Fuling Bian
Yichun Xie
Xiaohui Cui
Yixin Zeng (Eds.)

Communications in Computer and Information Science 399

Geo-Informatics in Resource Management and Sustainable Ecosystem

International Symposium, GRMSE 2013
Wuhan, China, November 2013
Proceedings, Part II

Part 2

Springer

Communications in Computer and Information Science 399

Fuling Bian Yichun Xie Xiaohui Cui
Yixin Zeng (Eds.)

Geo-Informatics in Resource Management and Sustainable Ecosystem

International Symposium, GRMSE 2013
Wuhan, China, November 8-10, 2013
Proceedings, Part II

Springer

Volume Editors

Fuling Bian
Xiaohui Cui
Yixin Zeng
Wuhan University
No. 37 Luoyu Road, Wuhan, Hubei 430079, China
E-mail: {flbian, xcui, yixinzeng}@whu.edu.cn

Yichun Xie
Michigan University, Institute for Geospatial Research and Education
125 King Hall, Ypsilanti, MI 48197, USA
E-mail: yxie@emich.edu

ISSN 1865-0929 e-ISSN 1865-0937
ISBN 978-3-642-41907-2 e-ISBN 978-3-642-41908-9
DOI 10.1007/978-3-642-41908-9
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013952958

CR Subject Classification (1998): C.3, C.4, G.1.6, H.2.8, H.3.1-4, I.4, I.5

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The 2013 International Conference on Geo-Informatics in Resource Management and Sustainable Ecosystem (GRMSE 2013) was held in Wuhan, China, during November 8–10, 2013. GRMSE 2013 aimed to bring together researchers, engineers, and students working in the areas of geo-informatics in resource management and sustainable ecosystem. GRMSE 2013 featured a unique mix of topics including smart city, spatial data acquisition, processing and management, modeling and analysis, and recent applications in the context of building healthier ecology and resource management.

We received 522 submissions from various parts of the world. The Technical Program Committee worked very hard to have all papers reviewed before the review deadline. The final technical program consisted of 136 papers. There were four keynote speeches, 5 invited sessions. All the keynote speakers are internationally recognized leading experts in their research fields, who have demonstrated outstanding proficiency and have achieved distinction in their profession. The proceedings are published as two volumes in Springer's *Communications in Computer and Information Science* (CCIS) series. Some excellent papers were selected and recommended to the *International Journal of Computational Science and Engineering, Sensors & Transducers Journal* and *International Journal of Embedded Systems*. We would like to mention that, due to the limitation of the conference venue capacity, we were not able to include many fine papers in the technical program.

We would like to express our sincere gratitude to all the members of Technical Program Committee and organizers for their enthusiasm, time, and expertise. Our deep thanks also go to many volunteers and staff for the long hours and hard work they generously gave to GRMSE 2013. We are very grateful to Wuhan University and Eastern Michigan University for their support in making GRMSE 2013 possible. The generous support from the Joint International Center for Resource, Environment Management and Digital Technologies (JIC-REDT), International School of Software, Wuhan University, is greatly appreciated. Finally, we would like to thank all the authors, speakers, and participants of this conference for their contributions to GRMSE 2013, and we also look forward to welcoming you to Michigan in 2014.

Fuling Bian

Organization

International Conference on Geo-Informatics in Resource Management and Sustainable Ecosystem (GRMSE 2013)
November 8–10, 2013 Wuhan, Hubei, China http://www.ggers.org

GRMSE 2013 Committee

Honarary Chair

Michael Batty	Bartlett Professor and Chair of the Centre for Advanced Spatial Analysis (CASA) at University College London, UK
Deren Li	Academician of Chinese Academy of Sciences and Chinese Academy of Engineering, Wuhan University, China

Steering Chair

Fuling Bian	Wuhan University, China

General Chair

Wenzhong Shi	The Hong Kong Polytechnic University, Hong Kong, SAR China
Xiaohui Cui	Wuhan University, China

Executive Chair

Guobin Zhu	Wuhan University, China
Yixin Zeng	Wuhan University, China

Technical Program Chair

YiChun Xie	Eastern Michigan University, USA

Technical Program Chair on Ecological Remote Sensing

Zongyao Sha	Wuhan University, China

Technical Program Chair on Geospatial Information

Richard Sambrook	Eastern Michigan University, USA
Yangge Tian	Wuhan University, China

Technical Program Chair on GIS

Qingwen Xiong	Polytechnic Institute of New York University, USA

Technical Program Committee

YiChun Xie	Eastern Michigan University, USA
Zongyao Sha	Wuhan University, China
George Christakos	San Diego State University, USA
Ping Fang	Tongji University, China
Kuishuang Feng	University of Maryland, USA
Nanshan Zheng	China University of Mining and Technology, China
Changsheng Cai	Central South University, China
Zhenhong Li	University of Glasgow, UK
Yuqi Bai	Tsinghua University, China
Sabine Baumann	Technische Universität München, Germany
Qinghui Huang	Tongji University, China
David Forrest	University of Glasgow, UK
Arie Croitoru	George Mason University, USA
James Cheng	Manchester Metropolitan University, UK
Paul Torrens	University of Maryland, USA
Stephan Mäs	Technische Universität Dresden, Germany
Gina Cavan	Manchester Metropolitan University, UK
Jan Dempewolf	University of Maryland, USA
Bor-Wen Tsai	National Taiwan University, Taiwan
Yu Liu	Peking University, China
Xiaojun Yang	Florida State University, USA
Yan Liu	The University of Queensland, Australia
Jinling Wang	University of New South Wales, Australia
Xiaolei Li	Wuhan University, China
Pariwate Varnakovida	Prince of Songkla University, Thailand
Manfred F. Buchroithner	Technische Universität Dresden, Germany
Anthony Stefanidis	George Mason University, USA
Chaowei Yang	George Mason University, USA
Xiaoxiang Zhu	Technische Universität München, Germany
Matt Rice	George Mason University, USA
Jianjun Bai	Shaanxi Normal University, China
Yongmei Lu	Texas State University, USA
Alberta Albertella	Technische Universität München, Germany
F. Benjamin Zhan	Texas State University, USA
Huamin Wang	Wuhan University, China
Edwin Chow	Texas State University, USA
Lin Liu	University of Cincinnati, USA
Shuqiang Huang	Jinan University, China
Weihua Dong	Beijing Normal University, China
Shuang Li	Wuhan University, China
Mengxue Li	University of Maryland, USA
Wenwen Li	Arizona State University, USA

André Skupin	San Diego State University, USA
Yun Zhang	Wuhan University, China
Alan Murray	Arizona State University, USA
Mike Worboys	The University of Maine, USA
Mu Zhang	Wuhan University, China
Amirhossein Sajadi	Case Western Reserve University, USA

Table of Contents – Part II

Session 4: Advanced Geospatial Model and Analysis for Understanding Ecological and Environmental Process

Session 5: Applications of Geo-Informatics in Resource Management and Sustainable Ecosystem

Table of Contents – Part I

Session 1: Smart City in Resource Management and Sustainable Ecosystem

Session 2: Spatial Data Acquisition through RS and GIS in Resource Management and Sustainable Ecosystem

Session 3: Ecological and Environmental Data Processing and Management

The Analysis of the Initiation Conditions of the Debris Flow in the Jiangjia Ravine Based on the Simulation of the Hydrology Response

Jiafa Tang

Dept. Remote sensing and geospatial information, Southwest Jiaotong University, Chengdu, P.R. China, 610031
tajava@home.swjtu.edu.cn

Abstract. Soil moisture is one of the key hydro reasons that caused the debris flow. In this paper, a simulation model, based on the physical process of hydro cycle and the distributed hydrology model: TOPMODEL, is built to simulate the soil moisture variation in Jiangjia ravine in 2001. In this model, the potential evapotranspiration is computed with the Penman-Monteith equation; the water movement in the unsaturated layer of the soil is described by the one dimension Richards equation and the saturated base flow is based on the principles of the TOPMODEL and the surface runoff is calculated by an experience equation. The soil column is divided into 5 layers and we calculate the water balance for each layer. Background data, include vegetation, soil texture and micrometeorology, are come from the interpretation of the Quick Bird image and the field surveying. The cell size of Dem data, with the scale of the 1:100,000, is 50m×50m. Compared the simulation result with the field data which is got by the oven-dried way from 20 June to 7 July and the 17 debris flow events of the 2001, we have some conclusions: 1) The simulated data has the same variation trend with the field data with the same magnitude during surveying time; 2) The history precipitation was accumulated in the soil near the land surface; 3) The average basin's saturation degree (ABSD) may be a quite good index to evaluate the soil water conditions and the value of the 70% of the ABSD is the critical soil water condition of the initiation of the debris flow; 4) When the soil is wetted enough(ABSD>=70%), the debris flow in Jiangjia Ravine can be easy trigged by the precipitation of bigger than 25 mm per day or by the precipitation between 5 mm to 25 mm with short time heavy rainfall which can generate more than 0.35 m^3/s unit discharge. The model built in this paper provides a physical basis for the understanding of the debris flow initiation conditions of Jiangjia Ravine.

Keywords: Debris flow, Soil moisture content, Initiation conditions, Modeling, Jiangjia Ravine.

1 Introduction

Many studies have reported the triggering rainfall of the debris flow in the different area of the world. In these reports, the triggering rainfall generally has two parts:

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 1–11, 2013.

antecedent rainfall and real-time rainfall. The antecedent precipitation is accumulated in the soil, the more water in the soil; the easier the debris flow will be triggered. The affection of the antecedent rainfall is not easy to be determined because of the complex hydrology process. While to get the water stored in the soil accurately is very helpful to determine the trigging threshold of debris flow, many works have been done for this purpose. There are some studies focuses on modeling the soil moisture content (SMC) variation in response to the antecedent precipitation. Wilson [1] simulated the relationship between the rainfall and the shallow-hillslope pore pressure based on "Leaky-Barrel" model, Fiorillo [2] used "Thornthwaite" model and "Leaky-Barrel" model to simulate the soil moisture levels associated with the antecedent rainfall. In Jiangjia Ravine, which is an active debris flow watershed located in the southwest of China, there are also several works have been done to get the relationship of the antecedent precipitation and the SMC. Cui Peng etc. [3] used oven-dried method to get the SMC and found that the soil moisture varies greatly and the attenuation coefficient of antecedent rainfall is 0.78, Wei Fangqiang etc. [4] draw a calculation method to evaluate the effective antecedent rain through the relationship analysis of the soil moisture content and the antecedent precipitation. The infiltration character of the soil is one of the main factors that affect soil moisture content, because the ability of infiltration of the soil determines the rate and the magnitude of the transfer of the rainfall to the soil water. Wang Yuyi [5] analyzed the process of infiltration through the experiments in the lab with soil of the Jiangjia Ravine. Chen Ningsheng [6] got a infiltrate equation through the experiments with the artificial precipitation in the field of the Jiangjia Ravine. Chen Xiaoqin etc.[7] found that the infiltrate rate s varied with the vegetation type in the Jiangjia Ravine. Another triggering condition of the debris flow is the real-time rainfall. Field observations show that the debris flows in Jiangjia Ravine are initiated by the slide of the saturated top soil or the unsaturated top soil with the short time heavy rainfall. Yubin's [8]report shows that in Jiangjia Ravine when the unique discharge is up to the 0.35 $m^3/(s \cdot m)$, the surface runoff, which is generated by the real-time rainfall, can initiate debris flow.

In this paper, a hydrology model is developed, more factors, such as topography, infiltration, evapotranspiration (ET), surface runoff and the subsurface flow, are involved. The objects of this study are to capture the hydrology response to the rainfall and hope to be helpful in providing a physical basis for the understanding of the debris flow initiation conditions of Jiangjia Ravine.

2 Model Description

The simulated area is limited in the boundary of a watershed. Vertically, the simulation extends from the bottom of the saturated zone to the surface of the soil. Within a watershed, the soil system is divided into basic spatial units according to the scale size of the digital elevation model (DEM), which is 50m×50m. Each cell is treated as a unique system, except the ground and runoff water exchanges. Basic simulation is carried out at the cell scale. According to the need of simulation, the soil column is divided into five layers, saturated zone and unsaturated zone with 4 layers. The ET, unsaturated water movement, saturated base flow and surface runoff have to be considered in this model. Precipitation, topographic parameters, soil properties, and

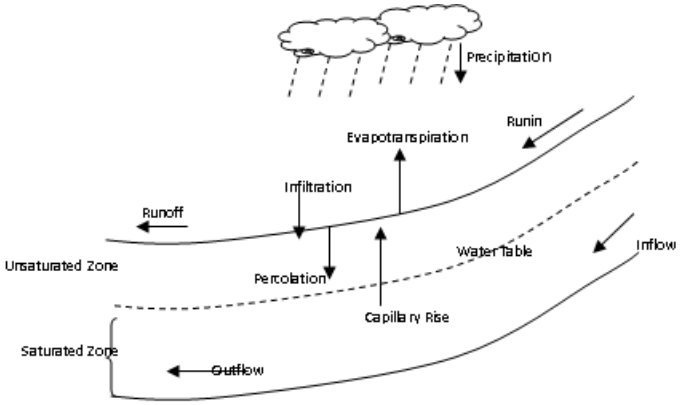


Fig. 1. Model structure

microclimate properties are the major inputs of the model. All input parameters are sampled to the same resolution as the DEM. The soil moisture and the discharge of the surface runoff are the outputs of the model (Fig 1).

2.1 Evapotranspiration

According to the interpretation of the Quick Bird image, the forestland, grassland, bare land and farmland, are the main landscapes in Jiangjia Ravine, four of them are mixed together and the grassland is the biggest one（31.24%）[9](Fig 2).

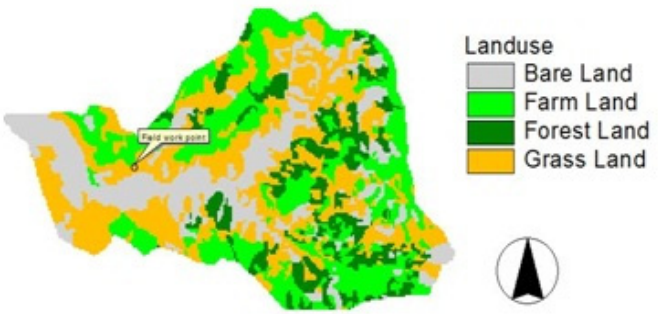


Fig. 2. The land use and the field work point in the Jiangjia Ravine

When we compute the ET, the four landscapes are simplified to the grassland because of the limitation of the input data. The potential ET is computed with the simplified Penman-Monteith equation, which is recommended by the FAO.

$$ET = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T_a + 273} u_2 (e_s - e)}{\Delta + \gamma(1 + 0.34u_2)} \tag{1}$$

Where, R_n is the net absorbed radiation; G is the soil thermal flux; $\Delta$is the rate of change of the saturated water vapor pressure with temperature; T_a is the air temperature;$\gamma$is the psychrometric constant; e_s is the saturated water vapor pressure; e is the actual air water vapor pressure; u_2 is the wind speed at the place where 2 m from the land surface.

2.2 Unsaturated Water Movement

The soil column is divided into five layers. The water movement in the soil is shown as Fig. 3.

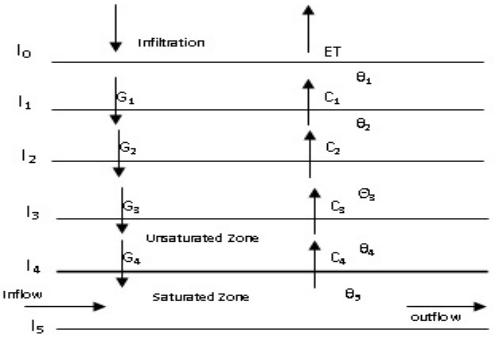


Fig. 3. The water movement in the different soil layers. I_i(i =1,2,…,5) is the interface of the different soil layers, G_i(i =1,2,…,4) is gravity water the moved in the different soil layers, C_i (i =1,2,…,4) is the capillary water in the different layers, θ_i(i =1,2,…,4) is volumetric water content of the different layers.

It is assumed that there is no lateral flow in the unsaturated soil. So the water movement in the soil can be described by the one dimension Richard's equation [10]:

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}[K(\theta) + K(\theta)\frac{\partial \psi(\theta)}{\partial z}] = \frac{\partial}{\partial z}[K(\theta) + D(\theta)\frac{\partial \theta}{\partial z}] \tag{2}$$

Where θ is soil moisture content, $K(\theta)$ is the hydraulic conductivity, $\psi(\theta)$ is the hydraulic suction, $D(\theta)$ is the hydraulic diffusion coefficient.

$K(\theta)$ is calculated by

$$K(\theta) = K_s (\frac{\theta}{\theta_s})^{2b+3} \tag{3}$$

Where θ_s is saturated soil moisture content, K_s is hydraulic conductivity of the saturated soil, b is a coefficient related with soil type.

$D(\theta)$ is calculated by

$$D(\theta) = \frac{K_s \psi_s b}{\theta_s} \left(\frac{\theta}{\theta_s}\right)^{b+2} \tag{4}$$

Where, b is the soil parameters related with soil type.

2.3 Saturated Base Flow

Each pixel of the watershed can exchange water with its adjacent neighbors. A moving window of 3×3 pixels was used to estimate the quantity of water exchange with its neighbors. The central pixel of the moving window was put on the any given pixel of the watershed, so, for this given pixel, there are up to eight directions of water output or input, and the flow direction was determined from the gradients in height from this pixel in question to down slope or upslope pixels. Along the direction with the steepest gradient among the eight neighboring pixels, the quantity of subsurface flow was the highest. For example, the rate of saturate subsurface flow at time t from pixel (i, j) to its downs lope neighbors is equal to

$$\begin{array}{ll} Q_{ij} = T_{ij}\beta_{ijk} w_{ijk} & \beta_{ijk} < 0 \\ 0 & \beta_{ijk} \geq 0 \end{array} \tag{5}$$

Where $T_{i,j}$ is the hydraulic transmissivity (m^2d^{-1}); β_{ijk} is the elevation gradient difference in the k direction; and w_{ijk} is the effective width of flow in the k direction. The total saturated subsurface outflow from pixel (i, j) to its downs lope neighbors ($Q_{out,ij}$) and total saturated subsurface inflow from upslope neighbors to pixel (i, j) ($Q_{in,ij}$) are approximated through the summation for all neighboring pixels

$$Q_{out,i,j} = T_{ij} \sum_{k=1}^{8} |\beta_{ijk}| w_{ijk}, \qquad \beta_{ijk} < 0 \tag{6}$$

$$Q_{out,i,j} = T_{ij} \sum_{k=1}^{8} |\beta_{ijk}| w_{ijk}, \qquad \beta_{ijk} < 0 \tag{7}$$

where the hydraulic transmissivity $T_{i,j}$ is defined as

$$T_{ij} = K_{sat}(z_{ij})[\exp(-f_{ij} z_{ij}) - \exp(-f_{ij} D_{ij})] / f_{ij} \tag{8}$$

Where $K_{sat}(Z)$ is the saturated hydraulic conductivity at water table $Z_{i,j}$; $D_{i,j}$ is the total depth; and $f_{i,j}$ is a parameter related to the decay of saturated conductivity with depth(m^{-1}). [11]

At each daily time step, this window is moved across the modeling domain to update the water table of each pixel as a result of the net lateral base flow in all eight cardinal directions.

2.4 Surface Runoff

Surface runoff discharge is calculated with the equation:

$$W_w = C_w I^2 t A \tan\theta R_1 r \tag{9}$$

Where, W_w (ml) is the surface runoff discharge, C_w is the runoff coefficient and is assumed to equal 5 in this area, I is the rainfall density (mm/min), t is the length of the rainfall time (min), A is the basin area(m^2),θ is the slope degree, R_1 is runoff coefficient associated with land use type (14 for the farmland and bare land, 1.4 for the grass land and forest land), r is volume soil moisture content.

The unit discharge is calculated by:

$$q_c = \frac{Q_0 + W_w / 60}{B} \tag{10}$$

Where q_c (m^3/s.m) is unit discharge which is generated by the heavyset density of 10 minutes in a rainfall, W_w is surface runoff discharge, Q_0 (m^3/s)is bare flow and is assumed to equal 0.28 m^3/s in this basin[8], B is width of the outlet point(m).

2.5 The Water Balance in the Soil

The water balance of each grid cell of the watershed is given as:

Water balance for the layer 1:

$$\Delta W_1 = Infiltration - G_1 + C_1 - ET \tag{11}$$

Water balance for the layer 2:

$$\Delta W_2 = G_1 + C_2 - G_2 - C_1 \tag{12}$$

Water balance for the layer 3:

$$\Delta W_3 = G_2 + C_3 - G_3 - C_2 \tag{13}$$

Water balance for the layer 4:

$$\Delta W_4 = G_3 + C_4 - G_4 - C_3 \tag{14}$$

Water balance for the layer 5:

$$\Delta W_5 = G_4 + Inflow - C_4 - Outflow \tag{15}$$

3 Data Preparation

Data used in the model include micrometeorology data, soil texture, topography and the vegetation type. The micrometeorology data used in the equation 1, including precipitation, temperature, humidity, radiation and wind speed, are from the Dongchuan Debris Flow Observation Station, Chinese Academy of Science. The soil thermal flux is supposed to be the line function of the net radiation (G=0.1R_n). DEM data with the cell size of 50m×50m, is from the topographic map with the scale of 1:100,000.

The soil in the Jiangjia Ravine is the wide graduation gravel soil, so the soil porosity varies greatly, here this parameter is determined by the land use type [11], the porosity of the bare land is supposed to be 48.9%, the farmland is 46%, the grassland

is 39.7% and the forestland is 49.7%. Wang [5] had a simple regression equation to describe the relationship of the hydraulic conductivity of the saturated soil (K_s) and the soil porosity(φ) in this area, which is: $K_s = 0.0126\times\varphi+0.0032$.

Most of the debris flows in Jiangjia Ravine are initiated by the slide of saturated topsoil with the thickness of 20-30 cm. Cui and Wei's [3,4] field data also showed that common density precipitation can only affected the topsoil with thickness of the 40 cm. So, the thickness of the each unsaturated layer is supposed to be 30 cm, the saturated soil layer is from the water table to the bedrock.

The water table and soil moisture before the wet season were initialized by the TOPMODEL principle. The wetness index (W_{ij}) for a pixel at (i, j) was calculated as:

$$W_{ij} = \ln(\frac{A_{ij}}{\tan\beta_{ij}}) \tag{16}$$

where A_{ij} is the contributing area which can be calculated from the DEM, β_{ij} is the slope. The initial water table ($z_{t,ij}$) was linearly related to the wetness index:

$$z_{t,ij} = \overline{z_t} + m(W_{ij} - \overline{W}) \tag{17}$$

Where $\overline{z_t}$ is the mean water table, supposed to be 0.8m, $\overline{W}$ is the mean wetness index, supposed to be 0.7. At the beginning of the rainy season, the soil moisture content is supposed to be at the wilting point.

4 Simulation Results and Discussion

Cui Peng etc.[3] measured the SMC at a grassy slope with the oven-dried method from 20, June to 8, July 2001 in Jiangjia ravine (Fig 2). Because the measured data is a mass water content, so it was changed to volumetric water content with the equation of $\theta_{v=}\theta_{m\times}\rho_s$, where θ_v is volumetric water content, θ_m is mass water content andρ_s is soil density, here it is assumed to equal to 2.0g/cm^3. Fig. 4 (a) is the sketch map of the field data and the modeled data. Generally speaking, the modeled data agrees with the measured data quite well, it has the same variation trend with the field data with the same magnitude during surveying time, but the modeled data at the beginning is quite lower than the field data, it may be caused by the following reasons: (1) the inaccuracy of precipitation data, it was found that precipitation data that used in the model had 5 days difference with data that were presented in Cui Peng's [3] paper during the field time (17 days) and this may caused by the mistake of statistic; (2) the inaccuracy of the change of the soil water content, ρ_s was assumed to be a stable value, but the fact is thatρ_s is always changed with soil water content; (3) the inaccuracy of parameters of the soil texture used in the model, some parameters of the soil are assumed to be stable associated with the soil texture, but character of the soil is quite different in Jiangjia Ravine.

The simulation results of the top 4 layers of the surveying point are shown in Fig 4. (b) The SMC of the 1st layer has the fastest response to the precipitation; the variation of the SMC of the 2nd layer lags the 1st layer, it begins to catch up with the 1st layer at DOY 152, but still lags in phase with narrow amplitude. The SMC of

the 3rd layer begins to rise at the DOY 165; The SMC of the 4th layer stays stable and has a little change after the medium of August and this shows that precipitation in the rainy season cannot affect this layer and the antecedent precipitations are accumulated in the top 3 layers (90 cm in the model). It is found in the fieldwork in Jiangjia Ravine that medium density rain can only affect the SMC of topsoil about 40 cm[12], and about 80 cm of the soil can be affected by the rain during the whole rain season[13].The curves in the Fig 4 (b) show that the variations of the modeled data in the different layers agree with the report of the fieldwork quite well.

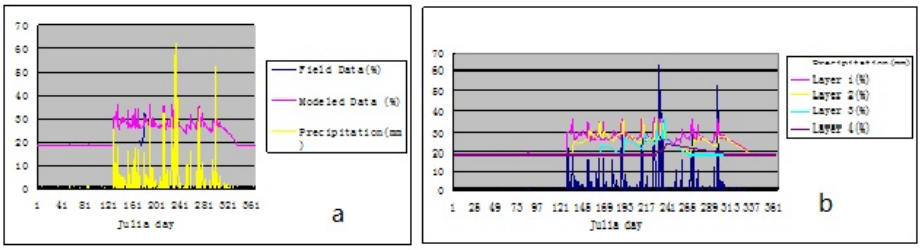


Fig. 4. The field data and the modelled data

5 The Initiate Conditions of the Debris Flow of the Jiangjia Ravine

Like other debris flow which is initiated by the rainfall, the initiation conditions of the debris flow in Jiangjia Ravine have two parts: the antecedent precipitation and the real-time precipitation. In the fact, the antecedent precipitation is accumulated in the soil, the real-time precipitation maybe generate the surface runoff, so the conditions can also be described as: the soil should be wetted enough by the antecedent precipitation and enough strong surface runoff.

5.1 The Initiate Conditions of the Antecedent Rainfall

The antecedent rainfall is accumulated in the soil column with the thickness of about 80 cm (in the field surveying) or 90 cm (in the model), so the water in the soil column with such thickness should be considered together to evaluate the accumulation of the antecedent rainfall. Here, the average basin's saturation degree (ABSD), which is induced from the SMC of the soil of the top 3 layers in the model, is used as the index to analyze the initiation conditions of the debris flow. Firstly, average layer's saturation degree (ALSD) is induced from the average saturation degree of all the cell in one layer based on the definition of the saturation degree: saturation degree = soil water volume/ soil pore volume, then, the ABSD is calculated with the average of the ALSD of the top 3 layers of the model. Fig 5 is the relationship of the ALSD of the top 3 layers, the ABSD, debris flow events and the precipitation. The ALSD of each layer has the different variation trend, and the ABSD begins to raise to high level

from the beginning of the rainy season to the medium of July, and tries to stay there with severe fluctuation due to the variation of the precipitation till to the end of August, then it begins to decrease slowly along with the reduction of the precipitation and stays stable in the low level at last.

Chen Xiaoqing's[14] point field experiment in this basin shows that when the soil's saturation degree is up to 70%, the soil will collapse and can be initiated easily to form the debris flow. This value is retrieved from the single point and it is also supported by the simulation result in this paper. There were 17 debris flow events in 2001 in this area and most of them occurred between the medium of June and the end of August. During this period, the ABSD keeps close to or higher than 70%, while before or after that, the ABSD keeps lower than 70% and there is no debris flow although it has medium to big rainfall in the days showed in the Tab.1. This means that debris flow can only happened when the soil was wetted enough by the accumulation of the antecedent rainfall. Thus, the ABSD maybe can be used as the index to evaluate the initiation conditions of the SMC for the debris flow, and 70% is the critical value of the initiation conditions of debris flow according to this index.

