supp.1), 46–50 (2010)

12. Song, P., Wu, K., Zhang, J., Li, W., Fang, X.: Sorting System of Maize Haploid Kernels Based on Computer Vision. Transactions of the CSAM 41(supp.), 249–252 (2010)
13. Quan, L., Zhu, R., Lei, P., Han, B.: Recognition Method of Maize Cultivars Based on K-L Transform and LS-SVM. Transactions of the CSAM 41(4), 168–172 (2010)
14. Xue, L., Qiu, B.: Boundary Points Detection Algorithm Based on Coefficient of Variation. PR & AT 22(5), 799–802 (2009)
15. Gonzalez, T.F.: Clustering to minimize the maximum intercluster distance. Theoretical Computer Science 38, 293–306 (1985)
16. Feng, S., Xiao, W.: An Improved DBSCAN Clustering Algorithm. Journal of China University of Mining &Technology 37(1), 106–111 (2008)
17. Yue, S.H., Li, P., Guo, J.D., Zhou, S.G.: A statistical information based clustering approach in distance space. Journal of Zhejiang University Science 6A(1), 71–78 (2005)

Author Index