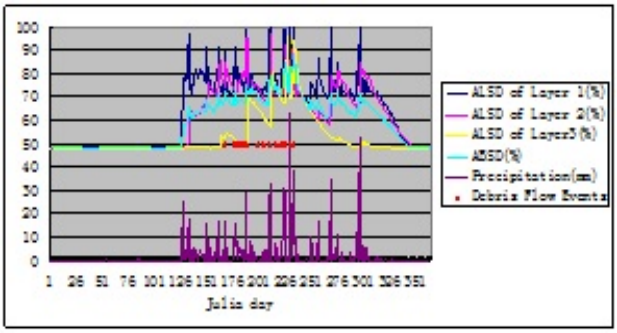


Fig. 5. The ALSD of the different layers and the ABSD

5.2 The Initiate Conditions of the Real-time Rainfall

Debris flow in Jiangjia Ravine can be initiated by the short time high intensity precipitation. The simulation result shows that when the precipitation is lower than 5 mm per day, it can only affect topsoil, while those bigger than 5 mm per day can cause obvious variation of the SMC of the 2nd layer, so 5 mm may be a critical point of the precipitation, which can affect deep soil. It also shows when the the precipitation is higher than 25 mm per day; the topsoil will be saturated by the real-time precipitation. It is recommended by Yubin's[8] paper that when the unit discharge in the Jiangjia Ravine is up to 0.35 m^3/s, the matters in the gully can be initiated by the surface runoff. The debris flows in the DOY 177,180, 182, 186,187, 200, 205, 217, and 222 belongs to this type. In these days, the soil has been wetted by the antecedent precipitation and the ABSD is nearly to 70%, and the most of precipitations are between 5mm and 25 mm per day with the short time high intensity rainfall which will generate more than 0.35 m^3/s unit discharge (Tab.2).

Table 1. Several big rain in 2001 which did not initiate debris flow

Date	Precipitation （mm）	The ABSD（%）
8,May	25.6	48.1
14, May	17.7	60.25
31,May	16.4	63.3
12, June	17.4	65.05
15, September	16.5	63.4
24, September	17.2	60.5
26, September	34.5	64.6
24,October	54.2	68.3

On the conditions of the soil is wetted enough by the antecedent precipitation, the debris flow in Jiangjia Ravine can be easy triggered by the precipitation which is higher than 25 mm per day. The debris flows of DOY 168, 178, 185, 189, 211, 225, 231, and 234 were initiated by this kind of precipitation. In these days, the soil is enough wet and the ABSD is higher than 70%, although some of the unit discharge generated by maximum 10 minutes heavyset precipitation is lower than 0.35 m^3/s. (Tab.2)

Table 2. The unit discharge and the ABSD of the debris flow events in 200

NO.	DOY	Unit discharge (m^3/s)	ABSD(%)	NO.	DOY	Unit discharge (m^3/s)	ABSD(%)
1	168	0.7	72.2	10	200	0.22	70.8
2	177	0.7	67.1	11	205	0.41	67.4
3	178	1.42	72.3	12	211	0.25	73.9
4	180	4.6	68.7	13	217	2.6	75.2
5	182	0.39	67.6	14	222	1.7	71.5
6	185	0.02	68.4	15	225	3.77	80.2
7	186	0.3	67.3	16	231	1.784	80.2
8	187	0.33	68.0	17	234	0.15	88.7
9	189	0.3	75.6				

6 Conclusion

A simulation model, based on the physical process of hydro cycle is built to simulate the the hydrology response to the rainfall in Jiangjia ravine in 2001. The simulation result shows that the history precipitation was accumulated in the soil near the land surface; the ABSD is a quite good index to evaluate the accumulation of the antecedent precipitation. When the soil is wetted enough(ABSD>=70%), the debris flow in Jiangjia Ravine can be easy trigged by the precipitation of bigger than 25 mm per day

or the precipitation between 5 mm to 25 mm with short time heavy rainfall which can generate more than 0.35 m^3/s unit discharge. The debris flow in Jiangjia Ravine cannot be easy triggered when the soil is fairly dry (ABSD < 70%) and when there has no enough precipitation (< 5 mm per day).

Acknowledgment. This paper is supported by the Ministry of education of Humanities and social science fund plan (12YJAZH124) and the Fundamental Research Funds for the Central Universities (A0920502051308-12).

References

1. Wilson, R.C., Wiezoreck, G.F.: Rainfall thresholds for the initiation of debris flows at La Honda, California. Environ. Eng. Geosci. 1(1), 11–27 (1995)
2. Fiorilloa, F., Wilson, R.C.: Rainfall induced debris flows in pyroclastic deposits, Campania (southern Italy). Engineering Geology 75, 263–289 (2004)
3. Cui, P.: Relationship between occurrence of debris flow and antecedent precipitation taking the Jiangiia Gully as an Example. Science of Soil and Water Conservation 1(1), 1–11 (2003)
4. Wei, F., Hu, K., Chen, J.: Determination of Effective Antecedent Rainfall for Debris Flow Forecast. Journal of Mountain Science 23(4), 453–457 (2005)
5. Wang, Y., Zou, R.: Interrelated research of relationship between debris flows trigger and permeation coefficient. Journal of Soil Erosion and Soil and Water Conservation 3(4), 76–82 (1997)
6. Chen, N., Zhang, J.: The research of permeability on lose gravelly soil in debris flow original area. Journal of Mountain Science 19(1), 169–171 (2001)
7. Chen, X., Cui, P., Hen, J., Feng, Z., Wang, D.: Infiltration Experiment of Wide-graded Gravel Soil in the Jiangjia Ravine. Journal of Mountain Science 24(7), 169–171 (2006)
8. Yu, B., He, S.F., Hong, Y.: Research on runoff and sediment in catchment of debris flow under rainfall. Journal of Soil and Water Conservation 15(3), 72–75 (2001)
9. Lin, Y.-M.: Study on Interactions of Landscape Change and Soil erosion in the Typical Watershed of Jinsha River, pp. 15–16, Ph.D Dissertation of Chinese Academy of Sciences (2008)
10. Liu, S.: Environmental Physics, pp. 128–140. Chemical Industry Press, Beijing (2004)
11. Chen, J.M., Chen, X., et al.: Distributed hydrological for mapping evaportranspiration using remote sensing inputs. Journal of Hydrology 305, 15–19 (2005)
12. Wu, J., Tian, L., Kang, Z.: Debris Flow and Its Comprehensive Mitigation, pp. 20–27. Science Press, Beijing (1993)
13. Wu, J., Kang, Z., Tian, L.: Debris Flow Observation and Research in Jiangjia Ravine, pp. 20–26. Science Press, Beijing (1990)

A City Construction Land Pile Foundation Adaption Fuzzy Comprehensive Evaluation

Chuang-ye Feng, Zhi-chao Zhao, Xue-mei Li, and Hui Li

Hebei Prospecting Institute of Hydrogeology and Engineering Geological
050021 Shijiazhuang, China
Fcy1970@163.com

Abstract. Through depth analyze and research the engineering geological conditions of a city construction land, use fuzzy comprehensive evaluation method, Quantitative analysis of engineering geological conditions of urban construction land suitability. this paper divided existing land into the suitability good area, good suitability District, General District suitability and appropriateness of poor area four categories, mainly Analyze how to choice low-rise building , high-rise buildings, and high-rise construction, Provide strong scientific basis for the design and planning.

Keywords: pile foundation, Fuzzy Comprehensive Method, evaluation suitability.

This study combined with Hebei water Institute survey of a city urban geological investigation and evaluation of projects carried out. Geological Survey of the city is a major project, The engineering services in a municipal social and economic sustainable development for the purpose, Tightly around a city facing and the urban geological problems to be solved, according to the needs of a city 's overall urban planning, tightly followed the process of urbanization pulse, Give full play to the advantage of the disciplines of geological work, To carry out comprehensive geological survey of the city three - dimensional geological structure, geological resources and geological environment , comprehensive evaluation of urban development resources and support capacity of the environment and urban security , to provide geological information for the overall urban planning and urban construction and management , industrial layout and material project site to provide basis for decision making, and to provide basic protection for sustainable urban development[1] .

In this paper, the theoretical significance is reflected in :explore a suitable Construction land evaluation method for Shijiazhuang, combined with types of pile foundation 15 - 20m depth evaluation, the evaluation results more guidance construction.

The scope of this study is a city of New Area, north of the Hutuo, East of the Beijing-Zhuhai Expressway, west of the Airport Road, south of the Southern cattle

[1] Xiaolan Yang. Nanchong geological evaluation of the environmental suitability of the urban construction[J]. Sichuan Geologica Acta1997, 17(4):298-302.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 12–19, 2013.

Township - North San board – lijiazhuang, the area is 142 square kilometers. Take full account of the status quo conditions, Combined with urban planning, Fuzzy comprehensive evaluation method to establish multi-level evaluation system.this paper aim at the specific circumstances of the study area[2], Considering various factors, Selected five primary factors and ten secondary factors to evaluate comprehensive, Tentative construction land suitability evaluation system engineering geological conditions for suitability, good, poor, poor in four grades.

We would like to draw your attention to the fact that it is not possible to modify a paper in any way, once it has been published. This applies to both the printed book and the online version of the publication. Every detail, including the order of the names of the authors, should be checked before the paper is sent to the Volume Editors.

1 Evaluation of Pile Foundation'S Determining Principle

Study area of pile foundation vertical bearing capacity determining method:First assume that the selected pile is reinforced concrete pile, Pile diameter is 800mm;then static Soil distribution the range of 15- 20m,and then check Hebei Province, engineering construction standard DB13 (J) 31-2001 " auger drilling pump pressure concrete piles composite foundation of technical regulations in Appendix A, Appendix B,determine the range of 15 - 20m depth the limits of the pile side resistance standard value q_{sk} and pile tip resistance standard value q_{pk}. According to the formula (1) to determine the vertical bearing capacity standard value . Pile side resistance calculation process which need to be coupled with 5 - 8m ,8 - 15m soil part .

<<Building Pile technical specifications>> Recommended standard values of ultimate bearing capacity calculation method as shown in equation (1):

$$Q_{uk} = u_p \sum q_{sik} l_i + q_{pk} A_p \quad (1)$$

Formula: Q_{uk}—singal pile Vertical limit bearing capacity standard value;

u_p—Pile cross-section perimeter;

q_{sik} —Pile side of the i - layer soil limit side resistance standard value, value in Appendix A;

q_{pk} —Limit end of the pile end soil resistance value of the standard value, in accordance with Appendix B;

l_i —I - layer thickness of the soil;

A_p —The pile side bottom area.

[2] Dewei Wang.Yibin City Geological Environment Evaluation[D].Chengdu University of Technology,2006,(12).

2 Establish Evaluation Index Set

The fuzzy comprehensive evaluation method is a very effective method which affected by many factors.To judge the pros and cons of the suitability of the land for construction,should first establish the appropriate evaluation factors set,To this end take in Figure 1 Suitability Evaluation Index System 12 index factors as the suitability evaluation factors set,namely:U={C1,C2, ……,C12}.

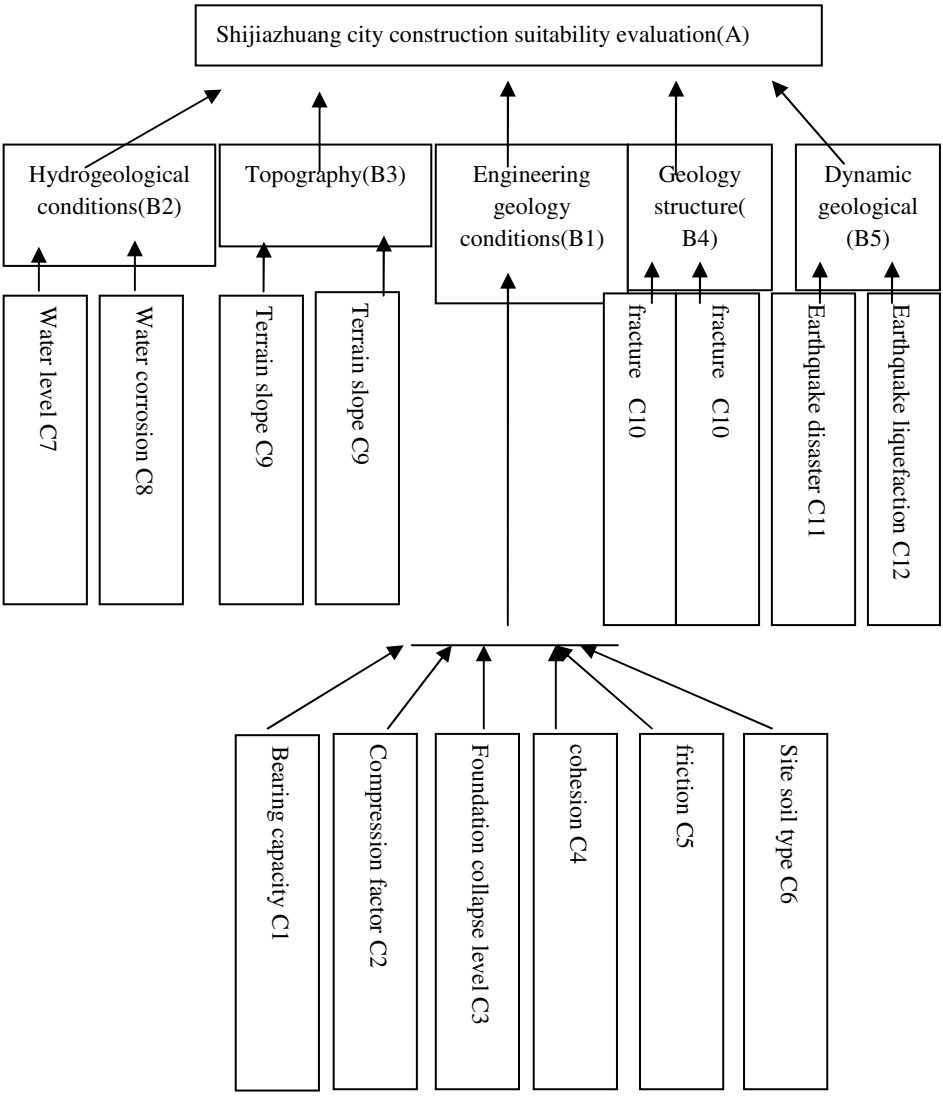


Fig. 1. Shijiazhuang construction land suitability evaluation analysis model

3 Establish Reviews Set V and Indicators of Membership on the Reviews Set

Construction land suitability evaluation set for continuous interval [0,1], urban construction land suitability level : suitable for (0.75 ≤Sj< 1); more appropriate (0.5 ≤ Sj< 0.75) ; less suitable (0.25 ≤Sj < 0.5); suitability Poor (0 ≤Sj < 0.25) .

Single factor evaluation for indicators set indicators,Resulting membership of the corresponding factors on the V :0≤ai≤1.ai Size indicates that the factors of construction land suitability how good or bad the vague concept of the degree of compliance[3], And provides a factor the greater the value of membership,That the factors the closer the goal of the best suitable areas , and vice versa away from the goal of the best suitable areas .

To establish the basis of membership function is to establish mathematical model,it is the key to part of the urban construction land suitability evaluation rating reflecting the individual factors,according to various factors quantized data demographics,In this paper, the trapezoidal membership function (Equation 2) mathematical model to determine the membership of these factors (Figure 2).

$$\alpha_i(x)=\begin{cases}0 & x\le a\\ \dfrac{x-a}{b-a} & a<x\le b\\ 1 & x\ge b\end{cases} \quad (2)$$

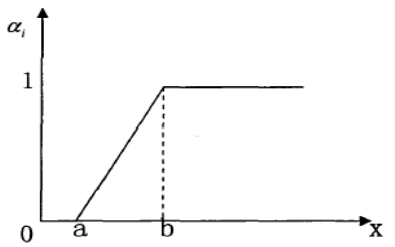


Fig. 2. Membership function curve

Note: X in the figure as the actual value of the evaluation factors, a, b, c for the factors of the upper limit or the lower limit value, ai is Membership of the factors.

4 Establish Fuzzy Evaluation Matrix

This article is divided Shijiazhuang urban construction land into suitability suitable more appropriate suitability poor , poor suitability of four levels of evaluation, Therefore, fuzzy matrix arranged in rows 12 a factor 4 the suitability level membership value[4], The formation of a 12 × 4 matrix of order, in the form as follows:

---

[3] Songbai Xu.The principles of the Analytic Hierarchy Process[M] .Tianjin:Tianjin University Press1988.

[4] Baosong Liang.Dianli Cao. Fuzzy Mathematics and Its Applications.Beijing.Science Press,2007.12.page 127-166.

$$R=(R_1\ \ R_2\ \ R_3\ \ R_4\ \ )=\begin{bmatrix}\alpha_{11}\ \alpha_{12}\ \alpha_{13}\ \alpha_{14}\\ \alpha_{21}\ \alpha_{22}\ \alpha_{23}\ \alpha_{24}\\ \cdots\cdots\\ \alpha_{j1}\ \alpha_{j2}\ \alpha_{j3}\ \alpha_{j4}\\ \alpha_{n1}\ \alpha_{n2}\ \alpha_{n3}\ \alpha_{n4}\end{bmatrix}$$

Among n=12, R_j Level j of the construction land suitable area (j = 1 , 2 , 3 , 4) fuzzy evaluation vector: $R_j=(\alpha_{1j}\ \alpha_{2j}\ \alpha_{3j}\ \alpha_{4j})^T$, α_{ij} is Control of the i-th this level impact factors on the level j suitability.

5 A Comprehensive Evaluation of the Suitability of the Natural Foundation of the Construction Land

To make each index construction land suitability is reflected, Must be integrated into the evaluation set each index weight and membership, Ordinary matrix multiplication to make construction land suitability level fuzzy comprehensive evaluation: $S=WR=(S_1\ S_2\ \_\ S_j)$, S_j (j=1,2,3,4) Corresponding geologic unit suitability to be evaluated fuzzy comprehensive evaluation results, The results take the S_j maximum the district for the evaluation unit suitable level area[5].

6 Pile Foundation Suitability Classification and Quantitative Criteria of Evaluation

According to the above evaluation of the principle of selecting establish numerical criteria, According to collected geological data and the actual extraction of each index value to establish a database, through scientific, systematic and comprehensive analysis to determine the value of each geological environment quality evaluation index grading standards, finally combined with the construction experience, Reference engaged in an urban geological environment monitoring, engineering geological exploration, urban planning expert opinions and views, Comprehensive analysis the development and distribution of the main geological factors of the nature of geotechnical engineering , groundwater , active faults and Geoenvironmental characteristics and the development and utilization of underground space , a city construction land suitability classification :suitable , comparable suitability, suitability comparable poor and suitability poor , as shown in Table 1 .

[5] Weishu Xiao.Fuzzy math basic and applied[M].Beijing:Aviation Industry Press, 1992:63-72.

Table 1. Pile foundation suitability influencing factors classification

sort	Evaluation factor	Suitability level			
		suitability	comparable suitability	suitability comparable poor	suitability poor
Engineering geology	15-20mPile vertical bearing capacity /(MPa)	>3.2	3.2-2.5	2.5-1.8	<1.8
	20-25mPile vertical bearing capacity /(MPa)	>3.9	3.9-3.2	3.2-2.5	<2.5
	25-30mPile vertical bearing capacity /（MPa)	>4.6	4.6-3.9	3.9-3.2	<3.2
	30-65mPile vertical bearing capacity /（MPa)	>5.3	5.3-4.6	4.6-3.9	<3.9
	Compression factor C2	<0.2	0.2-0.3	0.3-0.5	>0.5
	Pile bearing stratum thickness C3	>12	12-7.5	7.5-3	<3
	Cohesion C4	>25	25-15	5-15	<5
	Friction C5	>35°	35°- 25°	25°- 15°	<15°
	Site soil C6	Ⅰ level Assignment 10-7.5	Ⅱ A level Assignment 7.5-5	Ⅱ B level Assignment 5-2.5	Ⅲ level Assignment 2.5-0
Hydrogeological	Water level/(m) C7	>15m	15-10m	10-5m	<5m
	Water corrosion C8	no Assignment 10-7.5	weak Assignment 7.5-5	center Assignment 5-2.5	strong Assignment 2.5-0
Topography	Slope ratio C9	<5°	5°- 10°	10°- 20°	＞20°
Geological structure	fracture C10	no	little	Little more	Holocene active fault

Table 1. (*Continued*)

Dynamic geological	Earthquake influence C11	Assignment 10-7.5	Assignment 7.5-5	Assignment 5-2.5	Assignment 2.5-0
		no	weak	Little strong	More strong
	Earthquake Liquefaction C12	Assignment 10-7.5	Assignment 7.5-5	Assignment 5-2.5	Assignment 2.5-0
		no	weak	Little strong	strong
		Assignment 10-7.5	Assignment 7.5-5	Assignment 5-2.5	Assignment 2.5-0

7 15-20m Pile Foundation Suitability Evaluation Results

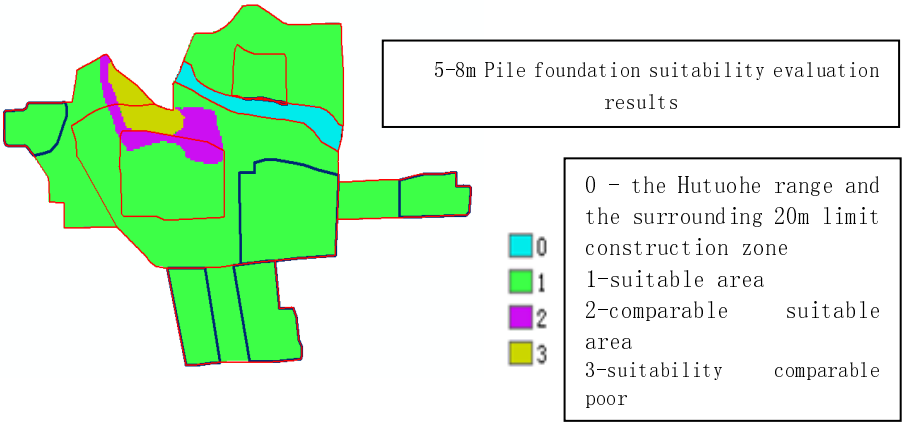


Fig. 3. 15-20m pile foundation suitability evaluation results

8 Conclusion

Through in-depth analysis and research on engineering geological conditions of the urban construction land, construction land suitability of the study area is a comprehensive evaluation of the system , the establishment of urban construction land suitability evaluation system, the fuzzy comprehensive evaluation model is introduced into the suitability of land for construction evaluation obtained engineering geological evaluation and prediction of the environmental quality of the evaluation results .

Natural foundation bearing stratum visual evaluation results can be seen, most of the study area of construction land area is more appropriate suitability poor are mainly distributed in the mountainous and Hutuo he Valley and the floodplain areas, to deal with these areas for afforestation, strengthen green.

References

1. Yang, X.: Nanchong geological evaluation of the environmental suitability of the urban construction. Sichuan Geologica Acta 17(4), 298–302 (1997)
2. Wang, D.: Yibin City Geological Environment Evaluation. Chengdu University of Technology (2006)
3. Xu, S.: The principles of the Analytic Hierarchy Process. Tianjin University Press, Tianjin (1988)
4. Liang, B., Cao, D.: Fuzzy Mathematics and Its Applications, pp. 127–166. Science Press, Beijing (2007)
5. Xiao, W.: Fuzzy math basic and applied, pp. 63–72. Aviation Industry Press, Beijing (1992)

Agronomical Zoning in the Municipal Urban Plan and Viticultural Predisposition

Pier Luigi Paolillo* and Giuseppe Quattrini

Abstract. The analysis of agricultural spaces is usually defined by disciplines and by the agronomic profession of which, in the municipal urbanistic tools, the result of its in-depth analysis is almost never recognized (at least in Italy), neither as a resolute factor that impedes the waste of soil for urbanization purposes nor as a selective support of agricultural/productive zoning. Furthermore, in contemporary times, the fine-tuning of Geographical Information Systems technologies is so advanced that it permits the full understanding of the physical interdependencies. These interdependences are useful in the evaluation of the agronomy of a territory within the municipal plan, adopting, through the most advanced analytical boundaries of the GIS applications, the multivariate geostatistical analyses both for the calculation of the correlations between the numerous environmental, landscape, and territorial components as well as for the elimination of the subjectivity of the attitudinal judgment attributed by the analyst to different indicators.In such cases, especially in the judgment of viticultural behaviors, there is a delicate built equilibrium in the relationship between the physical resources (soil, air, water) and the productive profile most suitable. For this reason, in the case study the available information is translated from the continuous dimension to the discrete, contriving a matrix where each homogeneous, static pixel corresponds to a spatial cell, finalized at the simultaneous comparison of all of the parameters necessary to evaluate the viticultural discipline and the consequential multivariate analysis from which emerge the basins of different behavioral levels. The successful result of this application thereby allows for critique of the all too mechanical modality of Overlay utilized by Boyer (1998) like the multi-criteria approach followed by territorial analysts such as Malczewski (2004), Riveira (2006), Jones (2004), Tomasi (2013) retaining multivariate analyses to be more lucrative in the resolution of the problem. Such multivariate analyses include the analysis of the correlations and principal components together with the final cluster analysis, which allow the groupings of variability of intensity to be extracted, evaluating them by the interdependence between them. In this way, a calculation of viticultural behavior is constructed that presents a procedure characterized by superior reliability in comparison to the most diffused hierarchical multi-criteria classification. It is therefore able to eliminate whatever arbitrary, discretional, and hierarchical attribution of judgments (and consequential points) to the single descriptive parameters of physical resources, able to further amplify by a great deal the entity and quality of the variables adopted in the calculation, and, in the end, able to pour the results of the geostatistical results into the corresponding cells of the

* Politecnico di Milano, Dipartimento di architettura e studi urbani, via Bonardi 3, 20133 Milan (Italy), pierluigi.paolillo@polimi.it, http://paolillo.professor.polimi.it/

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 20–31, 2013.

analyzed space and in the subsequent trajectory cartography of the municipal urban plan.

Keywords: vineyard zoning in the municipal urban plan, multivariate geostatistical analysis, Geographic Information Systems.

1 Introduction

The Italian agricultural territory and its physical resources (soil, air, water) necessitate a profound understanding to be utilized in the urbanistic municipal plan, both for the containment of negative perturbations of new settlements and for the guidance of the urban transformations toward those spaces already compromised without wasting good productive land, beautiful agrarian landscapes, bio/geo/diversity, and also to guarantee the parsimonious use of environmental capital handed down from the important national agrarian history. Therefore, the analyses of the agroproductive behaviors must become part of the plan to finalize the contributions of land sciences, of the climatology and the environmental disciplines to the adoption of the best decisions that are of both urbanistic and agronomical character.

In the presented case study, the construction of a behavioral analysis method in the evaluation of the soil for a particularly prestigious agricultural cultivation, that of the wine grapevine, intends to reach the goal of disaggregating a an agricultural space that has until now been considered undifferentiated and almost of mono-functional character, when, on the contrary, it masks innumerable specificities, such to influence in a perpetuating the variety of cultivations and their productive results along with the conservation of the environmental quality of the soil resource. Furthermore, in the current state the demonstrated scientific maturity of the territorial analysis techniques through Geographic Information Systems is able to satisfy the environmental instances of a parsimonious society in the use of the nonrenewable resource and awareness of the incalculable economic damage deriving from many phenomena of abandonment, instability, pollution and compromising of soil in the Italian territory. Thus, the subsequent comparison between the different research methods – hierarchical versus multivariate – is intended to stimulate the debate on the best treatment of environmental databases, declaring the absolute demand of sedimentary objective protocols that can be reimplemented in the construction of the decisions of the municipal urban plan (Paolillo, 2012).

1.1 The Peculiar Productive Functions and Attitudinal Cartography

More than once, it has been recognized[1]: (*a*) that the intensity with which the environmental complexity is expressed does not depend much on the general rules but rather on the interaction of local phenomena; (*b*) that therefore it is necessary to discover locally the specific physiognomies, the substantial differences, the particular modes of such complexity; (*c*) that different problematic intensities of the local basins therefore should be researched, highlighting those particular phenomena that attract environmental damage and evaluating the signals of attention.

[1] Cfr. in Paolillo 1990, 1994, 1995a, 1997a, 2000a, 2000b.

A possible explorative path must therefore identify homogeneous groupings of local phenomena through: (*x*) the application of the *analyse des données*[2] for the better exploration of available archives; (*y*) the discovery of the qualitative relations routed in the techniques of multivariate analysis; (z) the conclusive identification of 'area type' in which interventions can be tested that intend to reduce environmental damage[3] or to apply dedicated measures. This leads to the problem of so-called 'zoning' and thus, through multivariate analyses the archive will be explored in order to extract 'groupings' of differentiated dependent variables to vary the objective (independent variable) prefaced by zoning, after which the calculation of the interdependency between variables will suggest geospaces of similar aggregation of ties and, as a result, vocational intensity (in this case agronomic).

Such modalities of recognition/classification could prove to be adoptable in urban planning (where it is necessary to find isotropic settlements with respect to the apparent condition of citizen anisotropy), this is not very diffused for extra-urban[4] factors, even though it could be practiced with tremendous success in the study of the vine/environment interaction and the multiple variables in the surrounding conditions (altitude, exposure, arrangement, geologic origin, soil, typology of cuttings, form of the system, etc.) that characterize the enological quality. However, an appropriate 'zoning,' must reconcile such an intricate framework of needs (sometimes unethical) concentrating on the areas more inclined to the production of wine and dedicating those spaces of lesser soil aptitude (traces of trees and shrubbery in support of the organoleptic characterization of wines, spaces of renaturalization, characteristic vegetation of the specific local landscape), setting forth analyses dedicated first and foremost to the identification of 'vocational units' within the boundary of vegetative, productive and, most importantly, qualitative performance of the vineyards can be considered to be sufficiently homogeneous, contributing in this way (within the technical assistance to the winemaker) and providing counsel to optimize the use of productive factors, increasing the qualitative yield of the grapes and the wines produced and solving the errors commonly committed in the vineyard. Furthermore, not to be understated is the promotional repercussion of the viticultural zoning for the qualitative guarantee, against the uniformity of the product for more efficient politics

---

[2] This is a reference to the French pillars of the *analyse des données* (or 'multidimensional analysis' or 'multivariate'): among others, Benzécri, 1973; Bertier and Bouruche, 1975; Caillez and Pages, 1976; Fenelon, 1981, other than the trad. it. of Gruppo Chadoule, 1983; for the italian applications, one can distinguish in particular the pioneers, Bellacicco, 1975, 1979, 1983; Bellacicco and Labella, 1979; Fano, 1978, Palermo and Griguolo, 1983; also evidenced are the important studies of Merlin, 1974 and of Nijkamp, 1977, 1984); it is about '*a group superfluous of multivariate statistical techniques that allow for the analysis, through the employment of quantitative methods, of social realities and complex phenomena*' (Fraire, 1994).

[3] It is evident with regard to Parte II (*L'eco-programmazione: proposte di metodo applicativo*) in Paolillo P.L., *ed*., 2000, *Terre lombarde. Studi per un ecoprogramma in aree bergamasche e bresciane*, Giuffrè, Milan; in particular, *ivi* Paolillo, P.L., «Una modalità descrittivo-classificatoria d'individuazione dei "*bacini d'intensità problematica ambientale*" alla scala regionale», pp. 103-153.

[4] The experiments that were carried out for the extra-urban zoning, dependent on urbanizing factors and for the tutelage of the physical soil resource, are uncovered in Paolillo, 1995b; 1998.

of territorial marketing, to develop wine tourism, for the rise in employment, for the permanence in the locales of safeguarding of the environment and traditions, histories, and native cultures.

2 Multidimensional Analysis in Support of Classification: The Spatial Clusters of Different Levels of Viticultural Attitude

The urbanistic and environmental approach, as in the analytical analyses of the disciplines of the sector, usually employs a procedure of topologic overlay[5] of various informational strata, leading to hierarchical and multi-criteria analyses to connect the results to the decision on the territorial phenomena in question, for example the viticultural agronomic behavior[6].

Of another origin is the applied multidimensional analysis, in this case, in an Italian municipality (San Giorgio in the Province of Piacenza, in the Region of Emilia Romagna) over the Apennine margin of the Po Valley and historically inclined to prestigious agricultural production thanks to the division of profound soil which is amongst the most fertile in Europe, in the flatlands, and sloping clay-silt hills where the viticular cultivation has retained a constant historical quota. In such a situation, which still presents amplified margins of agricultural and agro-landscape resources of notable interest, the analyse *des données* of multidimensional sort appears to be a simultaneous application to assist the economic operators in business decisions, stimulating, moreover, the agricultural market to differentiate production by categories of product and qualitative levels.

2.1 The Informational Reorganization for Spatial Analysis: The Continual Dimension to the Discrete Space in GIS Domain

The first step in the method introduced thus lies in the identification of geospatial typologies that consent the analysis of the correspondence and subsequent nonhierarchical classification (carried out with the appropriate software[7]) in order to progressively reduce the complexity of the different variables considered in the multivariate treatment.

The identification of the 'finite' portions of the territory where the appraised complexity of the factors, active in the environmental prism, represents the moment of the initiation of the analyses and thus, the continual dimension of the field of study (equivalent to 49 kmq) was disaggregated into a matrix x, y of cells with side 25 m (dimensional choice compared to the analytical and planning objectives), representative of the statistical units included in the result of the calculation of the different variables and, in the final phase, of the interdependence existing between the various components.

---

[5] Possible with the normal techniques of geoprocessing by means of any application of GIS.

[6] Comparison in Tomasi *et al*., 2013, Jones G.V., 2004, e Boyer J., 1998, the latter is available on `http://scholar.lib.vt.edu/theses/available/etd-9219802524/unrestricted/Boyer.pdf`

[7] *AddaWin* is the togetherness of multivariate analysis routines developed by Silvio Griguolo (Iuav, Venice) that is targeted for use in territorial applications. The package can be downloaded for free from the website: `http://cidoc.iuav.it/~addati/addati.html`

2.2 The Treatment of Quali-Quantitative Information and the Launch of the Analyses of Correspondence

The morphologic characteristics of the field of study were examined through the reproduction of the Digital Terrain Model, which elevates the terrestrial surfaces, permitting the calculation *i*) of the *solar exposition* (the level of orientation of the sides with respect to the cardinal points); *ii*) of the *slope of the surfaces* (expression of the % decline with respect to the plan); *iii*) of the *level of altitude* (m above sea level); the *typologies of the use of soil* were derived from the data bases, Corine Land Cover, reworked at the municipal scale of the Region of Emilia Romagna. The information relative to the *typology and nature of the lands* result from the regional pedologic map, which is available for the entire municipal territory at the scale 1:50000 with 15 'Units of soil,' including the river channel, from which was extracted the variable of the '*Capacity of soil usage*,' that characterizes every unit through the attributed limitations: *i*) to the climatic conditions; *ii*) to the risk of erosion and/or landslide occurrence; *iii*) to the presence of water stagnation in the workable soil; *iv*) to the scarce quality of the lands, enough to reduce the range of cultivation options.

In summary, the municipal space was characterized in a multidimensional manner, first identifying the new explicative synthetic factors (principal components) and, therefore, extracting the basins of the same phenomenal level through nonhierarchical cluster analysis[8] with eight variables defined differently. In particular, the three of the superficial morphology each include 6 modalities:

a) the exposure of the arguments include: *Esp_basso* = to the north; *Esp_medio_basso* = to the north and northeast; *Esp_medio* = to the west and southwest; *Esp_medio_alto* = to the east and southeast; *Esp_alto* = to the south;
b) the slope concerns: *Pend_basso* = from 0% to 2.25%; *Pend_medio_basso* = from 2.25% to 8.91%; *Pend_medio* = from 8.91% to 16.27%; *Pend_medio_alto* = from 16.27% to 30.29%; *Pend_alto* = from 22.09% to 30.29%;
c) the altimetry is subdivided into: *Alt_basso* = from 82 m s.l.m. to 107; *Alt_medio_basso* = from 107 to 132; *Alt_medio* = from 132 to 159; *Alt_medio_alto* = from 159 to 191; *Alt_alto* = from 191 to 255.

The use of soil sees to the clustering of the typologies into distinct groups in: *Uso_basso* = areas of extraction and landfills; *Uso_medio_basso* = mixed forests, chestnuts, poplar grove, orchards, areas of shrub and/or grass vegetation with sparse trees, embankments; *Uso_medio* = buildings; *Uso_medio_alto* = vineyards; *Uso_alto* = arable land, heterogeneous agricultural zones, uncultivated areas in the urban.

The pedologic units are subdivided into homogeneous groups, in function of the geologic matrix and of the temporal evolution: *Pedo_basso* = soils of the alluvial plain; *Pedo_medio_basso* = soils in morphologically significant areas removed from the alluvial plain; *Pedo_medio* = soils of well-conserved terraced surfaces; *Pedo_medio_alto* = soils of transition between the most recent gravel substratum gravel and properly hilly areas; *Pedo_alto* = lands composed of ancient clays incisive gravel substratum.

[8] Cluster is a discrete areal of basins of geo-phenomenal characteristic; therefore, it must present descriptive elements that characterize it in its complexity.

The units of soil were then classified according to their usage capacity, attributing their distinct limitations: *Cap_basso* = limitations for workability, fertility, hydrological excess, erosion, climatic interference; *Cap_medio_basso* = risk of flooding, superficial rockiness and scares profoundness useful to the roots; *Cap_medio* = risk of erosion, milits in the workability and climatic interference; *Cap_medio_alto* = limits to the workability; *Cap_alto* = limits for the scare profoundness suitable to the roots and climatic interference.

The three climatic macro areas are: *Aclim_basso* = area of plains with medium precipitation around 800 mm annually; *Aclim_medio* = area of piedmont plains and primary hills, with elevated maximum temperatures and medium precipitation around 820 mm annually; *Aclim_alto* = hilly areas with elevated maximum temperatures and medium precipitation around 840 mm annually.

The analysis of the built space was conducted by selecting the polygons of factories and related buildings, calculating their coefficient of form to identify the most jagged and, thereby, attracters of further expansive additions (Paolillo, 2005) and in the end deriving six classes representative of the level of instability of the perimetral morphology: *Cf_assente* = unbuilt areas; *Cf_basso* = values of the coefficient of form between 0.45 and 0.51; *Cf_medio_basso* = between 0.40 and 0.45 *Cf_medio* = between 0.35 and 0.40; *Cf_medio_alto* = between 0.29 and 0.35; *Cf_alto* = between 0.08 and 0.29, the last of which is expressive of the maximum settlement fragmentation on the municipal scale.

The software AddaWin allows for the contemporary treatment of all the modalities assigned to the single statistical units (the cells of the matrix), studying the relations with the variable in the space R_n and calculating the absolute values of the Eigenvalues (*eigenvalues*, inertia explained by each component), the proportion (that is to say the rate of variance explained by each component with respect to the total variance), the cumulative proportion of the previous relationships (as the sum of the Eigenvalues), in such a way so as to evaluate from how many principal components a particular range of variance can be explained. In rare applications, all of the principal *n* components are considered, choosing rather to select the number of principal components that is retained to be optimal in basal analysis: *i*) the criteria of parsimony (adopting the minimum possible entity of principal components); *ii*) to the minimum loss of information; *iii*) to the minimum deformation of the representative quality; to this end the % of explained variance (that is fulfilled by the three requirements mentioned above) hereby corresponds to the threshold value between 80% and 90%, considering in this way a value of 'noise' from 10 to 20% (usual in this type of analysis). Therefore, 18 principal components emerge from the matrix of the *eigenvalues* that can explain approximately 80% of the variance (total inertia) of the cloud of originating points, respecting the selection criteria placed into evidence.

In our case, 31 of the 39 total factors move onto the procedure of classification, making it necessary to choose then three factorial axes for the description of the statistical units and, consequently, three factors for the variables and three for the factorial planes.

2.3 The Place of the R_n Units: The Correlations between the Variables and the Principal Components

In order to interpret in a complete manner the obtained results, one should first examine the R_n relationships (spaces of the variables) between the original ones and the synthetic new ones (*CP*), each obtained from the linear combination of all the original variables, so as to establish the contribution of each one of these to the variability of each *CP* calculating, in other words, the weight achieved by the original categorized variables in the variance determined by the *CP*. The significance of the factors, considered as newly constructed variables, should be derived from their correlation with those starting variables on the basis of the greater absolute contribution on the various factors. Generally, to interpret the principal components obtained, one must examine the existing correlation between each original variable and each *CP* and, for this reason, the factorial weights *f* (*Factor Scores*) were calculated from the *Factor Pattern Matrix*, comprised of all the pairs of correlations of Bravais Galton between variables and *CP*[9]. A high characterization of the principal component on behalf of one variable can be discovered in the factorial weights (correlation coefficients) in terms of r (zj, yj) $\geq$ |0,8| against a medium – high $\geq$ |0,5|. In support of the interpretation of the principal components, from the matrix of the factorial weights it is also possible to construct – for all the pairs of principal axes (principal plains) – the plot of the original variables with respect to the principal generated plan, whose coordinates merely represent the coefficients of correlation variables *CP* contained in the matrix of factorial weights.

From the analyses of the *factor pattern matrix* and the first two principal planes emerge as *i*) the first factorial axis should be explained by the two negative variables, the use of the medium – high grounds of the vineyards (– 2.453) and the medium-high slope (– 2.432), with the result that, at the diminishing of *X*, the statistical units are not distinct neither from the vineyard nor from the inclined sides; *ii*) the second axis is connoted by the direct correlation with extractive uses (1.764) and inversely proportional to the soils of medium-high capacity (– 1.563); *iii*) in the third and last axis emerges the dual positive relationship with the most elevated slopes (4.652) and with the clay soils of the hillside (4.540).

It is still necessary to note how the axes are characterized by a limited number of correlations with the 39 assumed variables, a phenomenon which is indicative of the necessity to subsequently consider an elevated number of axes to understand the greater number of factors, otherwise reduced to the margins of the nonhierarchical analyses that identify homogeneous clusters. In this way, all of the 39 axes will be cross-examined in such a way to obtain a whole of classes that effectively represent the subsistent interactive complexity in the statistical sample.

[9] The factorial weight represents the Euclidean distance of the original variable on the principal component and, therefore, the greater the distance, the greater the significance.

2.4 Non-hierarchical Classification

This involves a classification that stretches to group units of similar behavior into a limited number of clusters, with a similarity that is directly observable or calculable from the sum of the variables that offer an opportune description of the analyzed objects. Through non-hierarchical methods an initial partition is somehow established that has the number of expected classes, and its quality is then improved through the opportune and further attributions of the units next to the boundaries between the classes, if such increases the value of the objective function, continuing the process of reallocation until reaching a final configuration that can not be further improved through small local shifts[10].

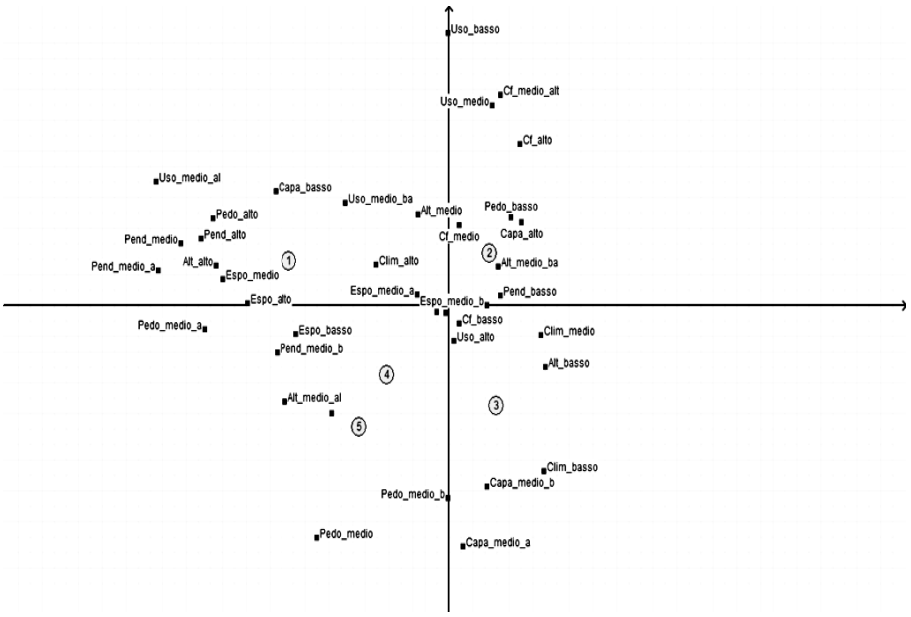


Fig. 1. Disposition of the statistical units and centers of class of the stable partition along factorial planes

In this case study, the individualization of the basins of similar level of viticultural behavior took place through a non-hierarchical analysis of the matrix of correspondence obtained previously, selecting 31 (basic) explorative partitions, among which the three best were then cross-analyzed to constitute the most compact groups that emerged from the analyses, the so-called stable classes, choosing ten under exploration since the number of those of the final division is effectively decided after the examination of the expressive graphic, from the diminishing of the objective function to the diminishing of the number of classes.

[10] The partition that is obtained constitutes in itself an optimal place, depending on the configuration initially assumed and on the number of required classes.

After the nonhierarchical classification, the portions of the territory (cells of the discrete matrix) were grouped into geospaces of similar behavior through a further simplification of complexity, determining, in this case, the cut of the curve in proximity of the sudden jump in slope, reaching a partition in 20 classes and arriving, through subsequent steps, at the aggregation of five stable profiles.

Cluster 1 marks 12.7% of the municipal territory, equal to 10.019 cells (remembering that each is 25 x 25 m), and it is characterized by hilly areas of elevated range, an articulation of diverse medium to high incline levels and well-exposed faces. With regard to the use of soil, arable soil, forests and vineyards, channels on pedologic horizons of terraced surfaces, and elevated agronomic limitations prevail. The built presents a medium level of perimeter fragmentation.

Cluster 2 constitutes 52.3% of the researched space (41.256 cells) mainly in foothills of low slope, soils at elevated agronomic capacity and a built surface of high perimeter fragmentation.

Cluster 3 accounts for 20.2% of the municipal surface (15.909 cells) in plain areas, with very low inclines, on soils of good agronomic capacity and a built surface of medium perimeter fragmentation.

Cluster 4 includes 10.4% of the municipal territory (8.184 cells), in hilly areas of medium/high range with medium slops on soils of good agronomic capacity (given the scarce limitations) and a medium perimeter fragmentation of the built surface.

Lastly, cluster 5 represents 4.5% of the researched space (3.516 cells) in hilly areas of medium-high range, with medium and elevated inclines and with the best exposure of the side, a use of soil dedicated primarily to forests, arable lands and vineyards, ancient pedologic substratum of gravel-clay patchwork and medium limitations to agro/cultivation conduct.

The result of the nonhierarchical analyses and the consequential affinity to the reference class was spatially reestablished in the GIS domain for each unit of the matrix of cells in which the entire municipal territory is disaggregated, as in figure 2.

3 The Results Reached: A Comparison between Different Modes of Calculation of the Degree of Viticultural Attitude

The multidimensional analysis conducted in the study was implemented to extract homogeneous basins of agronomic character in function of viticultural cultivation: the construction of a database of variables, that the agro-pedology accepts as factors of greater conditioning of viticultural cultivation, therefore represented the initial step in reaching the objective of zoning of agricultural space, as is the contribution of the morphologic/settlement variables will prove to be undoubtedly indispensible to the drafter of the municipal urban plans in order to correlate the agricultural qualities to the particular urban margins and hinges that, for the particular dispersive morphology of their perimeter, present incentives to urbanize for urban output reasons. In such a way, the urban/agriculture conflict for the use of soil will be contained as much as possible, at least in the spaces that tend toward the vocation of prestigious production and in the areas of landscape and environmental value.

Nevertheless, the application here presented distinguishes itself substantially from studies which are merely sectorial, like those on vineyard zoning: the role of the

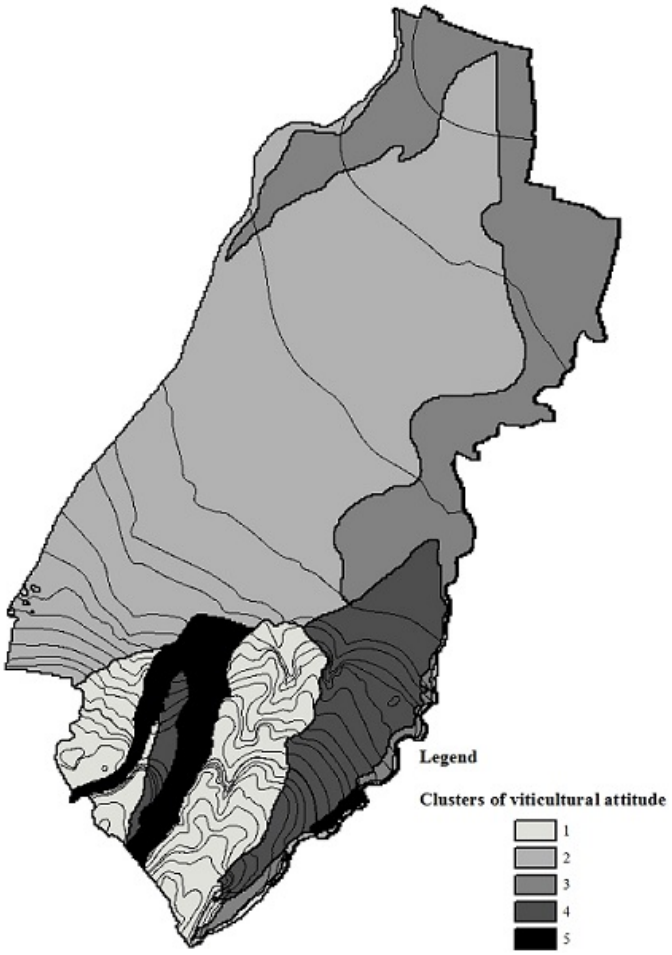


Fig. 2. Spatialization in GIS domain of the results of the nonhierarchical analyses: 1 = hilly ridges of elevated range, clay pedologic substratum with strong limitations, good exposures and medium slopes; 2 = plain areas, with soils of medium mixture and elevated productivity and absence of limitations; 3 = 'waterside' plain with soils of good agronomic capacity and scarce limitations; 4 = hilly ridges of medium range, with scarce slopes and soils of medium agronomic capacity; 5 = hilly slopes with strong inclines and prime solar exposure.

territorial planner should, in fact, collect and treat the *entire* cognitive foundation (Paolillo, 2010) in the construction of the urbanistic tool, not lingering solely on the configuration of urban, settlement, and infrastructural factors but also on those regarding agriculture, physical resources, the environmental asset. The complexity of the environmental asset, even once it has been divided into systemic components, must first be reduced (without however losing the significant calculations), thereby measured (with opportune quantitative modalities that in Italy are rarely practiced) and, in

the end, connected to the provisions of the government of the territory. Consequently, the ability to recognize the peculiar qualities of a space that appears to be anisotropic foresees the construction of a recognitive process that can be realized and is navigable without any allocation of weight and bias.

On the contrary, in Boyer (1998) and Jones (2004), every variable of calculation of the behavior of the winemaking territory is treated and weighted through the direct assignment (without detachment from the practice present in literature), with the result of a measure of the winemaking vocation calculated through the common law summation of bias to the variables, like the most recent Italian zoning practices (Tomasi *et al.*, 2013) that establish the quali/quantitative selection of grapes only in function of their agro-pedologic and climactic type, deriving the results always from hierarchical models that oversimplify (or do not even contemplate) the possibility of reading the interdependencies between variables.

The experiences hereby illustrated instead refuse to identify the viticultural potential through the overlay of single privileged factors. Contrarily, the multivariate statistical analysis that is not empirical but rather based on the multidimensional evaluation of the phenomena, aggregating into classes the statistical units (equivalent to the cells of the representative matrix of the municipal space) for the specific intensity and relational interdependence. In addition, it should be added as is the aggregation into classes of different viticultural behavior, obtained from the analyses, was proven by the verification of correspondence on the field, which has similar levels of intensity and correlation to factors proves an affective adherence to the real situation of the locales and offers significant guidance to the territorial planner.

References

1. Bellacicco, A., Labella, A.: La struttura matematica dei dati. La Goliardica, Rome (1979)
2. Bellacicco, A.: Metodologia e tecnica della classificazione matematica. La Goliardica, Rome (1975)
3. Bellacicco, A.: Una nota sul clustering condizionato. In: Ro, vol. 9 (1979)
4. Bellacicco, A.: Metodologie statistico matematiche per la identificazione di aree subregionali per le politiche dei servizi sociali. In: Palermo, P.C. (ed.) Modelli di Analisi Territoriale, Angeli, Milan (1983)
5. Benzecri, J.P.: L'analyse des donnée. Dunod, Paris (1973)
6. Bertier, P., Bouruche, J.M.: Analyse des données multidimensionnelles. Puf, Paris (1975)
7. Boyer, J.D.: Geographic analysis of viticulture Potential in Virginia. Thesis submitted to the Faculty of the Virginia Polytechnic Institute and State University, Master of science in geography (1998)
8. Caillez, F., Pages, J.P.: Introduction à l'analyse des données, S.m.a.s.h., Paris (1976)
9. Fano, P.L.: L'analisi fattoriale nell'identificazione del profilo socio-economico di una regione. Archivio di Studi Urbani e Regionali 9 (1978)
10. Fenelon, J.P.: Qu'est-ce que l'analyse des données? Lefonen, Paris (1981)
11. Fraire, M.: Metodi di analisi multidimensionale dei dati. Aspetti statistici e applicazioni informatiche. Cisu, Rome (1994)
12. Chadoule, G.: Metodi statistici nell'analisi territoriale. Clup, Milan (1983)
13. Jones, G.V., Snead, N., Nelson, P.: Geology and Wine 8. Modeling viticultural landscapes: A Gis analysis of the terroir potential in the Umpqua Valley of Oregon. Geoscience Canada 31(4) (2004)

14. Merlin, P.: Méthodes quantitatives et éspace urbain, Masson, Paris (1974)
15. Malczewski, J.: GIS-based land suitability analysis: A critical overview. In: Progress in Planning 62, pp. 3–65. Elsevier Ltd (2004)
16. Nijkamp, P.: Environmental economics. Wiley & Sons, London (1977)
17. Nijkamp, P.: Information system: a general introduction. In: Nijkamp, P., Rietveld, P. (eds.) Information System for Integrated Regional Planning. North Holland, Amsterdam (1984)
18. Palermo, P.C., Griguolo, S.: Nuovi problemi e nuovi metodi di analisi territoriale. Angeli, Milan (1984)
19. Paolillo, P.L.: Tipologie dell'assetto territoriale e consumo di suolo agricolo. In: Martellato, D., Sforzi, F. (eds.) Studi di Sistemi Urbani. Angeli, Milan (1990)
20. Paolillo, P.L. (con Prestamburgo, M.): Il sistema agro-forestale dello spazio regionale. Indirizzi territoriali in materia di zone agricole e forestali, Regione autonoma Friuli-Venezia Giulia, Trieste, collana Studi e ricerche per il Piano Territoriale Regionale Generale. In: Gli ambiti di densità rurale per il governo del sistema agro-forestale regionale, part. vol. III (1994)
21. Paolillo, P.L.: Un atlante cartografico per l'analisi multivariata delle risorse fisiche: la sponda orientale del Lago Maggiore. In: Il programma di Diana. Storia ambiente tradizione venatoria, alla ricerca di un modello condiviso. De Agostini, Novara (1995a)
22. Paolillo, P.L.: Procedure analitico-statistiche per la valutazione dei processi strutturali agricoli e per la loro localizzazione sul territorio comunale. In: Spazi agricoli a Cusago. Un esercizio analitico sul territorio extra-urbano: agricoltura, ambiente, paesaggio in un comune lombardo. Angeli, Milan (1995b)
23. Paolillo, P.L.: La sintesi del piano: le ragioni dello spazio rurale e il processo di riorganizzazione urbana. In: Aa, V. (ed.) Gli interessi agricoli nella pianificazione del territorio, Quaderni dell'Accademia dei Georgofili, Florence (1997)
24. Paolillo, P.L. (ed.): Al confine del nord-est. Materiali per il nuovo piano regolatore di Zoppola. Forum, Udine (1998)
25. Paolillo, P.L.: Una modalità descrittivo-classificatoria di individuazione dei bacini di intensità problematica ambientale alla scala regionale. In: Paolillo, P.L. (ed.) Terre Lombarde. Studi per un ecoprogramma in aree bergamasche e bresciane, Giuffrè, Milan (2000a)
26. Paolillo, P.L.: L'estrazione dei bacini di intensità problematica ambientale in Bergamasca e Bresciana. In: Terre Lombarde. Studi per un ecoprogramma in aree bergamasche e bresciane, Giuffrè, Milan (2000b)
27. Paolillo, P.L.: Acque suolo territorio. Esercizi di pianificazione sostenibile. Angeli, Milan (2003)
28. Paolillo, P.L.: Il contenimento della dispersione insediativa e l'uso degli indicatori nella valutazione ambientale strategica: un'applicazione in area vasta. In: Urbanistica, vol. 128, pp. 111–123 (2005)
29. Paolillo, P.L.: Sistemi informativi e costruzione del piano. Metodi e tecniche per il trattamento dei dati ambientali. Maggioli, Rimini (2010)
30. Paolillo, P.L.: L'urbanistica tecnica. Costruire il piano comunale, Maggioli, Rimini (2012)
31. Riveira, I., Crecente Maseda, R.: A review of rural land use planning model. In: Environment and Planning B: Planning and Design 2006, vol. 33, pp. 165–183 (2006)
32. Tomasi, D., et al.: Using Geospatial Technologies to Better Understand Terroir. In: The Power of the Terroir: the Case Study of Prosecco Wine. Springer, Basel (2013)

Proposal for Spatial Analysis in Web Mapping Applications. Computational Implementation

Romanuel Ramón Antunez[1], Lidisy Hernández Montero[2], and Keiver Hernandez Fernández[3]

[1] Instituto Superior para as Tecnologias de Informação e Comunicação, Luanda, Angola and Geoinformatics and Digital Signals Development Center, School 6, University of Informatics Sciences, Road to San Antonio de los Baños, Km 2 ½, Torrens, La Lisa, Havana, Cuba
{rramon,lhernandez,khfernandez}@uci.cu
[2] Geoinformatics and Digital Signals Development Center, School 6, University of Informatics Sciences, Road to San Antonio de los Baños, Km 2 ½, Torrens, La Lisa, Havana, Cuba
[3] Free Software Development Center, School 1, University of Informatics Sciences, Road to San Antonio de los Baños, Km 2 ½, Torrens, La Lisa, Havana, Cuba

Abstract. Based on the wide use of geographic information systems (GIS), and their impact on decision-making processes in organizations of several kind, the present study aims to develop an overall geospatial analysis web module for Web Mapping systems - based on international interoperability standards –this module constitutes the computational basis for the development of custom GIS applications to fit under the principle of technological sovereignty.The module was designed forenhancing the generality and reusability as a novel feature that allows the generation of thematic maps from statistics on default map databases. The study culminated in a proposal for the module to treat the main trends in algebra maps and hydrological analysis. It was also developed a generic algorithm independent of the coordinates system and/or the projection in which the data resides.

Keywords: Computational Geometry, Geographic Information System, Spatial Analysis, Thematic Map.

1 Introduction

The introduction of the increasing technological advances in communications, the development of powerful computing resources and in particular the release to the market of powerful software for storage, analysis and representation of geographic information, enable the modeling in a digital environment virtually any phenomenon capable of being referred spatially.

While there are several definitions of Geographic Information System (GIS), most agree that its main elements are the hardware, software, data and applications used by users to manage, analyze and deploy geographically referenced information.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 32–43, 2013.

Just as accessing other information via the Internet or the local network, it is also possible to use this one to access geographic information and work with it in a GIS. In the present context, it is not reliable to depend only on a SIG only with local data as files on the same computer on which you are working, but it is necessary to operate with remote data. Networking is the way to disseminate all kind of information including geographical information.

Web Mapping applications refers to the process of designing, implementing, generating and displaying or offering geospatial data through the World Wide Web (Mitchell, 2005). While a desktop SIG it is mainly intended for more specialized users, to provide to a Web browser viewing or editing capabilities of geographic information makes these systems to reach a different audience and opens new possibilities.

In recent years efforts in Cuba have been focused primarily on the development of GIS applications as well as in the implementation of the Spatial Data Infrastructure of the Republic of Cuba (IDERC, from its Spanish acronym), combining the efforts of several companies, research centers and universities in general. Despite the efforts and the progress made by various institutions in the development of this kind of applications, technologies associated with SIG systemswhich are used in the country come from different suppliers - mostly private companies -, have created a technological dependency unsustainable over time.

For these reasons, the design and implementation of solutions based on open source alternatives represents an ongoing challenge for the professionals of this fieldin Cuba, because of the impact of these technologies in the processes of decision-making in organizations whose field of action is of vital importance for our society.

Taking into consideration the defined elements are presented a set of questions that constitute the main motivation of this research:

1. How efficiently generate thematic maps to represent statistical data on a given cartographic base?
2. Is it possible to develop analysis on raster layers in these systems and minimize the effort to achieve it?
3. Can it be design a generic algorithm for geometrical analysis independent of the coordinates system and / or the projection in which the data resides?

Considering these questions, the present study aimed to develop a geospatial analysis module – based on international interoperability standards – for Web Mapping application, under the principle of technological sovereignty.

2 Materials and Methods or Computational Methodology

For the development of the solution a design based on a layer and component-based architecture - called plugin in this work - in a way to ensure the its scalability and reusability (Pressman, 2002), Fig. 1. Three plugins were definedfor its first version:

- Plugin for creating thematic maps.
- Plugin for analysis on raster layers.
- Plugin for geometric analysis.

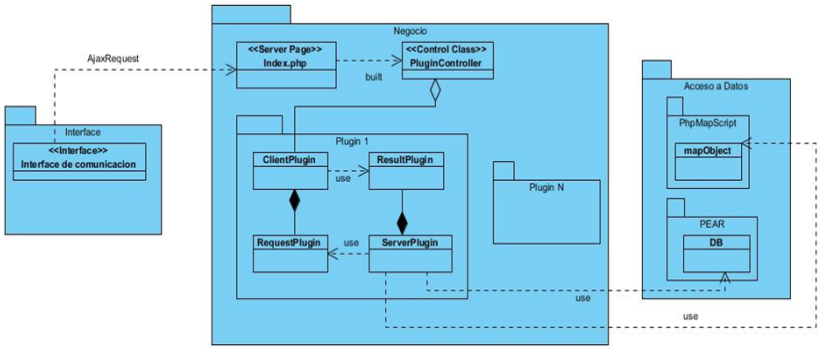


Fig. 1. Architecture of the Geospatial Analysis Module for Web Mapping Systems

3 Creation of Thematic Maps

The characteristics of thematic maps come from qualitative and quantitative data. In a qualitative map are shown the spatial distribution or the situation of a group of nominal data, one of its features is that in this kind of maps it is impossible to determine the relations of quantity. The data representing these maps are punctual, lineal and of surface (Aguirre et al, 1998):

The quantitative map is very used to represent all kind of themes such as social, economic and environmental benefits at a national, regional and local level. Different entities require to represent quantitative information geographically referenced, specialized and accurate.

Quantitative maps are the result of the application of quantum and experimental sciences, supported primarily on descriptive statistics. These maps generally represent punctual, lineal or area distribution data, even the combination of them.

The developed plugin is based mainly on working with expressions belonging to the classes of the maps layerswhich are being used. It allows the work with various formats such as ESRI Shape (.shp), MapInfo (.tab, .mif), GML and Postgres / Postgis, this latter being used as native to the general module.

Moreover it allows theming with data in these formats, enables the integration of spatial data stored in Postgres/Postgis with other alphanumeric data that are stored in another database related with the represented spatial information, being this the main innovation in a systems of this sort, since generally the existing solutions, mainly the free and open source applications only allow data theming on the ones stored next to the cartography and not in external sources (Ramon, et al 2011).

The figures 2 to 4 show some of the results obtained with the use of the developed plugin.

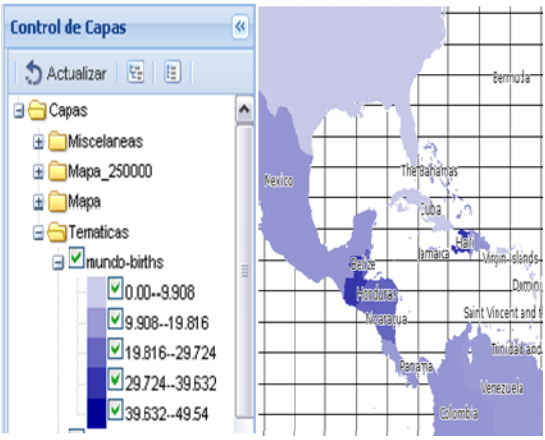


Fig. 2. Generated Chorochromatic Map Showing

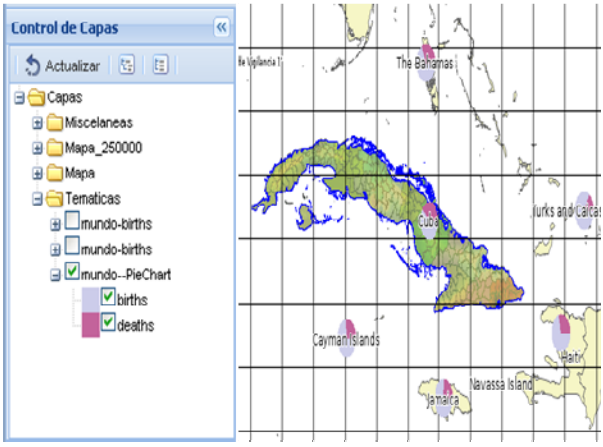


Fig. 3. Generated Map with Graphics

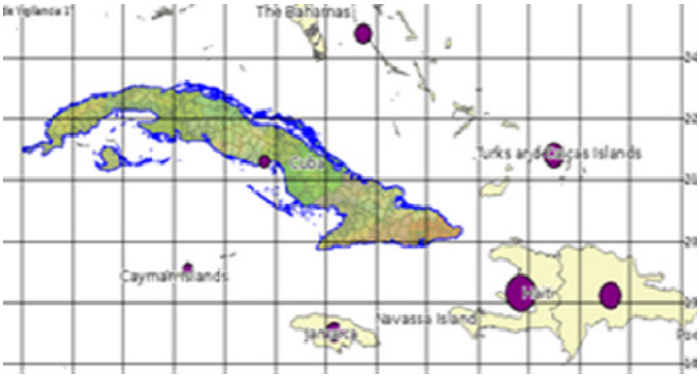


Fig. 4. Generated Proportional Symbols Map

4 Analysis on Raster Layers

For analysis on raster layers, the first version of the plugin emphasizes on the functionality for mapping algebra and hydrological analysis. To accomplish this a study is made of the major existing algorithms and models to solve these problems.

In (Hernandez, et al, 2012) are state the algorithms and models most referenced in the specialized literature to perform a comparison between them, according to complexity of development and problems which solve; throwing as a result the following proposal which was included in the first version of the module.

4.1 Solution for Algebra Mapping

The algebra mapping is an analysis tool that allows different operations between maps to gain new and different models from one or more maps as mentioned above. The amount of data used to perform operations, the application within the algebra mapping, the operations that apply to the calculation and the result, are critical parameters to achieve a comparison between the functions of algebra mapping (Tomlin, 1990) (Wood, 1996).

Once completed the comparison above, it was decided to present such proposal as a first version for Web Mapping applications to local and focal functions. The characteristics of these two functions allow the terrain analysis module to perform any transformation from one model to another using different operators established in the research, and the creation of a new model based on a layer of specific values for the construction of this one.

The local function has a great simplicity in the implementation of mathematical models and the operators used, allowing a rapid development and understanding by the module development team. Alongside the focal will be used for more robust analysis, it usually is a bit more complex than the local type but simpler than the other according to the input data. The input data can minimize or increase the processing time depending on the quantity required by the functionalities to be performed in the application

In the case of the local and focal function they use the least amount of possible input data, while the zonal depends on the classification criteria by classes and the generalas the name implies uses all data in the array of values. This feature gives Web Mapping systems the ability to accelerated response time to the customer or end user, since it is a web application and need to be as simple as possible to perform the calculations.

4.2 Algorithms for Hydrologic Analysis Solution

There are various algorithms and models that can be used for modeling and representing different hydrological parameters (Pilesjo, 1998) (Pilesjo et al, 1997) and for the calculation of topographic variables such as the curvature and slope, they are:

- Draw lines of flux.
- calculation flow using a flow matrix codes.

- Problem of the concavities.
- Delineation of watersheds.
- Alternate method based on routine flow striping.
- Cellular Automata.

The stated algorithms allow to generate, define and calculate three major features within the hydrologic analysis, flow lines, delineation of watersheds and maximum flow calculation. Allowing the modeling of these properties in a more simple and direct way (Belmonte et al, 2006) (Pizarro et al, 2005). Not all algorithms proposed are possible to use, so a comparison is made in terms of execution time and the problem it solves within the hydrologic analysis (Hernandez, et al, 2012).

The cellular automata by using states allows to go calculating the values taken by the hydrological variable when transiting different weather conditions. Obtaining an array of values which then allows to model and represent the different parameters of hydrology. This model calculates the three parameters in the same way, while other approaches focus on the calculation of a particular one. This shows that using the cellular automata model can be useful in time and effort when implementing this tool into the module.

5 Geometric Analysis

For the geometric analysis plugin of the module, a generic algorithm was designed which was independent of the coordinate system and/or projection in which the data resides. On which subsequently can be implemented a set of basic geometric algorithms, which help to build more complex analysis algorithms (Olaya, 2011) (Haining, 2003).

The proposed generic algorithm in (Ramón et al, 2012), parts of the property of bijectivity of the cartographic projections. So if working in projected coordinates is possible to obtain ellipsoidal coordinates for the reference ellipsoid of the beginning. Hence the idea presented is as follows:

- If you are working in a projected system then you can work with algorithms for planar coordinates or algorithms for ellipsoidal coordinates after a previous transformation of planar coordinates to ellipsoidal.
  - This decision may be taken by the user in case it is an expert, by default it would work converting ellipsoidal coordinates.
- In case it is not working on a projected system then it could work with algorithms for ellipsoidal coordinates.

This generic criteriafor geometric analysis inGIS was selected as a base to accordingly design the algorithms used to calculate distance, perimeter, azimuth and areas. Basically, wereproposed for the distance in the plane the classical Euclidean formulation, and in the case of spherical distances the Haversine formulation because they present more computational stability. Subsequently linked both proposals according to the presented generic variant (Ramón, et al, 2012).

```
functionCartesianDistanceCalculation (_pto1, _pto2: Point)
      return  sqrt(  pow(  (_pto2.x  -  pto1.x),  2  )  +
pow( (_pto2.y - pto1.y ), 2 ) )
end function

functionSphericalDistanceCalcutation(_pto1, _pto2: Point,
_radian: bool)
  _radio • 6 378 400
  _dlon • _pto2.x - _pto1.x
  _dlat • _pto2.y - _pto1.y
  _intcal•    pow(sin(_dlat/2),2)    +    cos(_pto1.y)    *
cos(_pto2.y) * pow(sin(_dlon/2),2)
  _intd • 2 * arcsin(min(1.0,sqrt(z)))

  if _radian do
      return_intd
  end if

  return_radio * _intd
end function

functionCalculateDistance (_pto1, _pto2: Point)
  if projection do
      ifcartesiando
            returnCartesianDistanceCalculation     (_pto1,
      _pto2)
      end if

      returnSphericalDistanceCalcutation(Transform(_pto1),
Transform(_pto2), true)
  end if
  returnSphericalDistanceCalculation(_pto1, _pto2, true)
end function
```

In the case of the area of polygons in the plane it is reported in the specialized literature a very efficient algorithm for solving it with complexity O(n), where n is the number of vertices. The algorithm uses the incremental technique, based on the calculation of the areas of the triangles formed by three vertices of the polygon starting from a fixed vertex (Chen, 1996). In the case of simple spherical polygons, the most widespread version is presented by (Bevis et al, 1987), which presents a constraint during digitizing of polygons and in case of being omitted it imposes a complexity of O (n log n) to the algorithm. For this reason, as one of the main contributions we designed a new algorithm that eliminated this restriction and maintained a complexity O(n) (Ramón et al, 2012).

```
functionLeftTurn( pto1, pto2, pto3: Point)
      return(pto1.x  *  pto2.y)  -  (pto1.y  *  pto2.x)  +
(pto2.x * pto3.y) - (pto2.y * pto3.x) + (pto1.x * pto3.y)
- (pto1.y *pto3.x)
end function
```

```
functionSurfaceCartesianPolygon (p: Polygon)
area • 0
fori • 0 to p.vertex – 2
      area • area + LeftTurn(p.vertex[0], p.vertex[i],
p.vertex[i+1])
end for

if area < 0 do
      return(area * -1) / 2
end if

returnarea / 2
end function

functionFcoseno (l1, l2, l3: double)
cose • (cos(l1) – cos(l2) * cos(l3)) / (sin(l2) *
sin(l3))

returnarccos(cose)
end function

functionSurfaceSphericalTriangle (pto1, pto2, pto3: Point)
radio • 6 378 100
l1 • SphericalDistanceCalcutation(pto1,pto2,false)
l2 • SphericalDistanceCalcutation(pto2,pto3,false)
l3 • SphericalDistanceCalcutation(pto1,pto3,false)

a1 • Fcoseno(l1,l2,l3)
a2 • Fcoseno(l2,l3,l1)
a3 • Fcoseno(l3,l1,l2)

epsilon • (a1 + a2 + a3) – Pi
area • epsilon * pow(radio, 2)

ifLeftTurn(pto1, pto2, pto3) < 0 do
      area • area * -1
end if

returnarea
end function

functionSurfaceSphericalPolygon (p: Polygon)
  area • 0
  fori • 0 to p.vertex – 2
      area•                    area                    +
SurfaceSphericalTriangle(p.vertex[0],p.vertex[i],p.vertex
[i+1])
end for
```

```
if area < 0 do
      returnarea * -1
end if

returnarea
end function

functionCalculateSurface(p: Polygon)
if projection do
      ifcartesiando
            returnSurfaceCartesianPolygon(p)
      end if

      returnSurfaceSphericalPolygon(Transform(p))
end if

returnSurfaceSphericalPolygon(p)
end function
```

The result was shown arriving to the following theorem:

Theorem 1. The area of a polygon may be computed, taking into account the coordinate system in which it is, on a time O(n), where n is its number of vertices.

To verify that the results in a real practice correspond to the theoretically obtained results, a set of tests were made to compare what difference there was among the results of the proposed algorithm for the calculation of spherical polygons area, and the results shown by the algorithms currently implemented in geographic information systems with open and close source code respectively, Fig. 5.

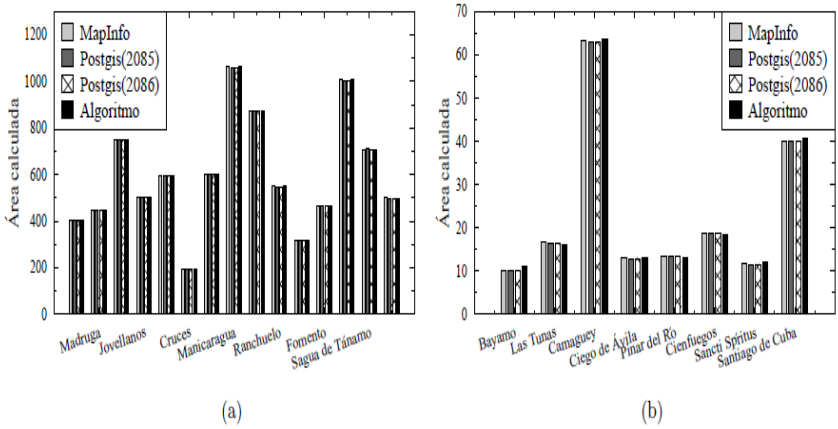


Fig. 5. Obtained results over scale mapping based on 1:500 000 (a) and 1:100 000 (b)

These tests were performed with geometries at different scales, as geographic information can be analyzed at different levels and depending on the level used, the results may be of different nature (Ramón, 2011) (Mandelbrot, 1983). It is verified that the results shown by the proposed algorithm are comparable to the classic variants, while reducing the response time to requests since the calculations are performed in O(n) in any case.

6 Results and Discussion

From the need to achieve technological sovereignty in the media representation on a digital map and the lack of an application with these requirements, it was decided to develop the GeneSIG platform, as a support for spatial data infrastructures of the Republic of Cuba and in defense areas. The presented module was incorporated to this system to enable it with the basic functions of a GIS, the system is developed in collaboration between the University of Informatics Sciences - GEySED Development Center - , the Business Group Geocuba and the MINFAR (Ministry of the Revolutionary Armed Forces) – XETID Center-; becoming in the technology package for the development of GIS of more use in such entities.

6.1 Scientific and Technological Benefits

- It is a package of the tools suite and technologies for developing GIS platforms, which is manifested in the number and variety of applications developed and registered in the Copyright National Center of Cuba (CENDA, from its acronyms in Spanish). It is also interoperable - by using OGC standards - with existing solutions, and able to integrate with the management software for decision making developed in the FAR(Revolutionary Armed Forces) and other entities in the country

- It was developed a GIS technology for the Web with a fully modular and scalable structure with minimum functional requirements and it is cross-platform which provides a set of functionality equivalent to the regularly found in traditional desktop GIS applications.

- Since its conception and during its development process there have been madea number of publications, Master thesis and scientific – technical exchanges, as evidenced in the references that are included in the article. All this has significantly contributed in the postgraduate training of the authors and the creation of GIS applications with high additional value, suitable for both domestic and exportation use.

6.2 Economic Benefits

- Elimination of costs associated with payment of software licenses.

- Reduced costs of production in the updating and management process of geospatial information.

- Increased efficiency in the production of geospatial information.

- Reduction of the scheduled time for the policies implementation of the computerization and migration to free software of data and applications.

7 Conclusions

The conception of the geospatial analysis module for Web Mapping systems, presented and its computational realization in the GeneSIGplatform, are an impact result which provides additional value in the field of innovation in such platforms.

As the most important topics and concluding it is remarked that:

- With the implementation of a plugin to generate thematic maps, it is ensured the creation of these from statistic data provided by the users and the selected geographic base. This, together with the techniques used for multidimensional analysis of data generated as a result of all transactions that take place in daily activity, is useful as a support for decision-making processes.

- The genetic algorithm designed to perform geometric analysis - specifically for calculating the area of a polygon - which is independent of the coordinate system and/or projection in which the data resides, ensures comparable results to traditional variants while reducing the response time to requests, because the calculations are performed in an O(n) order - where n is the number of vertices of the polygon in question - in any case.

The presented module, implemented with free tools and technologies, besides complying with OpenGIS specifications established by the OGC and consistent with the policies of migration to free software and technological sovereignty which drives our country, makes easy to respond efficiently and effectively to the needs of geospatial information, and the current requirements and perspectives of the national market, economic growth, the country's readiness to achieve military invulnerability, the integrated opening to markets in the area, in addition to ensuring the technological support and tools of the IDERC.

References

1. Aguirre, G.N., Ortíz, M.A., Nernand, O.: Principios Básicos de la Cartografía Temática. IGAC, Bogotá Colombia (1998) ISBN: 958-9067-32-8
2. Belmonte, S., Núñez, V.: Desarrollo de modelos hidrológicos con herramientas SIG. Instituto de Recursos Naturales y Ecodesarrollo. Facultad de Ciencias Naturales, Universidad Nacional de Salta (2006)
3. Bevis, M., Cambareri, G.: Computing the Area of a Spherical Polygon of Arbitrary Shape. Mathematical Geology 19(4) (1987)

4. Chen, J.: Computational Geometry - Methods and Applications. A&M University, Computer Science Departament, Texas (1996)
5. Haining, R.: Spatial Data Analysis: theory and practice. Cambridge University Press (2003) ISBN: 0-521-7743-73
6. Hernández, L., Ramón, R.: Propuesta De Algoritmos Para Análisis Hidrológico Yálgebra De Mapa En Sistemas Webmapping. Mapping Interactivo 142, 42–47 (2012) ISSN: 1131-9100
7. Mandelbrot, B.B.: The Fractal Geometry of Nature. Henry Holt and Company (1983) ISBN: 0716711869
8. Mitchell, T.: Web Mapping Illustrated: Using Open Source GIS Toolkits. O'Reilly, s.l (2005) ISBN: 9780596008659
9. Olaya, V.: Sistemas de Información Geográfica, p. 911. OSGEO:Your Open Source Compass, s.l (2011)
10. Pilesjo, P.: Estimation of flow distribution for hydrological modelling. In: Proceeding Geoinformatics 98 Conference (1998) [En línea], http://www.natgeo.lu.se/Personal/Lars.Harrie/GeoInf98.pdf
11. Pilesjo, P., Zhou, Q.: Theoretical estimation of flow accumulation from a grid-based digital elevation model. In: Proceeding Geoinformatics 97 Conference (1997), http://geog.hkbu.edu.hk/QZone/Research/Papers/FullPaper/1997_GInf_3.pdf
12. Pizarro, R., Soto, M., Farías, C., Jordan, C.: Aplicación de dos Modelos de Simulación Integral Hidrológica, para la estimación de caudales medios mensuales, en dos cuencas de Chile central. Bosque 26(2), 123–129 (2012), http://eias.utalca.cl/2publicaciones/paper/pizarro_soto_farias.pdf (Marzo 17, 2012)
13. Pressman, R.S.: Ingeniería de Software. In: Un Enfoque Práctico. McGraw-Hill, Madrid (1998) ISBN: 84-481-1186-9
14. Ramón, R.: Propuesta De Algoritmos Para Análisis Geométricos En Sig. Master'sthesis, Departamento de Geoinformática, Facultad 6, Universidad de las Ciencias Informáticas (Diciembre 2011)
15. Ramón, R., Hernández, L.: Análisis geométricos en SIG basada en un algoritmo genérico independiente del sistema de referencia. Revista Cubana de Ciencias Informáticas 5(3) (2012) ISSN: 1994-1536
16. Ramón, R., Hernández, L.: Sistema para la creación de mapas temáticos. In: Memorias de Informática 2011. s.n., La Habana (2011) ISBN: 978-959-7213-01-7
17. Tomlin, C.D.: Geographic information systems and cartographic modelling. Prentice Hall (1990) ISBN: 0133509273
18. Wood, J.: The Geomorphological Characterization of Digital Elevation Models. Ph.D. thesis, University of Leicester (1996)

Water Quality Models Sharing in Three Gorges Reservoir

Jingwei Shen, Limin Guo, Tinggang Zhou, Guowei Li, and Tiying Su

School of Geographical Sciences, Southwest University
400715 Chongqing, China
{jingweigis,guo_limin.bobo}@163.com,
{zhoutg,oolili}@swu.edu.cn,
237895772@qq.com

Abstract. To better reuse water quality models, it is necessary to implement the sharing of water quality models in Three Gorges Reservoir. The studying area, Three Gorges Reservoir, is illustrated. Water quality models are classified according to the type of pollutants. Network service architecture including water quality model service providers, service requesters and service broker are designed. The relationships of the different parts of network service architecture are set forth. The conceptual data model is designed to describe the water quality models. Web service technology is introduced to encapsulate water quality models. Web services are deployed on the server and web services are scheduled by the clients. Experiments are conducted and we can see the results executed by the different web services. At last, we can conclude that web service can effectively implement the water quality model sharing and promote the reuse of the water quality models.

Keywords: Three Gorges Reservoir, Water Quality Models, Encapsulated, Sharing, Web Service.

1 Introduction

Three Gorges Dam project was completed and fully functional as of July 4, 2012. After completion of the Three Gorges Dam, the main stream of up to 660 km long river-type reservoir was formed. The water flow rate significantly reduced and the self-purification capacity of the water decreased. Recent years, water accidents occurred frequently in Three Gorges Reservoir and significant ecological problems were caused.

Water quality models using mathematical simulation techniques were developed for predicting spatiotemporal distribution of the water pollution. Large water quality models were built in recent years. Water quality model to simulate pollution belt with dynamic features was built [1]. Dynamic behavior of the oil pollution was described by hydrodynamic model [2]. An advection diffusion model was developed to simulate the spread of chemical substances [3]. Pollutant dispersal using numerical simulation was displayed in the marsh [4]. A water quality model was established to simulate the

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 44–52, 2013.

physical, chemical and biological conversion processes of the Bohai Bay [5]. Three-dimensional numerical model was established to simulate the unstable transmission of COD [6]. Mathematical models and formulas, including large number of hydrology models, published in "Journal of Geographical Science" from 1934 to 1999 were analyzed [7]. "Resource Environment Mathematical Model Handbook" was published and large numbers of geographic models were collected [8]. There was also some water quality models proposed for Three Gorges Reservoir. Two-dimensional random forecasting model based on the finite element method was proposed and the probability distribution of water quality was calculated using Monte Carlo method in Three Gorges Reservoir [9]. One-dimensional and two-dimensional water quality mathematical models which can simulate the migration of pollutants diffusion process for Three Gorges Reservoir were studied [10]. Sudden water pollution accident model and the movement of pollutants in the water were studied [11]. The horizontal and vertical distribution and characteristics were simulated [12]. The above researches promoted the progress of water environment. However, most of models which were expressed in written form can't be directly used for computer-level simulation and computing.

To achieve the sharing of models, some related studies were conducted. A spatial decision support modeling framework based on agent was proposed to improve the access capacity of the distribution model [13]. A dynamic architecture for distributed geographical service was presented to solve geographic service sharing problem in the distributed, heterogeneous network environments [14]. A geospatial data processing based on SOA for an open network of services was proposed [15]. A sharing and integration framework for eco-oriented models was proposed [16]. An XML model was put forward and the web-based decision support system model was explored [17]. A service-oriented distributed geographical model integration framework was put forward and the sharing of heterogeneous distributed model was discussed [18]. Geographical models sharing methods based on web service were proposed [19]. The distributed shared services framework for geospatial model was proposed referring to geographic information services classification of OGC [20]. The above researches can provide some experience for water quality models sharing in Three Gorges Reservoir. However, the detail description and the application of the models were not described.

In this paper, water quality modeling sharing in Three Gorges Reservoir will be studied. The remainder of the paper is organized as follows: in Section 2, Three Gorges Reservoir area will be described and water quality models for the research will be introduced. Section 3 will provide methodology on how to implement the sharing of the models. In Section 4, some experiments will be conducted and the results will be presented. Lastly, Section 5 will describe the paper's conclusions and point to future work.

2 Study Area and Water Quality Models

Three Gorges Reservoir Region (Fig. 1), is the area directly or indirectly involved in the submersion of the water storage of the reservoir area of the Three Gorges Project.

The Three Gorges Reservoir Region stretches along the Yangtze River from Jiangjin District of Chongqing to Yichang City of Hubei, which is very narrow and where the geography is complex [21]. Fig. 1 shows our study area. Our research area covers the whole Three Gorges Reservoir Region.

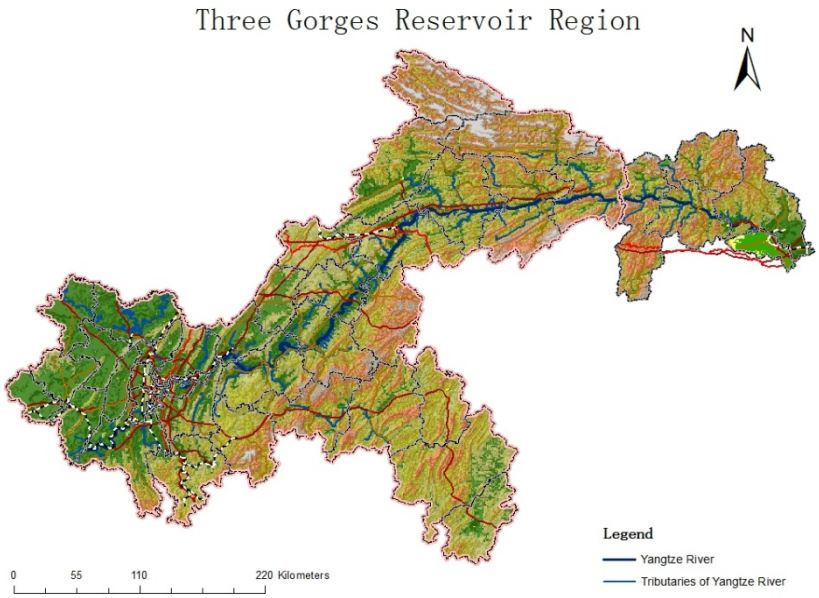


Fig. 1. Three Gorges Reservoir Region

Large amounts of water quality models in the Three Gorges Reservoir Region have been accumulated. There are many classification methods. According to the type of the pollutants, pollutants type can be divided into nine classes including persistent pollutants (such as heavy metals, toxic organic), non-persistent pollutants (general organic pollution), acid pollution (PH), thermal pollution, suspended solids, plant nutrients, the radioactive material, oil, pathogens [22]. Therefore, we can classify the models according to the pollutant type.

3 Methods

3.1 Network Service Architecture of Models Sharing

In this paper, network service architecture of models sharing, including water quality model service providers, water quality model service requesters and water quality model service broker, are designed. The network service architecture of water quality model sharing can be described in Fig. 2.

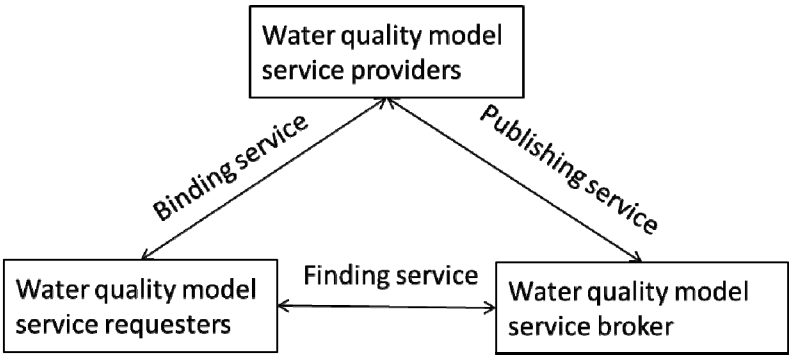


Fig. 2. Network Service Architecture of Water Quality Model Sharing

Service providers which define and implement water quality model publish the service description to the service broker. Model service requesters search service from service broker and bind web services with service providers. Service broker is the link of service provider and service requester. Both service providers publishing their services and service requestor searching services require to the help of service broker. In this paper, server plays the roles of service providers and service broker and client plays the role of the service requesters. Both indexing service and publishing service should be implemented by the server. The mentioned network service architecture can effectively implement the sharing of the water quality models.

3.2 Water Quality Models Description

To describe the water quality models, a conceptual data model is designed. The application conditions, mathematical formula and description of each parameter are considered. Therefore, the conceptual data model described by ER diagram is follows.

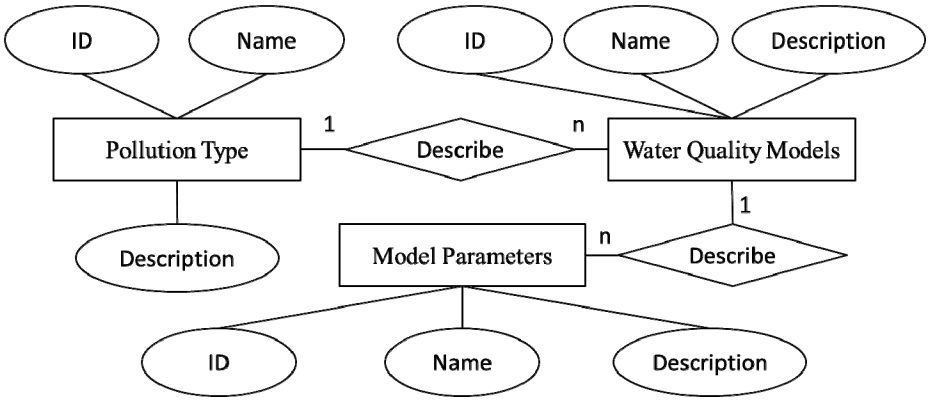


Fig. 3. Conceptual Data Model of Water Quality Models

In Fig. 3, pollution type entity, water quality models entity, model parameters entity, and their relationships are designed. Pollution type entity, water quality models entity and model parameters entity are the abstraction of the pollution type, water quality model and model parameters, separately. Each pollution type may relate to various models in different geographic condition. Each model may include many of parameters. Therefore, model parameters entity is designed to describe the water quality model. On the basis of the conceptual data model, the logical data model and physical data model are also designed to meet the needs of the model description. The client can index the model from the description information of the models. According to the description information, the client decides whether schedule this model or not.

3.3 Web Service Encapsulating of Water Quality Models

A web service is a software system designed to support interoperable machine-to-machine interaction over a network. With the characteristics of good encapsulation, it can shield heterogeneity and implement the loosely interconnection of the models. Therefore, this project intends to adopt web service technology to implement the service-oriented encapsulation of the Three Gorges Reservoir water quality model.

The process of web service encapsulation is follows (Fig. 4). Firstly, we should analyze the water quality model and acquire the input and output variables which are the input and output parameters of the web service. The model is abstracted into source code. In the paper, C# is taken as the source code language. Secondly, the source code is encapsulated into web service. Consistent with the source code, web service is also built in C# on the .NET platform. Thirdly, web service is executed and the input and output parameters are described. At the same time, the corresponding description information of the web services is stored into relation database. By these steps, models are encapsulated into executed web services.

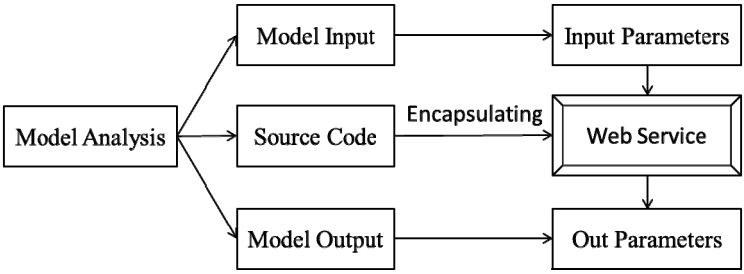


Fig. 4. The Process of Web Service Encapsulating

4 Results

4.1 Water Quality Models Encapsulating and Publishing

In the experiment, some water quality models are encapsulated by C# programing language. After the encapsulating operation, we can publish the water quality models (Fig. 5). From Fig. 5, we can see the some typical models are encapsulated and published.

Fig. 5. Web Service Encapsulation of Water Quality Models

After the web services are deployed on the server, the clients can search the service and schedule the models at any time and place by the internet. By this way, we can implement the sharing of the water quality models. This approach avoiding the duplication of work can greatly reduce workload.

4.2 The Simulation of Pollutants Movement

To demonstrating the simulation results of pollutants movement, an experiment is conducted. The scheduling process is follows. Firstly, the clients index the water quality models according to the description information of the web services. Secondly, the clients enter the parameters that are needed by the models and submit the parameters to server. Thirdly, the results are calculated by the web service in the server and the server sends the results to the client. Lastly, the execution results in the server are returned to the clients. In this paper, the simulation results are displayed in the graphical form. By the above process, we can see the results by different water quality models. In this experiment, two models including persistent pollutants model and non-persistent pollutants model are scheduled and the results are illustrated. Persistent pollutants are resistant to environmental degradation through chemical, biological, and photolytic processes. Therefore, they have been observed to persist in the environment, to be capable of long-range transport. In the contrary, non-persistent pollutant is decomposed or degraded by natural biological communities and removed from the environment relatively quickly. We can get the following results (Fig. 6 and Fig. 7) executed by persistent pollutant model and non- persistent pollutant model.

Fig. 6. The Simulation Result of Persistent Pollutant Model

Fig. 7. The Simulation Result of Non-Persistent Pollutant Model

From Fig. 6, we can see that the density of persistent pollutants do not change with distance and time changing. However, from Fig. 7, we can observe that the density of non-persistent contaminants reduces with the changing of distance and time. The density difference is expressed by different colors. So, according to the different conditions, we can select a suitable model to simulate the pollutant movement.

5 Conclusions and Future Work

Water quality models sharing in Three Gorges Reservoir is presented in this paper. Network service architecture of models sharing is described and water quality model service providers, service requesters, service broker and their relationships are designed. The conceptual data model is designed to represent the water quality model.

Web service technology is introduced to encapsulate water quality models. Based on the proposed methods, some experiments are conducted and the results are illustrated. We can conclude that the proposed methods can implement the sharing of the water quality models.

Clearly, what we have explained in this article can be regarded as an initial step for the sharing of the water quality models. The following are the research issues that should be considered in our further study:

1) The construction of sharing of water quality models is a long time process. Therefore, we will continue to encapsulate water quality models in the Three Gorges Reservoir region and provide the web service for the pollution simulation.

2) Water quality models are very complex and the conceptual data model is very simple. A more comprehensive conceptual model will be designed in future studies.

3) With the development of distributed computing, such as parallel computing and cloud computing, we will consider running the complex models on multiple computers.

Acknowledgments. We also thank Yikang Rui from Royal Institute of Technology for his help to polish the English. This research is jointly supported by the State-Sponsored Scholarship Program, Fundamental Research Funds for the Central Universities (No. XDJK2013C035), and Doctor Foundation Project of Southwest University (No. SWU111062).

References

1. Giri, B.S., Karimi, I.A., Ray, M.B.: Modeling and Monte Carlo Simulation of TCDD Transport in a River. Water Resource 35(5), 1263–1279 (2001)
2. Cekirge, H.M., Sollohub, J., Navon, I.M., Nnaji, S.: Models of Oil Spills in Florida Waters. In: Third International Conference on Development and Application of Computer Techniques to Environmental Studies, Montreal, Canada, pp. 357–362 (1990)
3. Kachiashvili, K., Gordeziani, D., Lazarov, R., Melikdzhanianc, D.: Modeling and Simulation of Pollutants Transport in Rivers. Applied Mathematical Modeling 31(7), 1371–1396 (2007)
4. Meyer, J.C., Diniz, G.L.: Pollution Dispersion in Wetland Systems: Mathematical Modeling and Numerical Simulation. Ecological Modeling 200(8), 360–370 (2007)
5. Yuan, D., Lin, B., Falconer, R.A.: Development of an Integrated Model for Assessing the Impact of Diffuse and Point Source Pollutant on Coast Al Waters. Environmental Modeling & Software 10(5), 1–9 (2006)
6. Chau, K.W., Jiang, Y.W.: Simulation of Trans Boundary Pollutant Transport Action in the Pearl River Delta. Chemosphere 52(3), 1615–1621 (2003)
7. Liu, C.M., Yue, T.X., Zhou, C.H.: Geography Mathematical Models and Application, the Compilation of Mathematical Models and Formulas in “Geographical Journal” from 1934-1999. Science Press, Beijing (2000)
8. Yue, T.X.: Resource Environment Mathematical Model Handbook. Science Press, Beijing (2003)

9. Xu, Q.G.: Research on Water Quality Prediction and Water Pollution Control Measures of the Three Georges Reservoir. Master Thesis, Beijing, Chinese Academy of Environmental Sciences (2004)
10. Zhai, J.: Study on Turbulence Diffusion Simulation and GIS-based Water Pollution Control Management System in the Three Gorges. Doctoral Thesis, Chongqing, Chongqing University (2008)
11. Si, H., Bi, H.P.: Numerical Analysis on Pollutant Decline in the Emergency of Water Pollution in Three Gorges. Environmental Science 29(2), 2432–2436 (2008)
12. Bi, H.P.: Study on Numerical Simulation and Risk Assessment of Water Pollution Scenarios in Three Gorges Reservoir. Doctoral Thesis, Chongqing, Chongqing University (2008)
13. Sengupta, R.R., Bennett, D.A.: Agent-based Modeling Environment for Spatial Decision Support. International Journal of Geographical Information Science 17(2), 157–180 (2003)
14. Tsou, M.H., Buttenfield, B.P.: A Dynamic Architecture for Distributed Geographic Information Services. Transactions in GIS 6(4), 355–381 (2002)
15. Díaz, L., Granell, C., Gould, M., Olaya, V.: An Open Service Network for Geospatial Data Processing. In: 2008 Free and Open Source Software for Geospatial (FOSS4G) Conference, Cape Town, South Africa, pp. 410–420 (2008)
16. Feng, M., Liu, S.G., Euliss, N.H., Younge, C., Mushetd, D.M.: Prototyping an Online Wetland Ecosystem Services Model Using Open Model Sharing Standards. Environmental Modeling & Software (26), 458–468 (2011)
17. Li, X.F., Liu, L., Wang, X.Y.: Research on Sharing Organism of DSS Decision Model Based on XML. Computer Integrated Manufacturing Systems 10(8), 903–907 (2004)
18. Wen, Y.N., Lu, G.N., Yang, H., Cao, D., Chen, M.: A Service-Oriented Framework of Distributed Geographic Model Integration. Journal of Remote Sensing 10(2), 160–168 (2006)
19. Yang, H., Sheng, Y.H., Wen, Y.N., Hu, Y.: Distributed Geographic Models Sharing Method Based on Web Services. Geomatics and Information Science of Wuhan University 34(2), 142–145 (2009)
20. Feng, M., Liu, S.G., Euliss, N.H., Yin, F.: Distributed geospatial model sharing based on open interoperability standards. Journal of Remote Sensing 13(6), 1067–1073 (2009)
21. Wikipedia, Three Gorges Reservoir Region (2013), `http://en.wikipedia.org/wiki/Three_Gorges_Reservoir_Region` (accessed on June 24, 2013)
22. Zuo, Y.H.: Environmental Studies. Higher Education Press, Beijing (2010)

Multivariate Applications in the Evaluation of the Discipline of Agriculture: Extra-Urban Spaces and the Resistivity Index

Pier Luigi Paolillo*, Massimo Rossati, and Mattia Andrea Rudini

Abstract. Agronomic studies represent a supplemental but inevitable element of the municipal urban plan, which must guarantee agricultural lands their ranking as a non-negotiable resource, thereby protecting its own maintenance for the irreversibility of its eventual transformation. Furthermore, in recent years in Italy, the exasperated conflict regarding extra-urban space, between building expansion and the practice of agriculture, combined with a growing environmental sensitivity, has shifted the epicenter of political and scientific attention to the unacceptable waste of soil that has occurred in recent decades. As a result, awareness has increased with regard to the finiteness of such a non-renewable resource and the implications of its safeguarding: that it results in the protection of important historical roots, the strong character of the landscape, and the tremendous quality of agricultural production in Italy. However, the attempt to protect agricultural grounds, within the municipal urban plan, is perhaps in vain, illusory, and unrealistic without the knowledge of the multiple variables that oversee the extra-urban structure. The extra-urban structure is viewed from the perspective of a protection that must rest – beyond the exonerative popular environmental ideologies – on the estimate of the agro-productive field and on the individual business capacity to resist the urbanization solicitations generating profit, investments, environmental production and qualified produce, multifunctionality, as new European agrarian politics intend. To these results we can attribute the careful use of Geographic Information Systems, allowing for the identification of the agricultural sector in extra-urban spaces and, at the same time, highlighting the areas that can resist pressure regarding land output better than others. The versatility of GIS tools allows for not only the (almost too banal) utilization of topologic overlay techniques but also – and most importantly – the evaluation of a synthetic indicator of resistivity of the agricultural enterprise, by means of geostatistical applications of multivariate analysis. In this way, it is possible to guarantee that urban decisions are sufficient in the conservation of the evermore scarce soil resource, while simultaneously safeguarding the country of Italy, its extremely evolved agricultural enterprise system, its unique landscape, and the services that it can offer in great variety, quantity, and quality.

Keywords: waste of agricultural soil, agricultural resistivity, geographical information systems, multivariate geostatistical analysis.

* Politecnico di Milano, Dipartimento di architettura e studi urbani, via Bonardi 3, 20133 Milan (Italy), pierluigi.paolillo@polimi.it, http://paolillo.professor.polimi.it/

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 53–65, 2013.

1 The Importance of the Topic of 'Waste of Soil' in the Conflict between Urban and Rural Uses

All too frequently, agricultural space has been intended exclusively for use as a 'tank' from which to draw installable ground for needs of the amebiform urban expansion that are evermore suffocating and uncontrolled. These needs have materialized in recent decades in Italy and have led to the consumption – and eventual waste – of optimal agro-productive soil and of spaces of notable environmental regeneration (Paolillo, 2002). Recognizing the importance of agriculture today signifies making informed decisions leading not only to the minimization soil wastage but also to the appreciation of the resource's multifunctional role in the direction of new European agricultural politics. Equally as important is its role in the appraisal of Italy, a country characterized by its truly evolutionary agricultural business and unparalleled landscape, and the services that Italian agriculture can offer in great variety, quantity, and quality. The case study hereby presented is that of Martinengo, a municipality of 10,000 inhabitants in the lowlands of Lombardy with profound historical roots and extensive agrarian production. The agronomical analysis conducted identifies the evaluative factors in terms of business resistivity, keeping in mind that the capacity of agriculture to compete with urban income for the 'soil resource' will be evermore connected to territorial appeal and to fixed environmental capital, to be managed virtuously and unitarily for the production of 'typical' agricultural produce, 'rare' community services, 'indirect' goods such as environmental quality, monitoring of the landscape, conservation, enrichment, and the transmission of traditional values and cultural peculiarities (Paolillo, 2003).

2 The Basis of the Research: The Survey of Agricultural Enterprises

The assumption of an investigation based on quantitative methods resides on the construction of an appropriate database. The database must allow for the development of analytical foundations, from which the complex characteristics of the studied space emerge. Consequently, the preliminary step involved the analysis of the existing databases to verify the thoroughness, level of update, quality, and usability of the data within Territorial Information Systems. From this exploration emerged the need to enrich the existing agronomical database through field research, geared toward understanding the agricultural businesses operating in Martinengo.

2.1 The Survey of Existing Agricultural Activities

The Excel model constructed to survey agricultural activities was structured in such a way to consent a detailed evaluation of business data, disaggregated into different categories corresponding to: *i*) general notes on the enterprise (denomination, tenant, form of jurisdiction, management system, measurements of support of rural development, information tools, surface, cadastral particles); *ii*) land use 2011 – 2012 (type of cultivation, presence of quality agriculture); *iii*) irrigation (surfaces included, cultivation method, supply source); *iv*) machinery; *v*) livestock (cattle, management

methods, storage of wastewater, application of wastewater, location of the breeding grounds); *vi*) labor and connected activities (categories of workers, equipment wholesaler, company-related activities, renewable energy implants, production of animal feed, commercialization of company products). In reference to the units of analysis of the cadastral particles, the data collected were opportunely related to and thus georeferenced for the subsequent production of the descriptive map and the appropriate GIS analyses.

Fig. 1. Classes of vastness of agricultural enterprises in the municipality of Martinengo (Italy)

2.2 The description of the Indicators in Concordance with the Judgment of Agricultural Resistivity

Following the enhancement of the existing database, a set of adoptable indicators was identified to calculate the agricultural resistivity of extra-urban lands. From the bibliographic reference, indicators were selected that are effectively calculable in the objective case. These indicators were aggregated so as to identify three macro themes on which to base the interpretation of the statistical sample.

Table 1. List of indicators adopted for the evaluation of business resistivity

Macro area	*Indicator*
	Company stability
Socio-occupational aspects	Effective company employment
	Typology of agricultural enterprise
	Unit of economic dimension
Economic aspects	Indicator of economic return
	Endowment of exploitable living capital
	Proprietary incidence
Morphologic aspects	Productive potential
	Company compactness

3 Topologic Overlay and Multivariate Analysis: Two Modes of Comparative Evaluation

3.1 The Limits of Topologic Overlay in Its Application to Territorial Phenomena

The topologic overlay represents a procedure saturated with spatial analyses that permits, by the simple means of layer overlap, the construction of new themes characterized by the presence/absence of a determined attribute. However, this distinguishes it tremendously from multivariate analyses that, instead, are able to calculate not only the contemporary presence of the multiple characteristics of the variables in question but also their latent characteristics and corresponding ties and interdependencies. The ability of multivariate analyses to calculate these characteristics leads to the rise of spatial basins of homogeneous intensity levels of the links between the dependent variable and the independent variables that define it (to recognize the limits and potentials of the two applications, the comparison of respective potentials and limits is demonstrated in the following table).

Table 2. A comparison between topologic overlay and (nonhierarchical) multivariate analysis

	Advantages	*Disadvantages*
Topologic overlay	Limited methodological capacity Validity limited to preliminary research	Restrained capacity to interpolate between the data (limited to the Σ of points or of present phenomena)
Multivariate analysis	Elevated level of detail of the obtainable results Reduction of the complexity of the model	Complex method that can be developed upon the provision of very advanced statistical applications

Therefore, multivariate analysis, a method of treatment of data that is undoubtedly more lucrative than the superimposition of information layers, which is the method usually adopted in urbanistic/environmental elaborations, allows for the identification of more articulated geospatial representations. As will be demonstrated in the following sections of this paper, the analyses of the typologies and the correspondences and nonhierarchical classification represent the steps (undergone with appropriate software) for the progressive reduction of complexity in the multivariate treatment of the different variables considered. However, this reduction of complexity is not synonymous with the reduction of precision, which is actually maintained sufficiently elevated in the identification of the descriptive classes.

The employment of quantitative measures for the understanding of phenomena and contexts of evermore complexity, and for the coinciding of consequential necessities of contemporary urban studies, is rooted in my most recent scientific works. These works present the indispensability of the adoption of quantitative measures and explain that other operative models cannot replace these measures. The applications of multivariate analysis that will be examined thereby represent a tool that – coupled

with Territorial Information Systems – guarantees parsimonious and prudent objective models for the environmentally sustainable allocation of destined soil usage.

3.2 The Determining of Interdependencies through Multivariate Analysis: Calculating the Classes of Settlement Interference

The multivariate method is preferable to topologic overlay because of the variety of its results and lowering of the level of uncertainty. The associated analytical/evaluative process is reported in the following chart, which shows the steps from the identification of the statistical units of analysis to the spatialization of similar basins on the basis of unit interdependence.

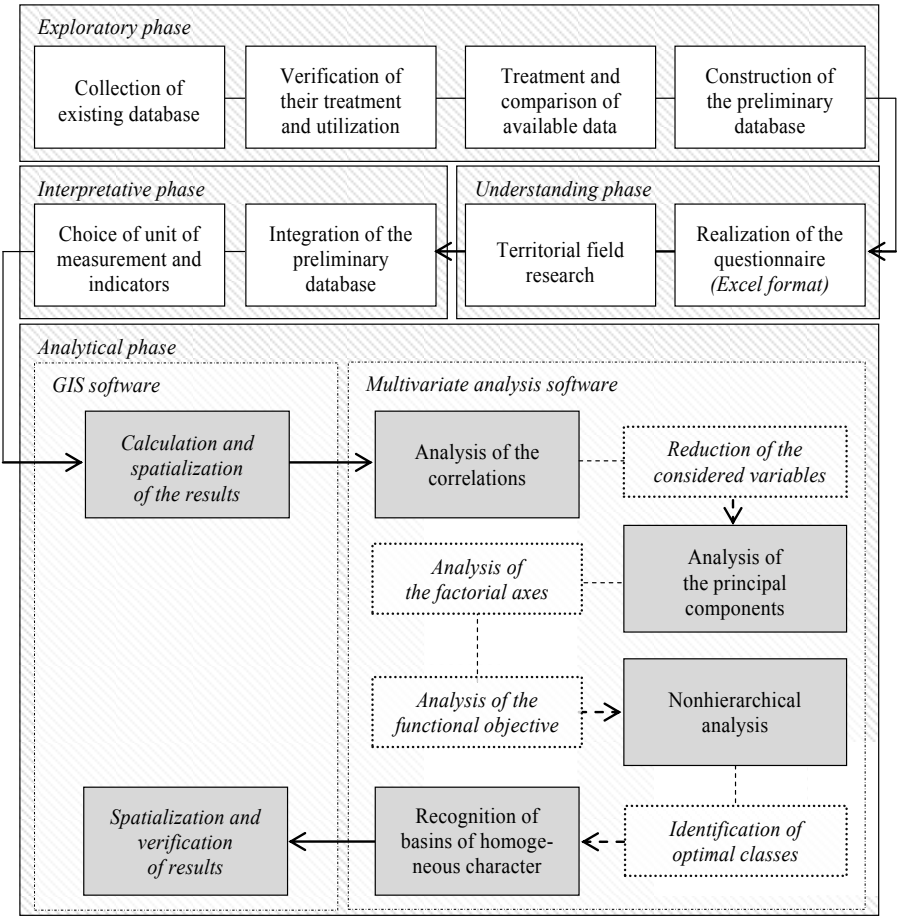


Fig. 2. Diagram of multivariate, nonhierarchical geostatistical analytic steps

4 The Calculation of the Indicator of Agricultural Resistivity to Transformation: How to Express an Objective Judgment

Consequently, in the examination of physical resources in the municipal space of Martinengo, the calculation of resistivity of agricultural enterprises took place through multivariate analysis utilizing the indicators, acquired through GIS, to examine the characteristics of the cadastral particles and of the business capital. These characteristics were also thereby assigned to a gradation based on classes of resistivity to urban transformation of the agricultural lands. One can thus observe in the following research which parameters were used in the study.

4.1 The Stability of Socio/Occupational Aspects

The first parameter refers to the structural component of the 155 agricultural enterprises examined in 2011. These businesses occupy 1,538 ha of useful agrarian surface and divide it into two halves, the first being grain growth and horticulture and the second being animal husbandry (6,087 cattle for meat and milk, 15,298 swine, 146,000 chickens for eggs, 39,000 turkeys). The indicators listed below highlight eventual latent weaknesses in business conduct through the calculation of the interdependence between stability of the business (*Sa*), enterprise employment (*Oa*), and typology of enterprise (*Ti*).

Table 3. Socio/occupational indicators

Indicator	*Formula*	*Variables utilized*
Sa = business stability	*Sa* = *f* (*a, b, c, d*)	*a* = age of the agricultural entrepreneur and eventual relatives employed fulltime. *b* = level of occupation of the entrepreneur and the familial components in the business *c* = level of involvement of the familial components of the entrepreneur *d* = enterprise defined by capitalistic conduct
Oa = labor units employed	*Oa* = *f* (*U.L.*)	*U.L.* = labor units employed (at least 287 days/person/year)
Ti = typology of agricultural enterprise	$Ti = \frac{U.L.fam}{U.L.tot}$	*U.L. fam* = labor units belonging to the family of the agricultural entrepreneur *U.L. tot* = total labor units employed in the company

The framework that results from such indicators characterizes the situation in Martinengo as follows: *i*) the greater part of the enterprises are run full-time by relatively young owners, fully operational in agriculture and even capable of competing and resisting in the oscillations of the market; *ii*) there is, however, a certain fragility with regard to the number of employees per business, different expressions of company management among which the cases of large companies with significant occupational

relevance stand out, in a positive manner; *iii*) limited presence of businesses of capitalistic operation, in a reality where family-run business is still the most diffused model.

In the face of such considerations, a cumulative parameter of the local socio/occupational reality was generated by:

$$Pso = \frac{Sa + Oa + Ti}{3}.$$

whose results generated 6 groups of enterprise with specific characteristics that were subsequently supplanted into the production of multivariate applications, which is the conclusive analytical step.

4.2 The Dynamism of Economic Factors

It is not necessary to reiterate how important the economic factors of the companies are to their competitiveness and thus also their possibility to continue to operate in the short/medium-term, through the examination of the areas of interest of the units of economic dimension (*Ude*), of the productive economic output (*Ir*), and of the possession of live exploitable capital (*Daz*).

Table 4. Economic indicators

Indicator	*Formula*	*Variables utilized*
Ude = unit of economic dimension	$Ude = \frac{R.L.S.}{a}$	*R.L.S.* = standard gross income *a* = 1.200 €/hectare
Ir = index of economic output	$Ir = \frac{Ude}{Sau}$	*Ude* = economic dimension of the company *Sau* = agricultural surface utilized
Daz = possession of live exploitable capital	$Da.z = \frac{Uba}{Sau}$	*Uba* = unit of adult cattle *Sau* = agricultural surface utilized

The synthetic framework demonstrates that: *i*) the majority of the municipal space is comprised of businesses of significant agricultural output, as opposed to those of reduced dimension and scarce economic significance; *ii*) nevertheless, many enterprises (of grain cultivation diffusion) present low values of economic output, as opposed to higher values in enterprises of animal husbandry and of specialized cultivation orientation; *iii*) the businesses of elevated livestock only represent ¼ of the total but cultivate more than half of the agricultural surface utilized and, vice versa, those businesses without livestock represent the majority of the total but only cultivate 1/3 of the land; *iv*) it is therefore a problematic framework for small enterprises that lack livestock, for the limited economic output from grain cultivation while, on the other hand, the remaining businesses are prized and competitive because of their specialized cultivation and breeding of livestock.

4.3 The Solidity of the Morphological Aspects

The third block considered the morphologic characteristics of the enterprises by inserting in the analytical/evaluative implantation other factors of some importance

such as the proprietary incidence (*Ip*), the productive potential (*Pp*), and the compactness of the (*Ca*). From these factors emerge: *i*) a company consistency of scarce proprietary incidence, for the fact that more than 1/3 of the companies integrate the land property with cultivated grounds that are rented; *ii*) a prevalence of businesses conductive to grounds of medium/high productive potential, derived from the general elevated agronomic propensity of the terrains; *iii*) a dominance of enterprises of significant morphologic compactness, with the concentration of bodies in adequate dimensions for the running of funds.

Table 5. Morphologic indicators

Indicator	*Formula*	*Variables utilized*
Ip = proprietary incidence	$Ip = \frac{Supprop}{Sat}$	*Supprop* = surface owned by the company *Sat* = total company surface
Pp = index of productive potential	$Pp = \frac{\sum \text{ha}\,(y_x)}{Sat}$	*ha* = extension of areas of diverse valence of Land capability (Ersaf) *Sat* = total company surface
Ca = compactness of the business	$Ca = \frac{Kn + Kj + Kd}{3}$	*Kn* = number of business bodies *Kj* = distance of the bodies from the company center Kd = dimension of the bodies

4.4 The Multivariate Applications for the Deriving of Agricultural Georesistive Basins

The classes of agricultural resistivity are derived from the application of multivariate geostatistics, in order to recognize the diverse profiles of statistical units (geospaces) with homogeneous characteristics. Once the set of nine variables/indicators has been determined, the software for multivariate analysis AddaWin[1], constructed specifically for territorial analysis and the first step, was used in the analysis of the correlations between the variables. This analysis facilitates the identification of the interconnected phenomena: the sole variables of correlative significance between them involve the economic dimension and yield (0.662), as well as the economic dimension and company occupation (0.614). Nevertheless, these values do not seem sufficient to be able to exclude some of the variables from the analytical process.

Given the absence of significant correlations (with a value ≥ 0.800), the second step (analysis of the principal components) was sustained by maintaining 9 initial variables whose reference indicators can be used to characterize the units of analysis. For each of these indicators, the Eigenvalue Decomposition was provided and the corresponding proportion (quota of variance explained by each component with respect to the total), accumulated as the sum of the Eigenvalues, as a way in which to

[1] *AddaWin* is the togetherness of multivariate analysis routines developed by Silvio Griguolo (Iuav, Venice) that is targeted for use in territorial applications. The package can be downloaded for free from the website: `http://cidoc.iuav.it/~addati/addati.html`

CORRELATIONS (*1000)

	A	B	C	D	E	F	G	H	I
A	1000								
B	57	1000							
C	330	-434	1000						
D	-59	371	-241	1000					
E	259	614	-314	292	1000				
F	176	504	-147	242	662	1000			
G	-134	-66	125	-32	-33	81	1000		
H	-78	-65	65	11	63	246	193	1000	
I	221	436	-328	132	562	408	89	20	1000

Fig. 3. Analysis of the correlations

evaluate how many resulting principal components are explicable through quota variance. This process allowed for the evaluation, by parsimonious criteria (minimum possible number of principal components, minimal loss of information, minimum deformation of the quality of representation) of the *n* components to consider in the analysis. Approximately 82% of the model (5 factorial axes) was adopted as significant, and the decisive step of the nonhierarchical analysis was launched.

9 DETERMINED FACTORS OF SIGNIFICANCE – EXPLAINED INERTIA:
TOTAL INERTIA = 9.000000

#	EIGENVALUE	EXPLAINED INERTIA (%)	CUMULATIVE INERTIA (%)	
1	3.0043111	33.381	33.381	******************************
2	1.4287684	15.875	49.256	**********
3	1.3136386	14.596	63.859	**********
4	0.8958135	9.953	73.806	**********
5	0.8079587	8.977	82.783	**********
6	0.5639517	6.266	89.049	********
7	0.3806026	4.229	93.278	*****
8	0.3774564	3.750	97.028	*****
9	0.2674990	2.972	100.00	****

Fig. 4. Inertia of the factorial axes

The treatment of the chosen variables allowed for the obtainment of an objective curve capable of describing the entire statistical sample with respect to an optimal number of classes, reducible or incremental through the subsequent aggregations/disaggregations. With regard to nonhierarchical classification, in this case, the cadastral particles grouped into geospaces on the basis of similar behavior reduces the model's complexity by assigning a quota of inertia to the number of classes included.

In this case, the objective curve was considered to reach a level of inertia (and, therefore, precision) of up to 70%, in coherence with the statistical tradition that

provides for limiting the consideration of the function where the trend does not present linear characteristics. In such a way, a number of stable classes equal to eight are derived. Below one can examine the relations between classes (in order) and variables (in abscissa), and the interpretation of the returned output is based on the symbols (+) and (–), where positive implicates a direct correlation and negative an inverse correlation.

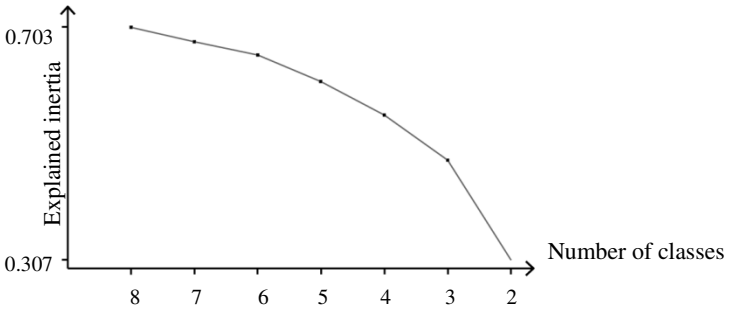


Fig. 5. Objective curve

- PROFILES of the 8 STABLE CLASSES -

			----	- -	~ ~	++		++++		Value of the
----------->|<----------->|<---------->|<---------->|<----------										ratio R
			-1.00	-0.20	0.20	1.00				
CLASS	NUM	IMPORT	A	B	C	D	E	F	G	H	I
1	18	11.6	2.94	2.72	1.78	2.50	2.89	2.56	2.44	2.28	2.33
			++	++++	~~~	++++	++++	++++	++	++	++
2	38	24.5	3.00	1.03	3.00	0.76	1.18	1.03	2.05	2.11	1.03
			++	----	++++	--	--	--	++	~~~	--
3	20	12.9	1.35	1.65	1.00	1.45	1.35	1.10	2.45	2.15	1.10
			----	--	--	~~~	--	--	++	++	--
4	26	16.8	2.69	1.88	1.12	1.73	1.77	1.23	1.12	2.00	1.04
			~~	~~~	--	++	~~~	--	--	~~~	- -
5	6	3.9	3.00	2.33	1.00	1.83	2.00	0.33	1.67	0.33	2.33
			++	++	--	++	~~~	----	~~~	----	++
6	13	8.4	2.85	2.00	1.00	1.23	2.08	1.46	2.77	2.15	2.23
			++	++	--	~~~	~~~	~~~	++	++	++
7	31	20.0	3.00	2.32	1.19	1.29	2.97	2.16	1.29	1.97	2.58
			++	++	--	~~~	++++	++	--	~~~	++++
8	3	1.9	2.67	2.00	1.67	0.00	0.00	0.00	0.00	0.00	0.00
			~~	++	~~~	----	----	----	----	----	----
	155	100	2.71	1.86	1.65	1.39	1.94	1.46	1.85	1.98	1.63

Fig. 6. Profiles emerging from nonhierarchical analyses

Table 6. Recoding of the classes of the nonhierarchical analysis based on agricultural resistivity

Classification of agricultural resistivity	*Multidimensional classes*
Class 1 – High resistivity	1, 7
Class 2 – Medium/high resistivity	5, 6
Class 3 – Medium resistivity	4
Class 4 – Medium/low resistivity	2
Class 5 – Low resistivity	3
Class 6 – Services for agriculture	8

An overall framework of significant resistivity of agricultural enterprises operating in the municipal context emerges, with the three best classes covering approximately 80.80% of the agricultural surface of Martinengo and including over 60% of total businesses. In the majority of cases, no particular criticalities emerge, and the businesses at medium/low resistivity are spread out in a uniform manner across the territory, located amongst those that are more resistive.

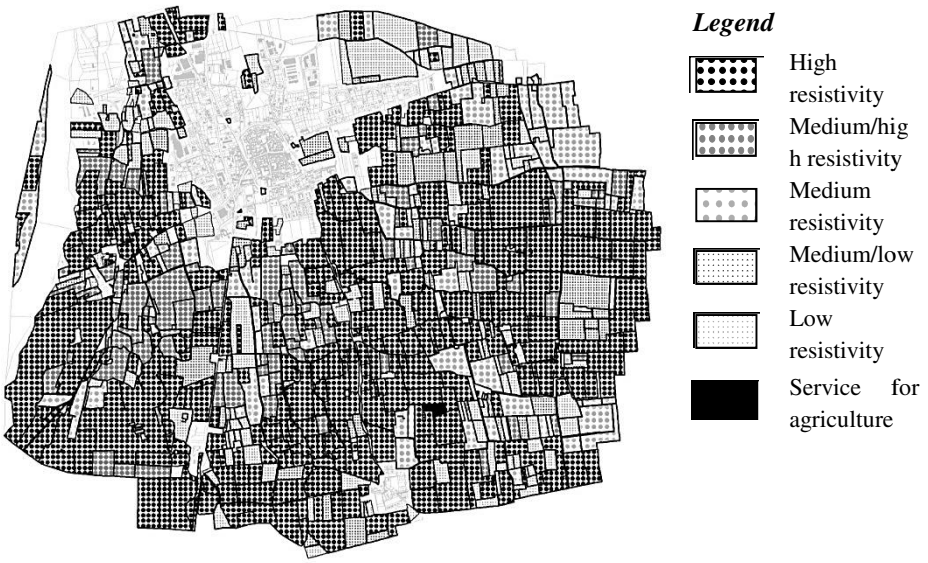


Fig. 7. Map of resistivity of enterprise

5 Conclusions: Possible Operational Implications

The results of the analyses conducted on the case study of the municipality of Martinengo show the prevalence of extra-urban geographies of a high and medium nature - high agricultural resistivity, highlighting contemporarily those areas of lesser agricultural aptitude – productive and, consequently, more suitable to a possible functional reclassification in the management of urbanization.

In such a way, the public decision-maker is in the position to operate with greater incisiveness and awareness, minimizing the waste of precious agronomical soil and limiting decisions regarding urban transformation solely to those spatial areas where the values of agricultural resistivity are such that urban transformation would not cripple the quality of the agrarian enterprise.

Simultaneously, the complexity of the in-depth analyses has allowed for the identification of those enterprises that operate actively and efficiently in agriculture, for optimal reception and for better investment of financial resources originating from the Common Agricultural Policy.

The experiment undergone in this case presents itself in continuity with the use of geostatistical methods in the domain of Geographical Information System, dedicated to the interpretation of territorial phenomena not only at the level of landscape but also at an agronomical and urbanistic level, in light of a more encompassing recognition, a more effective description, and a more incisive interpretation of their reciprocal interdependence within the environmental "prism" complex.

References

1. Casati, D., Maggiore, T.: Impresa, territorio e qualità pilastri del futuro agricol. In: Bollettino dell'Agricoltura, Giornale della Società Agraria di Lombardia, Atti. Fascicolo no 3 – 4 July – December, Milan (2009)
2. Fabbris, L.: Statistica multivariata. Analisi esplorativa dei dati. McGraw Hill, Milan (1997)
3. Fraire, M., Rizzi, A.: Statistica. Metodi esplorativi e inferenziali. Carocci, Rome (2005)
4. Griguolo, S.: Addati. Un pacchetto per l'analisi esplorativa dei dati – Guida all'uso. Istituto Universitario di Architettura di Venezia, Venice (2008)
5. Nijkamp, P.: Soft multicriteria analysis as a tool in urban land – use planning. Environment and Planning B 9(2), 197–208 (1982)
6. Nijkamp, P.: Multicriteria analysis; a decision support system for environmental management. In: Archibugi, F., Nijkamp, P. (eds.) Economy and Ecology; Towards Sustainable Development. Kluwer, Dordrecht (1989)
7. Paolillo, P.L., Borachia, V., Moretti, A., Tosi, A.: (a cura di): Il parametro suolo: dalla misura del consumo alle politiche di utilizzo. Grafo edizioni, Brescia (1988)
8. Paolillo, P.L. (ed.): Problematiche del parametro suolo: uno sguardo preoccupato alla situazione regionale. Angeli, Milan (2002)
9. Paolillo, P.L.: La misura della sostenibilità dei vincoli insediativi: un'applicazione di supporto alla Vas. Territorio 25, 65–76 (2003)
10. Paolillo, P.L. (ed.): Acque suolo territorio. Esercizi di pianificazione sostenibile, pp. 11–147. Angeli, Milan (2003)
11. Paolillo, P.L.: New survey instruments: studies for the environmental assessment report of the general plan in a case in Lombardy. In: INPUT 2008, Conferenza Nazionale in Informatica and Pianificazione Urbana and Territoriale, pp. 1–10. Politecnico di Milano, Facoltà di Ingegneria, Polo regionale di Lecco (2009)
12. Paolillo, P.L.: New survey instruments: studies for the environmental assessment report of the general plan in a case in Lombardy. In: Rabino, G., Caglioni, M. (eds.) Planning, Complexity and New Ict, pp. 215–224. Alinea, Florence (2009)

13. Paolillo, P.L.: Sistemi informativi e costruzione del piano. Metodi e tecniche per il trattamento dei dati ambientali. Maggioli, Rimini (2010)
14. Pieri, R., Pretolani, R. (eds.): Il sistema agroalimentare della Lombardia – Rapporto 2011. Angeli, Milan (2011)
15. Thurstain-Goodwin, M.: Data surfaces for a new policy geography. In: Longley, P.A., Batty, M. (eds.) Advanced Spatial Analysis, The CASA Book of GIS. ESRI Press, Redland (2003)
16. Voogd, H.: Prescriptive analysis in planning. Environment and Planning B: Planning and Design 12(3), 303–312 (1985)

The Overall Framework and Process Design of Active Service of Geographic Information System

Qiong Liu[1,2,*], Rong Zhao[1], and Lijian Sun[1]

[1] Research Center of Government GIS, Chinese Academy of Surveying and Mapping, China
[2] Cartography and Geographic Information Engineering, Shandong Agricultural University, China
apipip@126.com

Abstract. To solve the problem of inconvenient and difficult knowledge acquisition owing to the rapid expansion of geographic information and knowledge on the Internet, with the geographic information service mode, there's a brief introduction to the key technologies of active service system of geographic information used in data mining, information recommendation and information push, on this basis, it presents geographic information active service system, describes the active service function module and workflow, which lays the foundation for the Integration of geographic information system and the active service and solve the problem of information overload to some extent.

Keywords: Active service, data mining, information recommendation, information push.

1 Introduction

1.1 Background

The modern surveying and mapping geographic information technology, represented by geographic information systems, aerospace remote sensing technology and so on, penetrated into all aspects of national economy and social development, has becoming the basis means of sustainable development issues in China's population, resources, environment and disaster and other major social field after entering the twenty-first Century. In China's urban and rural planning, land management, environmental protection, disaster relief and national defense construction and other fields, it needs lots of basic geographic information and spatial data analysis service to support scientific decision. However, Internet is highly opened and the spatial information is updated lacking unified management, so how to quickly and accurately find the information needed becomes a big problem, this is the so-called "Rich Data Poor Information".

* Corresponding author.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 66–74, 2013.

Early, people use the search engine to obtain the required knowledge. However, this kind of search engine "pull" mode works like "looking for a needle in the ocean" with low efficiency, meanwhile, a large number of "information rubbish" makes people upset [1]. People hope to find a suitable method to get automatic acquisition, simple and refined information. The personalized service has become an important way to solve these problems.

1.2 Analysis

Long-term since, domestic and foreign scholars have been trying to explore and study personalized geographic information models and methods of active service. In response to the trends of user' intelligent, personalized and integration application needs, academic institutions and enterprises increased the speed of research and the efforts of development, the Micorsoft.Net platform from Microsoft, Web Sphere from IBM and Sun One from Sun is developed to face these challenges, to provide users with better Internet service and support; UDDI (Universal Description, Discovery and Integration), WSDI (Web services Description Language), SOAP (Simple Objects Access Protocol), XML (Extensive Markup Language) and other protocols and standards continue to be developed to support this demand for Internet services.

At present, there are still a variety of problems of domestic and foreign research in practice. Firstly, the existing services can't be based on a fixed calculation module function expansion in the implementation level. Secondly, existing services can't be aware of the changing needs of users and meet the needs of the user-oriented on-demand "services" based on static deployment.

These problems make geographic information active service can't be personalized based on user requirements, it's difficult to play a good role in the automatic learning, independent mining and updating information and obtain knowledge of active recommendation. To solve these problems, we need to change the application pattern of the current service that user can only use the existing services to enable them to select the appropriate feature set according to the specific needs, lead to the service mechanisms of active service discovery, customization, loading and using, to support new applications and services, on-demand creation, change the application pattern of the current service that user can only use the existing services. Also, it can help users obtain on-demand services and personalized service from the Internet whenever and wherever. The new application is the active service model.

2 Geographic Information Active Service System

2.1 The Concept of Active Service

Active service is a new computing model, which works according to the user's demand to provide user service requirements of information and knowledge from search, mining the Internet or the local network. The principles of geographic

information active service system is studied by mining different users' interest and valuable information from the huge information resources to meet the need of users. Its characteristics can be summarized as active information gathering and processing information, active mining knowledge, rule extraction, active dissemination of information, active demand forecasting, active human-computer interaction and actively adapt to the user [2]. Even they are same kinds of service, active service can also be personalized based on users requirements and characteristics, to customize service adapted to the requirements of users, which changes the Web service cannot be based on user needs and dynamic change.

2.2 Key Technology for Active Services

Research on geographic information active service mainly includes three aspects which are the active service model, service discovery and service customization. The service realization relies on the support of data mining, recommendation technology and information push technology.

1) Data Mining Technology.

Data mining is usually known as knowledge discovery in database [3], it extracts patterns which represents knowledge hidden in large databases, data warehouses or other information storage. At present, data mining is an important topic facing the mining complex types of data, including complex objects, spatial data, multimedia data, time series data, text data and Web data [4]. Its serves two purposes, one purpose is to use Web Mining to know the field of personalized data mining that users interested, and the other is to get the potential links hidden in spatial data.

a) Web mining techniques

Web mining is a technique combines traditional data mining techniques with the Web techniques, to extract useful patterns and hidden information from Web documents and Web activities [4]. Web Mining can be divided into three categories in accordance with the mining, Web content mining, Web usage mining and Web structure mining, which can be seen in Figure 1.

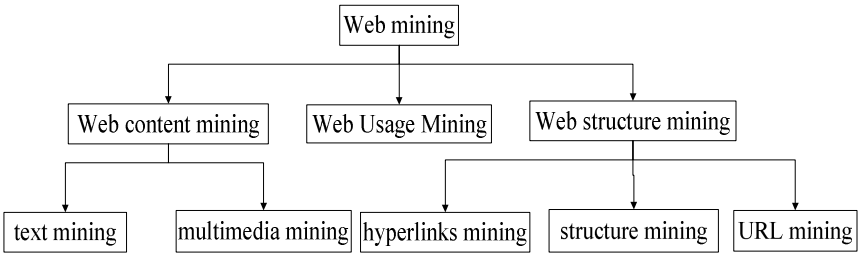


Fig. 1. Web mining classification

① Web content mining

Web content mining is to summarize, classify, cluster, analysis associated the collection of a large number of documents on the Web "content", and forecast the

trend with Web documents, is the process of extracting knowledge from the Web document content or description. It can be divided into two parts, namely text mining and multimedia mining, which depends on the content of processing.

②Web Usage Mining

Web usage mining can help find the users' interest from the Web page mode with Mining Web log record. The Web server usually store on the Web pages of each log entry (Weblog). The Weblog database provides a wealth of information about Web dynamic. As an important part of Web mining, Web log mining can bring users better analysis and understanding of the behavioral characteristics of users and find the agent and regular pattern hidden in users' behavior, so it is important to study the complex Weblog excavation technology. It helps to build custom Web services for individual users by analyzing such Weblog files.

③Web structure mining

Mining the potential Web link structure of the model, Web structure mining excavates information and knowledge derived from the Web organizational structure and their relationship of the link. This idea comes from the citation analysis, which creates a Web link structure mode by analyzing the number of objects in a web page link. This mode can be used for web page classification, therefore we can get the information of similarity and related degree between the different pages and related information. Web structure mining is consisted of hyperlinks mining, structure mining and URL mining.

In active service system, Web mining technology is mainly used to obtain users' interest information. At present, there are mainly two methods to get users interest information, the explicit collection and implicit collection [5].

The explicit collection requires the user to directly select the information they are interested in, such as web pages. This collection can obtain users' interest information directly, which needs users' feedback.

Implicit collection is an indirect form of information collection which is obtained from the analysis of user's Internet data rather than directly from the user's feedback.

b) Spatial data mining technology

Based on the data mining, the spatial data mining is combined with geographic information systems, remote sensing image processing, global positioning systems, pattern recognition, visualization and other relevant research fields. It's also process to extract from the spatial database implied, the interest to the user space and non-space model, common characteristics, rules and knowledge [6]. Knowledge of spatial data mining technology that can be tapped includes association rules, characteristic rules, classification rules, clustering rules, sequential patterns, integration of data and summarize, summarize the rules, trend analysis, variance analysis, pattern analysis.

There are many ways used in Data mining techniques, such as statistical analysis methods, genetic algorithms, rough sets, decision trees, artificial neural networks, fuzzy logic, rule induction, cluster analysis and pattern recognition, nearest neighbor techniques, visualization technology [7]. Spatial information representing approximately 80% of the total amount information of Digital Earth, so spatial data mining is an effective tool to converted information into useful knowledge. As a

result, it plays a very important role in the massive spatial data processing and has a broad application prospects.

2) Active recommend Technology.

The active recommendation technology is produced to meet the development of electronic commerce and information overload of network. Now the recommendation technologies are consist of recommendation based on content, collaborative filtering recommendation, recommendation based on association rules, recommendation based on utility, recommendation based on knowledge [8]. Among them, the collaborative filtering recommendation technology is one of the most successful recommendation technology [8].

The concept of collaborative filtering was first proposed by Goldberg, Nicols, of Oki and Terry in 1992 [9]. The principle uses history of the user's preference information to calculate the distance between them and then predicts the target users' preferences with the weighted evaluation value of geographic information assessment of nearest neighbor target user's, system recommended information to the target users by the preferences of high and low degree. Collaborative filtering recommendation technologies have mainly two types of algorithms, user-based collaborative filtering recommendation algorithm [10] and project-based collaborative filtering recommendation algorithm [11]. The former one is usually played in the nearest neighbor technology looking for a neighbor user, and the latter one in the project point of view, looking for certain items with the item.

3) Active pushing technology

As a new technology on the Internet, information active push service serves through regular transmission of information that users needed by certain technical standards or protocols. It can send the most up-to-date information categorized information based on the needs of users, which substantially increases the efficiency of the Internet information using [11].

Typically, the basic process of the active service technology begins by mining the users' needs, then searching the thematic information targeted online, and last regularly pushing information to the users [12].

3 The Framework of the System and Process Design

3.1 Footnotes

Be built on top of the Internet service technical specifications, active service provide service user-oriented in a certain function within the framework by searching and mining on the Internet. The function modules of Geographic Information active service system include data preprocessing module, data mining module, information recommendation module and push information module. The data mining module is divided into user interest mining module and spatial data mining module, which is shown in Figure 2.

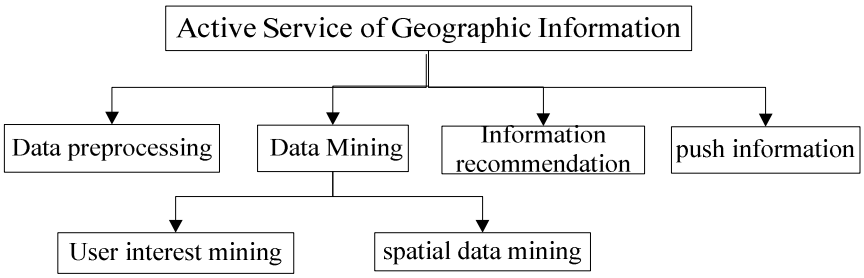


Fig. 2. The function module of geographic Information active service system

3.2 System Workflow

To achieve active service, the workflow is shown in Figure 3, the specific process is as follows.

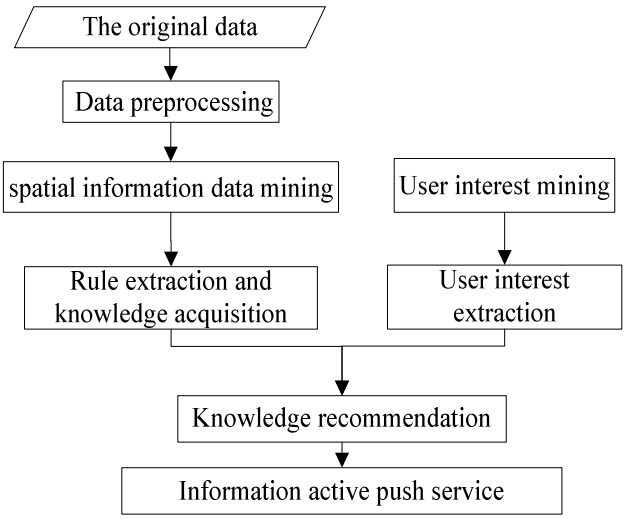


Fig. 3. The workflow of geographic Information active service system

1) Data preprocessing

Presence of incomplete, noisy and inconsistent spatial data is a common feature of most geographic information database. Therefore, in the implementation of the data mining algorithms need to preprocess the data. The data preprocessing mainly includes the following aspects. Firstly, it is data cleaning whose purpose is to remove noise or inconsistent data. Secondly, it is data integration which synthesizes the data in different data sources, including data integrity and consistency checks and noise filtering of data and incomplete information to fill, and so on. Then, it is data selection, to selectively extract related data according to the mining task from a

synthetic database. Data selection is intended to narrow the scope of the treatment and improve the quality of data mining.

2) Data Mining

Data mining is used for user interest and spatial data mining. Mining user interest is in order to provide users with better personalized geographic information services, spatial data mining is to improve the utilization of the data and update the data real-time.

a) Web mining user interest

Explicit way of collecting information can timely, accurate, but it requires a user to actively cooperation. Implicit method is obtained through the analysis of historical data on the user's access, which avoids the timely feedback problems from users. However, implicit collection must be built on a large number of data. So the geographic information active service system combines the explicit and implicit collection to excavate users' information of their interest.

b) Spatial information data mining

The process of spatial information data mining generally consists of 3 main stages [13], which is data preparation, mining operations, expression and interpretation of results. Firstly, system selects data mining algorithms on the data after preprocessing, the corresponding results (rules), and then analyzes the obtained results dialectically in forecasting or forecasting modeling work. To provide decision support, the stage of presentation and interpretation of results makes modeling results easily accepted by the users.

c) Active recommendation information

Before the active push, it's needed to classify the results according to different needs and interests of the users. This is the reasons why we need active recommendation. Recommendation algorithm is the most critical part, which determines the recommendation system [14].

Collaborative filtering techniques are based on the following assumptions [15] that user A interested in similar information of interest to the user B that user A is interested in. Its specific implementation process is as follows.

Firstly, system gets users' interest information and then extracts the interest model according to these information.

Secondly, system uses the similarity calculation method or statistical techniques to search for the category of users or information of nearest neighbors by analyzing and finding the characteristic pattern between the user and the information.

Thirdly, system timely outputs the recommendation list according to the current user's access or stage.

3) Information active push

The technology of active push information can provide users the knowledge of geographic information collected through mining and information recommendation technology treatment according to the user's interest. At present, information active push service implementations are mainly news, agent and channel three ways.

a) Web push media

In addition to select PC, internet Information active push can also compress the information and use SMS, micro blogging, mail pushed to provide the appropriate

geographic information services according to different environments and different subscription crowd.

b) Push feedback

The system will provide push results tracking function to guarantee the reliable information push and provide feedback to inform the failure of pushing information if the network is abnormal. At the same time, pretreatment is indispensable in the information push to reduce the probability of transmission failure, to improve the users' satisfaction accordingly with the system.

4 Summary and Outlook

This article proposes a new paradigm called Active Service of Geographic Information that aims to significantly enhance the customizability and reusability of Internet services. It presents a comprehensive framework and the associated technologies to realize the paradigm, which lays the foundation for the Integration of geographic information system and the active service. The ideas proposed are not only forward-looking but also deeply grounded in real world, and has great application prospects in the timeliness requirements (such as military, disaster prevention, disaster relief, etc.) real-time system.

It can be predicted, active service system will not only promote the development of space science, computer science, and will enhance human understanding of the world, the ability to find knowledge, in order to better transform the world, the service of human society.

Acknowledgment. This research was funded by National High-tech R&D Program of China (863 Program) under grant No. 2012AA12A309, the National Science & Technology Pillar Program under grant No. 2012BAH24B02, respectively.

References

1. Gu, Y.-Y.: A Hybrid Information Push Method Research and Its Application in EISP. Zhejiang University, Hangzhou (2008)
2. Wang, Z.-M., Chen, W.-W., Yang, S.: Research on Active Information Service and It's System Design. Computer Engineering and Applications (21), 110–113 (2003)
3. Linoff, G.S., Berry, M.J.A.: Mining the Web: Transforming Customer Data into Customer Value, pp. 66–115. Publishing House of Electronics Industry, Beijing (2004)
4. Liu, T.-M.: Data mining technology and its application, pp. 17–58. National Defence Industry Press, Beijing (2001)
5. Aggarwal, C., Yu, P.: Finding Generalized Projected Clusters In High Dimensional Spaces. In: Proceedings of ACM International Conference on Management of Data (SIGMOD 2000), pp. 70–78 (2000)
6. Estivill-Castrov, Houle, M.E.: Robust distance-based clustering with applications to spatial data mining. Algorithmica, 216–242 (2001)

7. Lv, A.-M., Lin, Z.-J., Li, C.-M.: Technique Methods of Data Mining and KDD. Science of Surveying and Mapping 25(4), 36–39 (2000)
8. Li, C., Zhu, Z.-M., Ye, J., Zhou, J.-Y.: Survey on research in personalization service. Application Research of Computers 26(11), 4001–4005 (2009)
9. Goldberg, D., Nicols, D., Oki, B.M., et al.: Using collaborative filtering to weave an information Tapestry. Communication of the ACM 35(12), 61–70 (1992)
10. Yu, C., Xu, J.R., Du, X.Y.: Recommendation algorithm combining the user-based classified regression and the item-based filtering. In: Proceedings of the 8th International Conference on Electronic Commerce: The New E-commerce: Innovations for Conquering Current Barriers, Obstacles and Limitations to Conducting Successful Business on the Internet, pp. 574–578. ACM Press, New York (2006)
11. Kim, B.M., Li, Q., Park, C.S., et al.: A new approach for combining content-based and collaborative filters. Journal of Intelligent Information System 27(1), 79–91 (2006)
12. Zhang, Y.-T., Wu, W., Zeng, X.: Analysis of the Information Push Technology. Journal of Yunnan Agricultural University 3(2), 116–120 (2009)
13. Geng, Y.-S.: The studying of the individualized information service operational mode. Researches in Library Science 9, 65–67 (2005)
14. Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large database. In: Proc.1993 ACM-SIGMOD Int. Conf. Management of Data (SIGMOD 1993), pp. 207–216 (1993)
15. Shi, Y.-Z., Zhen, H.: Research on personalized recommendation system based on collaborative filtering techniques. Electronic Design Engineering 20(11), 41–44 (2012)

A Markov-Kalman Model of Land-Use Change Prediction in XiuHe Basin, China

Huiqiong Xia[1], Hai Liu[1,*], and Chunyan Zheng[2]

[1] Faculty of Resources and Environmental Science, Hubei University, Wuhan 430062, China
liuhai11191@163.com
[2] Department of Geography, Jiaying University, MeiZhou City, China

Abstract. A study of forecasting the dynamic change of regional land resources will reveal the characteristics and laws of the land use structure. Three periods of TM images of XiuHe basin in year 1990, 2000 and 2010 were used as data source, and a united model of markov and Kalman method, which called Markov-Kalman filter model, was applied to simulation the dynamic changes of land use in the study area. When land use predictions were gotten based on Markov model, a measurement value with error from the TM images was utilized to calibrate the forecast result in Kalman filter model, and the optimal estimate as accurately as possible to the real value was gotten, and quantitative forecasts and analysis of land change were acquired. The experiment results show that the method can effectively improve the forecast precision, and analysis of the forecast results will help the government realize the future evolution trend of land use structure in XiuHe basin.

Keywords: Land use, Markov model, Kalman filter, Dynamic simulation.

1 Introduction

Land use/Land Cover Change (LUCC) is one of the important cause for today's global environment Change, the typical area of case study is the necessary way to meet global environmental Change, also is the key to the regional sustainable development research [1]. Study on land use/land cover change in ecological fragile region, and understanding the region ecological environment change trend, can help people to reveal the nature and humanities factors that affect on sustainable development of land use, and provide the basis for government to make scientific decisions[2]. Markov model has been presented in simulation of land use/land cover changes for a long period and achieved outstanding effect[3][4][5][6] , But Markov model assumes that the transition probability matrix stays the same, namely in the forecast period, land use structure must be stable. However, the stability of the land use structure will be affected inevitably by the factors of social, economic, and ecological systems. As a result, the predictions of land use from Markov model always contain all kinds of error, including random error, systemic error, or the integrated error. If some

* Corresponding author.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 75–85, 2013.

measurements can be available in the periods, the forecasts from Markov model may be calibrated by some means. Kalman filter method is a kind of linear unbiased minimum variance estimation, and can be used for any linear random system. Applying the method, the model error and measurement error can be processed , and the optimal estimation of system state variables will be acquired[7].

In this paper, TM remote sensing images in XiuHe basin in 1990, 2000 and 2010 were used as data sources, when the average transition probability matrix with the step of 10 years is available through the raw data, the Markov model is presented to forecast the land use changes in the study area. During the course, measurements from the TM images in 2000,2010 were introduced to calibrated the forecasted results via Kalman filter method, and the optimal estimates as accurate as possible to real value was obtained as the expected value, and land use patterns of the XiuHe basin in the next 30 years were predicted.

2 Study Area and Data Source

2.1 Study Area

XiuHe, also known as xiushui, is one of the five major rivers in JiangXi province, located between 113°55'E～116°01'E and 28°40'N～29°30'N. XiuHe basin is surrounded by high mountains in north, west and south. The whole basin includes seven counties, such as XiuShui, WuNing, YongXiu, TongGu, JingAn and AnYi. The area is about 14700 km^2, and there is nearly 2 million people live there[8].

There are many mountains and rivers in XiuHe basin and the terrain is complex, Mountain accounts for 15%, hilly accounts for 15%, and the terrace plain accounts for 37%. The river valley climate belongs to the southeast Asia monsoon region, the average temperature is 17°C, frost-free period is about 260~280 days, annual average rainfall is 1604 mm, it is suitable for plant growth.

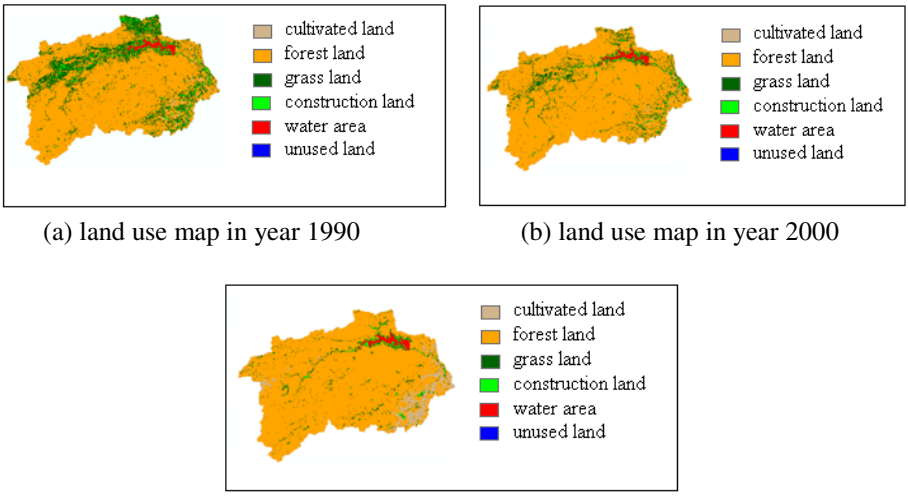
Hydraulic engineering in XiuHe basin has been repaired by the government after 1950s, and water resources are plentiful. Total effective irrigation area is about 1100km^2 , and the security area is more than 900 km^2.

There are rich plant species, intricate fauna components and various vegetation types in the XiuHe basin. The wetland of XiuHe basin is richer in plant fauna in JiangXi province and it has great scientific research value and protective value[9].

2.2 Data Source and Data Processing

In this paper, TM images in XiuHe basin in 1990, 2000 and 2010 were used as primary data source, supplemented 1:10 0000 topographic maps of the region。 The TM images were preprocessed in ERDAS IMAGINE 9.2 software operating platform. On the basis of land use classification system promulgated in 1984 by the government, using the TM imagines as the base map, and considering the land utilization condition of XiuHe, three periods Land use map were obtained as data source in the following experiments.

In the course of image preprocessing in ERDAS software, works of geometric correction and image registration of TM images was proceed on the basis of the l: 100000 topographic map in XiuHe basin. Then image Mosaic was accomplished based on geographical coordinates, and according to vector data of XiuHe basin's boundary, graphics clipping was carried on. When all works of TM images preprocessing completed, band 4, 3 and 2 of TM images were used to composite false-color images, and image was enhanced, and the images was used as the base map for supervised classification; Different types of land were extracted also by using ERDAS software, supervised classification was used in which maximum likelihood classifiers was run on the TM images. When the classification course finished, post-processing was carried out. According to the relevant data of XiuHe basin, further work of artificial visual interpretation was proceeded to modify the classification results for a higher interpretation accuracy. Final six types of land use were determined, i.e. cultivated land ,forest land, grass land, construction land, water area and unused land, then land use maps in 1990,2000,2010 were generated(Fig. 1).



(a) land use map in year 1990 (b) land use map in year 2000

(c) land use map in year 2010

Fig. 1. Land use maps in study area in1990,2000,2010

According to the land use maps, development of the region in recent years and the trend of land use change was revealed. Calculation of the area in square kilometers of the resulting land cover types for each study year was done and subsequently comparing the results as show in Table 1. The figures presented in Table 1 represents grass land, construction land and water area have been increasing, cultivated land and forest land have been decreasing respectively. It can be noted that construction land and grass land have been increasing 47.9% and 9.1% respectively from 1990 to 2000, unused land has been decreasing 62.9%, it implied that development of unused land was quickly in study area. During 1990-2010, construction land and water area has been keep up increasing, at the same time, cultivated land and forest land has been decreasing slowly, and unused land has been remained stable. During1990 and 2010,

unused land and cultivated land has been decreased quickest, the former decreased 62.9% and the latter decreased 7.6%. It suggests that the protection of cultivated land in XiuHe basin has been insufficient and exploitation progress of unused land has been rapidly.

Table 1. Area in Land use for years 1990, 2000, 2010 (unit: km^2)

Land use categories	1990	2000	2010
Cultivated land	3318.435	3285.765	3247.164
Forest land	8715.015	8856.247	8609.117
Grass land	1413.630	1542.668	1552.415
Construction land	76.747	113.513	300.863
Water area	393.975	396.360	485.610
Unused land	441.495	164.745	164.128
Total	14359.297	14359.297	14359.297

3 Methodology

In this paper, a Markov-Kalman filter model was introduced to simulate the changes of land use in XiuHe basin. First Land use data in several periods was used as reference data, and transition probability matrix and the area of each type land was computed via Arcgis 9.2. Base on the land use data at interval t_i, Markov model is presented to obtain the prediction area in the next interval t_{i+1}, the measurement at t_{i+1} from the land use map was used to calibrate the prediction value in a Kalman filter course to obtain an optimal estimator, which would be a substitution of real value.

3.1 Markov Model

Markov chain models are particularly useful to geographers concerned with problems of movement, both in terms of movement from one location to another and in terms of movement from one "state" to another. "State", in this context, may refer to the size class of a town, to income classes, to type of agricultural productivity, to land use, or to some other variable[10]. To model a process of land-cover change by a Markov chain, the land-cover distribution at t_2 is calculated from the initial land use/cover distribution at t_1 by means of a transition matrix. The Markov chain can be expressed as:

$$X_{t+1}=f(X_t) \tag{1}$$

Where, X_t and X_{t+1} denoted the condition of land use at time t+1 and t respectively.

Markov model applied to the field of land use change could be described as follows[11]:

$$n_{t+1} = Pn_t \tag{2}$$

$$P = (P_{ij}) = \begin{Bmatrix} P_{11} & P_{12} & \dots & P_{1m} \\ P_{21} & P_{22} & \dots & P_{2m} \\ \dots & \dots & \dots & \dots \\ P_{m1} & P_{m2} & \dots & P_{mm} \end{Bmatrix} \tag{3}$$

Where n_t is a column vector, whose elements are the fraction of land area in each of m states at time t, and P is an $m \times m$ matrix, whose elements, p_{ij}, incorporate the birth, death, and change rates of each state during the time interval (or 'time-step') from t to t +1,i, j=l,2…m; the sum of column vector is 1:

$$\sum_{i=1}^{n} P_{ij} = 1, \quad i,j = 1,2, \ldots m. \tag{4}$$

3.2 Markov-Kalman Filter Model

Kalman filter method runs by taking the minimum mean square error (mse) as the best estimate criterion, using the state space model of signal and noise, and utilizing the estimates of the moment and the observation value of the moment before to update the estimate of the variables. The algorithm will get an estimate satisfied the minimum mean square error for signals based on the system equation and observation equation[12].

The realization process of Kalman filter is as follows:

Step 1: Getting the estimation of land use structure at time k. The equation is as follows:

$$X(k) = A\tilde{}X(k-1) + Bu(k) + W(k) \tag{5}$$

Where $X(k)$ is the state of land use at time k by using the state at time k-1,A is the transition probability matrix of land use, $\tilde{}X(k-1)$ is the optimal value of the system at time k-1, and $A\tilde{}X(k-1)$ reflects the prediction process of Markov model, $u(k)$ is the control value at time k. If land use change is assumed as a natural process, and there is no human factors during the course, namely no control value, hence $u(k)$=0;$W(k)$ is process noise, the process is a Gaussian white noise for linear stochastic differential systems.

Step 2: Calculating the error related matrix P, estimating the accuracy of measurement.

$$P(k-1) = A\tilde{}P(k-1)A' + Q \tag{6}$$

Where, $P(k-1)$ is the covariance of X(k-1),~P(k-1) is the covariance of $\tilde{}X(k-1)$, A'is the transposed matrix of A, Q is the transposed matrix of $W(k)$, it can also be considered as the system noise variance. System noise variance can be computed by considering all kinds of error in the process of land use change.

Step 3: Computing Kalman Gain $Kg(k)$.

$$Kg(k) = P(k-1)\,H'/(H\,P(k-1)\,H' + R \tag{7}$$

Where, H can be assigned to 1 because the measurement value of land use is from TM images, and it is corresponding to the types of land use directly. R is the noise covariance of the measurement, here to simplify the processing procedure, it can be treated as a constant matrix. Considering the system is a single input and single output course, R is turned into a 1 x1 matrix, that is, a constant. The formula as above can be simplified as follows:

$$Kg(k) = P(k-1) / (P(k-1) + R) \tag{8}$$

Step 4: Computing the error fed back from state variable of Land use.

$$e = Z(k) - H\,X(k) \tag{9}$$

Where $Z(k)$ is the measurement with noise. The land use map was obtained through TM images, and some factors, such as sensor error, influence from the observation environment, cloud cover, geometric correction of image processing and classification accuracy, from satellite images will bring out error into the land use data, so the measurement was followed by error inevitably. X_k is the state value at time k, namely the predicted value. The equation above can be simplified as:

$$e = Z(k) - X(k) \tag{10}$$

Step 5: Update the state variables of land use

$$\tilde{X}(k) = X(k-1) + Kg(k)\,(Z(k) - X(k-)) \tag{11}$$

Where $\tilde{X}(k)$ is the optimal estimate value at time k.

Step 6: Update the error correlation matrix P, in order to obtain the optimal value of land use at time k+1.

$$P(k) = (\mathrm{I} - Kg(k))\,P(k-1) \tag{12}$$

Where I is an identity matrix, $P(k)$ is covariance of $X(k|k)$ at time k.

4 Results and Discussion

4.1 Analysis of Land Use Change in Study Area

In the course of simulation of land use change in XiuHe basin, transition probability matrix was acquired by computing the average of transition probabilities ,step for 10 years, from land use change data in the interval of 1990 to 2000, and the interval of 2000 to 2010, based on the land use map in 1990, the change process of land use in study area was simulated, and the simulation value of land use was calibrated by the measurements in 2000 and 2010 via Kalman filter model, meanwhile the structure of land use in study area at the next thirty years was predicted.

The Markov model was introduced to simulation the change course of land use in study area, and the transition probability matrix was the key for the model. Transition probability matrices at the interval 1990 to 2000 and 2000 to 2010 were calculated respectively, and then an average of transition probability matrices was computed (Table 2).

Table 2. The average transfer probability matrix within 1990-2010(step for 10 years)

Land use categories	Cultivated land	Forest land	Grass land	Construction land	Water area	Unused land
Cultivated land	0.4462	0.3340	0.1775	0.0114	0.0064	0.0245
Forest land	0.1355	0.7313	0.1083	0.0033	0.0156	0.0060
Grass land	0.2237	0.6046	0.1006	0.0155	0.0497	0.0059
Construction land	0.2354	0.1934	0.1031	0.3420	0.1048	0.0213
Water area	0.0724	0.0789	0.0470	0.2803	0.5020	0.0194
Unused land	0.3056	0.2470	0.1502	0.2281	0.0102	0.0589

Land use map in 1990 was used as base data n_t, and the average transition probability matrix showed in table 2 was applied as p_{ij}, the prediction value of land type in 2000 and 2010 was computed by formula (2) above. The predicted results and error is reported in Table 3.

Table 3. Predicted results and error in study area in 2000,2010

Land use categories	2000		2010	
	Predicted area (km^2)	Error	Predicted area (km^2)	Error
Cultivated land	3159.310	3.85%	3118.280	3.97%
Forest land	8491.305	4.12%	8475.058	1.56%
Grass land	1767.812	14.59%	1739.019	12.02%
Construction land	175.885	54.94%	366.021	21.66%
Water area	437.771	10.45%	496.266	2.19%
Unused land	177.214	7.57%	164.653	0.32%

Error showed in table 3 is the deviation value compared with prediction and measurement. The error of construction land in 2000 is 54.94%, bigger than other land type, and remains bigger in 2010. The reason causing the phenomenon is that urbanization in study area has been speeded up between 2000 and 2010, and Land area of grass land and construction land has increasing so quickly that the speed surpassed the period of 1991-2000. The structure of land use in study has been changed, and the deviation between predicted value and measurement displayed larger.

In view of the measurement was not very accurate because of some errors, hence errors showed in table 3 can't truthfully reflect the accuracy of prediction value of land type via Markov model. Now we presented the Kalman filter model to obtain the optimal estimate value of land type, which would be a substitute of desired value.

Kalman filter model combined the predicted value from Markov model and measurement from land use map, utilized measurement to calibrate the predictions and an optimal estimate value was obtained. When next prediction course processed, the optimal estimate value was used.

In our study by utilizing Kalman Filter model, we assume the initial deviation of predicted value of land use in 2000 is 10% of the predicted area of each land type, and the deviation of measurement is always 5%. To simplify the simulation course, we ignore the process noise, namely $W(k)$ is 0.

The Kalman gain, optimal estimate value, and deviation of optimal estimate value is showed in table 4.

Table 4. The output value of Kalman filter model

Land use categories	2000			2010		
	Kg	Optimal estimate area (km^2)	deviation of optimal estimates (km^2)	Kg	Optimal estimate area (km^2)	deviation of optimal estimates (km^2)
Cultivated land	0.937	3207.651	110.397	0.771	3209.527	85.679
Forest land	0.891	8605.857	286.636	0.769	8539.451	248.558
Grass land	0.941	1641.354	52.184	0.770	1601.856	44.857
Construction land	0.938	281.335	3.834	0.516	327.859	10.800
Water area	0.930	415.324	13.290	0.752	490.586	13.785
Unused land	0.935	171.542	5.555	0.770	164.447	4.753

Precision of prediction in Markov model and Markov-Kalman filter model were analyzed by comparing the predicted value with measurement value in 2010(Table 5).

Table 5. Comparison of results in Markov model and Markov-Kalman filter model

Land use categories	2010			
	Predicted area in Markov model (km^2)	Compared with measurements	Predicted area in Markov-Kalman filter model (km^2)	Compared with measurements
Cultivated land	3118.280	3.97%	3190.174	1.76%
Forest land	8475.058	1.56%	8502.340	1.24%
Grass land	1739.019	12.02%	1632.081	5.13%
Construction land	366.021	21.66%	345.312	14.77%
Water area	496.266	2.19%	493.761	1.68%
Unused land	164.653	0.32%	164.630	0.31%

We predicted the land use in 2020, 2030 and 2040 based on the optimal estimate value in 2010 with the Markov model to validate its applicability in the study area, and transfer probability matrix in Table 2 was utilized. The forecasting period was 30 years. The result is showed in Table 6.

Table 6. The output value of Kalman filter model

Land use categories	Predicted area in 2020(km^2)	Predicted area in 2030(km^2)	Predicted area in 2040(km^2)
Cultivated land	3110.473	3095.117	3090.439
Forest land	8428.098	8407.120	8387.144
Grass land	1737.220	1727.575	1725.075
Construction land	376.747	401.344	415.392
Water area	515.680	537.768	550.521
Unused land	165.508	164.802	165.154

In order to understanding the trend of land use change clearly, the statistics of area ratio of each land type from 1990 to 2040 was showed in Fig.2.

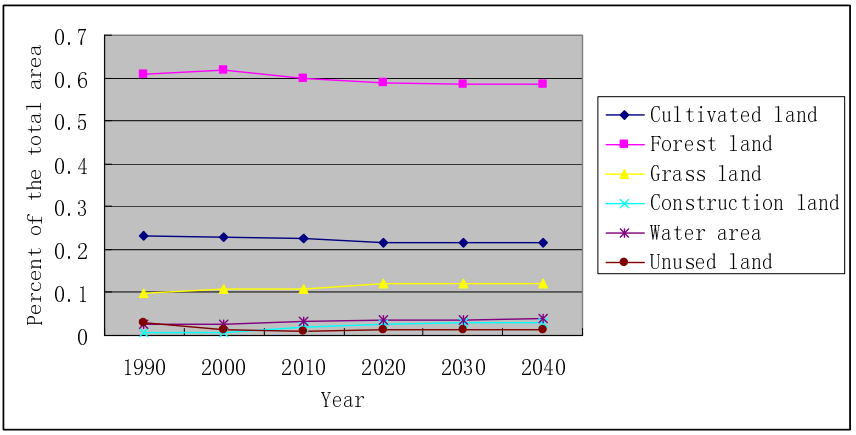


Fig. 2. Land use change trends within 1990-2040

4.2 Results

From Table 4, precision of predicted value in Markov-Kalman filter model is universal higher than in Markov model, which implies the method of forecasting land use change using Markov-Kalman filter model is feasible.

According to Table 6 and Fig.2, cultivated land, forestland and grassland were still the main landscape categories in the study area, covering more than 93% of the total area in next 30 years (Table 6). Forest was the largest land cover type, but it will be decreasing from 60.6% of the total area in 2010 to 58.5% in 2040.On average, 109.29 km^2 forest will be lost per 10 years from 2010 to 2040 due to deforestation. Cultivated land will continue decreasing, declining by 153.73 km^2, from 2010 to 2040. The area of Construction land will increase steadily throughout next 30 years, the growth is 0.5% from 2010 to 2030 increased 12.61% from 1974 to 2000. The water area will also increase slowly throughout the study period, while the area of unused land will remain relatively stable.

The proportion of construction land continues to rise means the industrialization and urbanization in XiuHe basin will be accelerated. The destruction of the forest caused by industrialization, urbanization and people's production activities has become increasingly serious in study area. The area of unused land decreased dramatically from 1990 to 2010, which shows that utilization of the unused land previously was efficient, but with fewer unused land, People will have to the face with the shortage of land.

5 Summary

This paper presented a Markov-Kalman filter model to improve the precision of forecasting of land use in XiuHe basin. Error of measurement value from land use map was considered in this method. When predicted value from Markov model was obtained, the measurement value with error was used to calibrate the predicted value in Kalman filter, then an optimal estimate value would be acquired, and the next prediction continued based on the optimal estimate value. The results from Markov-Kalman model revealed that the method could improve the precision of predicted value. Based on the optimal estimate value in 2010, The prediction of land use in 2020,2030 and 2040 was simulated, but the measurements in 2020,2030 and 2040 couldn't be gotten, and those predictions weren't calibrated in Kalman filter.

Results from the study in XiuHe basin indicates that the forest area and cultivated land will remain decreasing, construction land will increase rapidly, water area will have a slight increase and unused area decrease quickly from 1990 to 2010 and will be stable in next 30 years. With regional population growth, economic development and urbanization process accelerated, forest land continues decreasing and utilization of unused land is speeded up, which will causes forest deterioration and the shortage of cultivated land. The time period looked at here is too short to adequately reflect the long-term effects of such changes in XiuHe basin.

References

1. Peng, F.: Research on Land Use/Land Cover Changes in Yinchuan City Based on RS and GIS. LanZhou University, LanZhou (2011)
2. Zhou, X.S., Sun, B.P., Zhao, Y., et al.: Study on the Change and Prediction of land Use/vegetation Dynamic-taking the She County as Example. Chinese Agricultural Science Bulletin 26(19), 306–311 (2010)
3. Michael, I., David, L., Ahmed, E.-G., et al.: A Markov Chain Model of Land Use Change in the Twin Cities, 1958-2005. In: Image Segmentation (2012), http://en.wikipedia.org/wiki/Image_segmentation (retrieved September 30)
4. Levinson, D., Chen, W.: Paving new ground: a Markov chain model of the change in transportation networks and land use. In: Levinson, D., Krizek, K. (eds.) Access to Destinations, ch. 12, pp. 243–266. Elsevier (2005)
5. Ye, B., Bai, Z.: Simulating land use/cover changes of Nenjiang County based on CA-Markov model. In: Li, D. (ed.) Computer and Computing Technologies in Agriculture. IFIP, vol. I, pp. 321–329. Springer, Boston (2008)

6. Md, S.I., Raquib, A.: Land use change prediction in DHAKA city using GIS aided Markov chain modeling. Life Earth Sci. 6, 81–89 (2011)
7. Tian, Y., Lei, X.-H., Jiang, Y.-Z., et al.: Summary of Study on Real-Time Correction Technique of Flood Forecasting. Yellow River 33(3), 25–28 (2011)
8. Kong, Q.J., Fang, G.H.: Water allocation plan research in XiuHe Basin. Yangtze River 40(1), 27–30 (2009)
9. Chen, X.J.: Preliminary Study on Wetland Plants of Xiuhe Valley in Jiangxi Province. Acta Agriculturae Jiangxi 22(11), 164–166 (2010)
10. Lyndhurst, C.: An Introduction to Markov Chain Analysis, pp. 1–33. The Invicta Press, Ashford Kent (1995)
11. Baker, W.: A review of models of landscape change. Landscape Ecology 2, 111–133 (1989)
12. Peng, D.-C.: Basic Principle and Application of Kalman Filter. Software Guide 18(11), 32–34 (2009)

Modeling of Urban Post-disaster Reconstruction Using Extended Cellular Automata

Yun Cheng[1], Zhaoyang He[2], Yichun Xie[3], and Huifang Deng[4]

[1] School of Information Science and Technology, Zhongkai University of Agriculture Engineering Guangzhou, Guangdong, China
[2] School of Economics and Management, China Three Gorges University Yichang, Hubei, China
[3] Institute for Geospatial Research and Education, Eastern Michigan University 125 King Hall, Ypsilanti, Michigan 48197, USA
[4] School of Computer Science and Engineering, South China University of Technology Guangzhou, Guangdong, China

Abstract. A city is the symbol of human beings civilization. A sudden natural disaster may let this thousand-of-years civilization be destroyed in a minute, so the post-disaster reconstruction is an inevitable task of human encounter. Based on the dynamic mechanism for urban reconstruction and the features of city renewal, we presented the principle and technical procedures for constructing an extended cellular automata (CA) reconstruction model, used the model to simulate and analyze the extended urban post-disaster reconstruction of Great Tangshan, and made a comparison with the city's status and overall planning results. The model considered the overall interacting relationship among spatial elements affecting cellular transition rules. These elements include policy and planning, cellular desire (demand), population, degree of destruction and so on. The simulation results show that the CA model can predict the direction of urban development and reconstruction, and provide theoretical basis for authorities to effectively make the land management rules and policies in urban post-disaster reconstruction.

Keywords: Model, Cellular Automata, Urban post-disaster, Reconstruction, Land use, Urban planning.

1 Introduction

A city is the symbol of human beings civilization, and also the center of the production, accumulation and dissemination of material and spiritual wealth. A sudden natural disaster may let this thousand-of-years civilization be destroyed in a moment, For example, the horror Tangshan Earthquake on July 28, 1976, the Wenchuan earthquake on May 12, 2008, and Ya'an earthquake on April 20, 2013, and so on, caused huge damages to people's lives and fortunes. So the post-disaster reconstruction is an inevitable task that human has to carry out.

The city's evolution is impacted by the natural, social, economic, political and other factors and evolution behavior is highly complex. How to construct an urban

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 86–97, 2013.

reconstruction model to describe, simulate and analyze the complex dynamic behavior during the course of the city's reconstruction and varying expansion after the disaster, is a fundamental issue of the studying urban rebuilding, and of great scientific and practical significance. Because the city is a complex grand system, many traditional urban models are lack of direct numerical simulation of dynamic city system. Cellular Automata (CA) is a "bottom-up" dynamic simulation modeling framework [1], and capable of modeling temporal-spatial evolution of the complex systems. For almost two decades, the cellular automaton (CA) has been proven to be a popular and sometimes effective modeling tool for studying the complex urban systems [2].

This paper is organized as follows: In Section 2, we described characteristics of urban reconstruction. In Section 3, we specified the study area; In Section 4, we presented the design of urban reconstruction model. In Section 5, we gave the model predictions, compared and analyzed the simulated results, and In Section 6, we drew the conclusions.

2 Characteristics of Urban Reconstruction

2.1 Dynamical Mechanism for Urban Reconstruction

Urban rebuilding after the disaster is an interactive result between macro and micro factors such as natural, economic, policy- oriented, and other aspects, and the scale and pattern of reconstruction are closely associated with the defined inner functions of the city [3], therefore, understanding of the dynamic mechanism for rebuilding will help us to accurately master temporal-spatial discrepancy of urban reconstruction. Main factors influencing the urban post-disaster reconstruction included: 1) natural element such as topography, geomorphy, climate and geological conditions around a city; 2) traffic factors; 3) population growth; and 4) government's policy and planning control; 5) the rate of economic growth; 6) the extent of the damage after disaster.

2.2 Mode of Urban Renewal

The expanding mode of urban reconstruction is divided into the reconstruction of the original destroyed urban land and spreading boundary growth [4].

Reconstruction on the original destroyed land is often constrained by regional economic and natural conditions. Inside the city, some open land, grass or destroyed land gradually become urban land, making a sparse city become a dense city gradually, and also exhibiting randomly generated local urbanization in the outer urban. This process reflects the scale effect and the clustering effect of urban development. That is, when a regional urban development reaches a certain scale, there will appear a new urban unit (cell) within and around this region.

A spreading boundary growth is one of the most common patterns of urban growth. Due to urban population growth, industrial and commercial developments, the demand for urban land use increases, and the city extends to the surrounding agricultural land. Urban spreading means growth of new urban cellular around the original city cellular, reflecting the urban agglomeration (clustering) effect of the development.

3 The Study Area

3.1 An Overview of Area Selection

The Tangshan earthquake of 1976 was one of the largest earthquakes in recent years, one of the top ten natural disasters during the 20th century. And the earthquake took place in the highly industrialized and urbanized region of Beijing, Tianjin and Tangshan, The magnitude of earthquake is high and the disaster is severe. The Great Tangshan earthquake caused serious damage and loss of life in this densely populated industrial area. More than 280 km of roads were severely damaged, 71 large and medium bridges, 160 small bridges and more than more than 1000 road culverts were collapsed and destroyed, the roadbed of trunk road to Tianjin, Beijing, north-east and the coast were collapsed or cracked, the traffic is basically cut off, East-West Rail way was cut off, Jing-Shen railway was paralyzed.

3.2 Collection and Pre-processing of Data

(1) Data Source

The scale of our study is within the municipal area of Tangshan, whose area is 725.16 square kilometers. Due to limited data sources, key data used for the models include:

1) Raster (grid) data: TM images of land use Map (jpg format);

2) Vector data: the topographic map (scale of 1:20,000) of the year 1976, urban land use map, urban plan (urban overall planning blueprint for 1976-2003).

(2) Remote sensing image preprocessing

Using American remote sensing image analysis software (Erdas 8.7) for image processing and analyzing, we extracted information through computer automatic interpretation combined with artificial interpretation.

1) Geometric correction of remote sensing images;

2) Band combinations and integration;

3) Enhancement processing of remote sensing images.

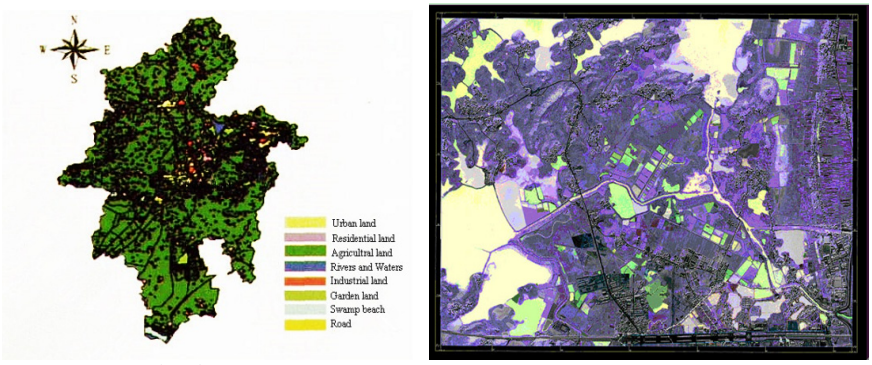
(3) Extracting the scale of urban land use

1) Extraction from classified remote sensing image

The land use category includes urban land use, rural land use and rivers. And then we transform the classified data into a GRID format, which is further converted into the SHP format, modify some land use codes incompatible with the actual land use in the ArcGIS, and then, convert the modified data back to the GRID format which is used as the input data of model. The land use category in Tangshan in 1993 is shown in figure 1(a).

2) Extraction from land use status map

Using land use data of 1993 and 2003, first we make a match (i.e., registration) of a remote sensing image with the land use map, and then vectorize the urban land area and rivers; the rest area belongs to the rural land, and finally converts the vectorised data into the GRID format. The status of the land use in Tangshan in 1993 is shown in figure 1(b).



(a)Tangshan land use type in 1993 (b) Tangshan land use map in 1993

Fig. 1. Tangshan urban land use category (a) and land use map (b) in 1993

According to the data of economic development and spatial variations of Tangshan City along with resolution of remote sensing image data, we choose the cellular grid size as 30m×30m. The whole area is 3,874 square kilometers, in which the urban construction area accounts for 187 square kilometers, constituting a cellular space of 1143×846=96978. Classified statistical data are given in Table 1.

Table 1. Tangshan urban land use statistics (unit: number of grids)

Land use \ year	1976	1993	2003
Urban land	832	2536	4677
Rural land	89448	87918	86287
The river water	6698	6524	6014

(4) Acquisition of slope data

Slope data are taken from Tangshan City 1:20,000 scaled topographic map in 1993. We converted a contour format map to SHP format, then conduct a 3D analysis in order to build a digital elevation model diagram in a TIN format, which is then converted to the GRID format for the slope analysis, and, finally generate the gradient layer, represented in percentage.

(5) Acquisition of traffic layer

Traffic layers are extracted from remote sensing image combined with land use map using Erdas and ArcGIS. We use a buffer representing the impacted range by roads to the city, with the buffer range of 120m for the main road, 60m for the secondary roads. We transform the layers into the GRID formatted file using a grid size of 30m×30m.

(6) Acquisition of map layers from the restricted construction land and the planned construction land

The restricted construction land belongs to the layer of macro-controlling factors. There are two ways for the data acquisition: 1) Water zones of remote sensing images; 2) The controlled construction land zone from town plan, such as some natural conservation areas, or large protected sites having special restrictions, scenic

areas, military sites, railway marshalling station and so on, which are not changed in the simulation. The same grid size of 30m×30m is used

The planned construction land is taken from the coordinated/calibrated city plan, including short-term and long-term plans Using planning data of 1976, 1993, and 2003 we get planning scale after registration and vectorization and convert it to the GRID formatted file. Again, the grid size is set to 30m×30m.

(7) Acquisition of initial land damage degree

The system divides damage degree into 5 levels, the lower the level, the smaller the damage. No damage is set as level 0. According to the statistical data of city damaged degree after disaster, we rate each cellular damaged degree, and finally change the dataset into a GRID format using the grid size of 30m×30m. Degree of damage is only effective in the initial phase of the simulation. When the damage degree of each cellular (grid) reaches the level 0, the implication of urban reconstruction is exactly the same as the one of urban development.

4 Design of Urban Reconstruction CA Model

4.1 Basic Cellular Automata (CA) Urban Model

CA systems are originally designed to study self-replication in natural sciences, originally as computable systems in general and then in fields such as biology and physics [2]. The models proposed by Tobler (1979) can be described as follows:

$$c_{xy}^{t+\Delta t} = F(c_{x\forall i, y\forall j}^{t}) \quad (1)$$

Where c_{xy}^{t} is the land use category such as urban and rural at the cell location x, y at time t, and $c_{xy}^{t+\Delta t}$ is the land use category at the same location in the future. The model (equation 1) suggests an application of a typical CA system where land use depends on the land use of its neighbors in previous time, here $x\forall i$ and $y\forall j$ represent the cell neighbors of x and y called i and j respectively [2].

The CA model has not simply remained as a new methodology. Linked to the complexity sciences, it has provided a much broader knowledge framework in which to understand urban systems in terms of interactions between their components, their spatial structure, and their temporal dynamics (Batty, 2005).

4.2 Model Data and Parameters

We extend basic CA model to include the geographical features. The goal of the extended CA model is logically to get the result on the urban land use conversion after reconstruction, through the use of CA dynamic evolution to simulate the urban renewal process.

(1)Input data

Initial data for the model to run is obtained by remote sensing images and statistics, and they are converted to the GRID format as the initial data for model running through the pre-processing and statistics. They are the source of various attribute values of cellular entity.

Initial data for the model running include: land use category, traffic, slope, non-construction land, planned construction land, the population of cellular regions, and damage degree of cellular regions.

(2) Model parameters

We set the following parameters for the model: desire (demanding), planning factors, population, slope and degree of damage. By changing the combination of the above parameters and running the model, we get a combination of the most realistic parameters.

1) Desire (demanding) parameter

In the natural growth rules, the desire parameter is to determine the possibility that a randomly selected cellular converts into urban cellular. For example, when a residential area reaches a certain scale, there will appear a commercial area nearby. The phenomenon, complying with the law of urban development, is called natural growth, also called desire growth.

Table 2. A table of evolutionary parameters that indicate the preference of various types of land to the nearby land unit use types

	The desire value	Neighbor cellular of all types of land use					
		Industrial land	Commercial land	Agricultural land	River	Street	Planned land
Central cellular	Industrial land	1	0.25	0.5	0.5	1	0.25
	Commercial land	0.25	1	0	0.75	1	0.25
	Agricultural land	0.5	0	1	0.75	0.5	0.25
	River	0.25	0.5	0.75	1	0.5	0.5
	Street	0.5	0.5	0.5	0.5	1	0.5
	Planned land	0.25	0.25	0.25	0.5	1	0.5

2) Planning factors affecting growth parameter

The planning factors that will impact the growth parameters determine the prior possibility of the cellular in the planned area becoming urban Cellular. The following are the factors table of various land use affected by the planning policy. It should be noted that the magnitude of the parameters values does not matter, only the comparison of them with other types of values matters

Table 3. Planning parameters table of all types of land use

Land use	Industrial land	Commercial land	Agricultural land	River	Street	Planned land
Planning factor	1.3	1.7	1	1	1.25	2

3) Slope parameter

Slope parameter will affect all the growth rules. We choose 35° for the gradient threshold, and the construction with the slope above this value is prohibited.

4) Parameters of the urban population and the damage degree

We set the lower bound of the population and upper bound of the damage to determine whether a cellular is capable of randomly generating another cellular. At start, the cellular with an excessive damage is considered as vacant land, and its neighbor cellular determines the outcome of its evolution.

4.3 Construction of the Model

(1) 5 steps to build a reconstruction CA model:

1) Determine the elements for the CA reconstruction model: which include cellular space, cellular state, cellular neighbors and time interval (step);

2) Analyze the data required for modeling the urban reconstruction which include seeds data and their attributes and macro controlling factors;

3) Determine the running parameters of CA model;

4) Set the transition rules for all kinds of cellular;

5) Select developing environment and developing model to implement the modeling software, and continuously debug and modify the running parameters, to make the modeling closer to the actual results.

(2) The composition of extended CA reconstruction model

1) Cellular space

The research area for the model is post-disaster urban and peri-urban regions, namely the cellular space. In this study, we define the size of each cellular space as 30m×30m.

2) Cellular state

We divide the urban land use into industrial use, commercial use, agricultural use, rivers, streets, and planned construction use of six types, and cellular's state is set to { 1,2,3,4,5,6 }.

3) Selection of the neighborhood

The model uses the extended Moore neighborhood configurations of cellular Space. Based on traffic and land destruction, we choose a radius of r=3, a total of $((2\times3+1)^2-1)$ cellular neighbors and then check the traffic factors and damage degree of all cellular, using 0 for no-through traffic roads, 1 for the ordinary-through traffic roads and 2 or above for main traffic roads through. We choose the cellular with the traffic factor of 1 or above and the damage level of 3 less as the neighbor of the central cellular.

4) Time step

It is a little bit complicated to determine the time step for the extended CA reconstruction model. We choose the initial data as the base year data in the model. After a lot of repeated trials and comparisons, we choose 3 months as the best time step in the model.

5) The transition rules of cellular

The transition rules of extended CA reconstruction model include: natural growth, growth along the traffic network, growth on planning factors and growth on central

point–wise spread. All growth patterns are affected by the factors of population and land destruction.

The city is a living system. The urban land units divided by certain rules are the cellular of urban life. A prerequisite for the continuation of urban living systems development is harmonious coexistence between the all living cellular, to ensure maximally meeting the desire of urban life cellular. The changes in desire of all types of cellular reflect the evolution of the land attributes during the reconstruction.

Let us denote the cellular desire as:

$$\mathrm{Demand_i} = \{\mathrm{Demand_{(i,1)}}, \mathrm{Demand_{(i,2)}}, \ldots\ldots \mathrm{Demand_{(i,M)}}\} \quad (2)$$

Where m stands for the total number of classified land use category, $\mathrm{Demand_{(i,1)}}$, $\mathrm{Demand_{(i,2)}}$, ……$\mathrm{Demand_{(i,M)}}$, respectively, represent the preference or need degree of type i cellular to type l~m of land use, whose value can be chosen as 0, 0.25, 0.5, 0.75, or l, which respectively denotes, disgusting, dislike,, do-not-care, like and love etc., those emotional factors,. Therefore, the desire of the neighbor cellular to the central cellular is:

$$\mathrm{Demand_{Avg}} = (\mathrm{Demand_{(i,1)}} + \mathrm{Demand_{(i,2)}} + \ldots + \mathrm{Demand_{(i,M)}}) / M \quad (3)$$

Integrating its own weighting model of macro and micro factors of the central cellular, the cellular evolution rules governed by the attributes are as follows

$$F: S_i^{t+1} = \mathrm{Demand_{Avg}} * F(A_i^t, S_i^t, S_n^t, A_n^t) \quad (4)$$

Where, t stands for time; i stands for cellular identity; S stands for cellular state; n stands for neighbor cellular; A stands for attributes.

4.4 Design of Model Framework

This model consists of three layers: 1) land use layer; 2) traffic layer; 3) the controlling factor layer. They are associated with each other through raster (grid) structure of the unified spatial resolution. The land use layer is at the core of the model, and is a CA model based on many types of land use. Interaction and dynamic change of different land use units create dynamic development of the city. The traffic layer is a layer of dynamic changes. The control factors layers are consisting of layers of spatial area with different attributes influence and control the behaviour of the above land use and traffic unit. Three layers are synchronized running in the unified time axis and, coupled with each other.

Under the unified framework of three layers, the model incorporated the external macro geographical model into micro CA model, and significantly improved the actual applicability of the model

According to the structure, the model can be denoted as:

$$CA = \{L, T, C\} \quad (5)$$

Where, L stands for the layer of the land use category; T stands for the traffic layer; C stands for the controlling factors layer. It is shown as in figure 2

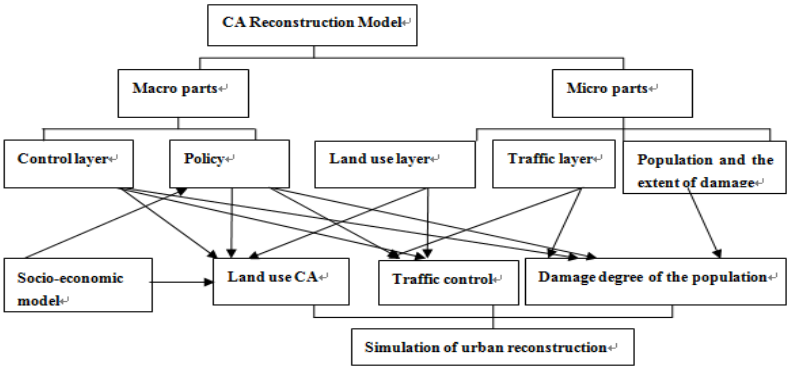


Fig. 2. The framework of CA urban reconstruction model

All the factors, which influence the urban disaster rehabilitation, will be input to cellular entity, and appear as attributes of the cellular entity. We fully consider the impact of macro and micro factors onto urban disaster, including micro factors such as neighbor selection and transition rule determination, and macro ones such as population size, economy, culture, and destruction degree, and use AHP method to determine the impact weights of all factors. We continually debug and rectify the running parameters, with adjustment and optimization of results, to make the modeling results closer to the actual ones.

4.5 Implementation of Modeling Software

The Operating System used for model development and execution is Microsoft Windows XP; the developing tools are Arc-objects and VC 6.

The procedure of the model application includes cellular initialization, data conversion, production of seeds, generation of controlling factors layers, specification of neighborhoods, and setup of status transition rules, determination of the model parameters, and simulation and analysis of the results of nine basic steps.

1) Cellular initialization. According to the spatial pattern and geographic elements of the study area, we determine the size and shape of cellular, cellular space, and set up descriptive indices of the cellular status.

2) The data conversion. Convert slope maps generated by maps and topographic maps after classification of remote sensing image, and other vector graphics into a grid format raster data.

3) The generation of seed points. In the model, the seed point is the initial state of urban development, and the urban growth is based on seeds. Therefore, the distribution of seeds should truly reflect urban spatial pattern after disaster.

4) Generation of control factor layers. Control factors are obtained through analysis of various data such as from general topography; land use, urban plans and so on.

5) Specification of the neighbors to the central cellular. According to the region in study, the simulation system determines neighbor's model for different geometric pattern and enables automatic search for the neighbors.

6) Creation of transition rules for the cellular state. The attribute value of cellular state is introduced into the CA model to determine the transition rules of the cellular state

7) Determination and modification of the model parameters. We choose different periods of urban land use status as a reference and testing data to optimize and rectify parameters. All parameters are manually input in the simulation system.

8) Simulation and output of model. Depending on the transition rules of cellular state, the modeling system calculates the next state of the cellular, and gives a visualized output according to its status values

9) Analysis of the results. The accuracy of the simulation results is evaluated, and analysis of the predicted (simulated) results let us to obtain a general trend of the city post-reconstruction.

4.6 Model Calibration

There are two purposes for model calibration: one is to determine its optimal parameters, i.e., using historical data to verify the model results through repeated testing, and finally to obtain its best running parameters; The second is to map the system time to real time, for the use of planning and decisions making. We use the method of verifying historical data, i.e.,, take the historical data at two or more different times, use the data at the first time as the initial state of the model, and compare the modeling data with the data at other time. If they match, the model parameters at this time are saved as optional parameters and further simulation continues.

The initial data are set to be those of 1976. The classified data for the land use map of 1993 and the TM remote sensing image of 2003 are used to verify the model simulation results. Using the future land use planning data as a control factors of urban development pace, the links are established between the simulating time and real time.

5 Model Predictions and Results

5.1 Simulation Process and Results

The assumptions we made for modeling urban renewal include: 1) the urbanized land in the reconstruction will not change its land use type; 2) the planned construction land data is used as a control layer, and input into the model for simulation

We use classified data of Tangshan land use of 1976 as initial simulation data, take 1993 as the first simulation period, and 2003 as the second simulation period, and compare the simulated data of 1993 and 2003 with the corresponding real data, and then we approximately determine the rectification scheme of the model parameters through analyzing and comparing the results, and re-run the model. Both the slope layer and the restrained construction land use layer in control factors are taken as a static layer and will not vary in the simulation.

For a given combination of parameters, after repeated runs, comparisons and adjustments, a valid set of running parameters is found for the model, and then using this set of parameters, we modeled the forthcoming urban development, and obtained a predictive analysis of the urban renewal. Simulation results are shown in the table 4.

Table 4. The table of Tangshan urban land use simulation results (number of grids) (base data of 1976)

Land use \year	1976	1993(modeled)	1993 (real)	2003 (modeled)	2003 (real)
Urban land	832	2536	3239	4277	4989
Rural land	89442	87918	83878	86687	82264
Rivers	6698	6520	6312	6014	6431

5.2 Analysis of the Results

According to the analyzed and simulated data, we can see the basic features of Tangshan urban renewal.

1) The industrial land use increased rapidly in the urban land.

2) A large scale of residential land use occurred in the urban land use

3) Tangshan city urban renewed very fast in the first decade, and then slowed down.

4) In terms of the actually extended profile in the city, through comparison with the existing data, the model reproduced the major trends in Tangshan urban renewal in general. There is a little difference between the simulated results with real data for 1993, but the difference for 2003 is larger in some region. The main reason is due to its protected state of the historical sites which belongs to the restricted construction zone in the model, so there is no change in the simulation result. There are two more reasons: One is the error in remote sensing image classification, another is the difficult handling of policy factors. As a whole, the model is able to deliver an acceptable and reasonable result.

It should point out that the model simulation is not the accurate prediction of urban development; rather it reflects the possible trend of urban development in the reconstruction. In fact, urban development is a very complex process with a great deal of uncertainty and chaos. Although there exist certain rules, it is very hard, even impossible, to give an absolutely correct prediction. The merit of the model is to help the geographer or city planners better understand the characteristics of the urban development.

6 Conclusions

Based on the features of city extension and reconstruction, we presented the principle and technical procedure of constructing CA reconstruction model. And further we used the built model to simulate and analyze the extended reconstruction of Tangshan

urban land use, and made a comparison with the city's status and overall planning results. Through this study, we can draw the following conclusions:

1) In this paper we analyzed and summarized the dynamical mechanism for urban development, the model of urban reconstruction and expansion, and simulation of complex behavior of urban expansion. This laid a foundation for establishing the dynamic model of reconstruction and expansion of urban.

2) Use of extended CA renewal model to simulate the variations in Tangshan urban land use could provide the basis for land use planning after the disaster, and help making the effective land management rules and policies. The simulation accuracy reached eighty percent and above, which showed the reasonability, and the applicability of the model.

3) From the simulation results, the model cannot fully make the accurate predictions of the extended reconstruction of the city, but could approximately provide possibilities of urban expansion and reconstruction. There are three reasons for it: 1) the data quality and reliability; 2) the transition rules playing a crucial role in the simulated result; 3) very difficult to accurately predict the variation in a developing city because of the natures of its chaos, self-organization and uncertainties, as well as other typical factors within a complex urban systems, in spite of existence of some rules.

References

1. Batty, M., Xie, Y., Sun, Z.: Modeling Urban Dynamics through GIS-Based Cellular Automata. J. Computers, Environments and Urban Systems 233, 205–233 (1999)
2. Kim, D., Batty, M.: Calibrating Cellular Automata Models for Simulating Urban Growth: Comparative Analysis of SLEUTH and Metronamica. UCL Working papers series: Paper 176 (2011), http://www.casa.ucl.ac.uk
3. Liu, M., Chen, P.: Model and Simulation of Urban Spatial Diffusion of Growth. J. Human Geography 19(2), 1139–1143 (2004)
4. Li, S., Zheng, D.: Application of Artificial neural network model in geological research progress. J. Advances in Earth Science 18(2), 68–76 (2003)
5. Cheng, Y., He, Z.: The Study on Geographic CA-Based Dynamic Model of Urban Evolution. In: 2011 International Conference on Software Engineering and Multimedia Communication, pp. 315–317. IERI Press, Qingdao (2011)
6. He, Z.: The Research on Model of City Post-disaster Reconstruction Based on Cellular Automata. Master's thesis (Supervised by Yun Cheng): China Three Gorges University (2009)
7. Xie, Y., Batty, M.: Integrated Urban Evolutionary Modeling. UCL Working papers series: Paper 68 (2003), http://www.casa.ucl.ac.uk
8. Xie, Y., Batty, M., Zhao, K.: Simulating Emergent Urban Form Desakota in China. UCL Working papers series: Paper 95 (2005), http://www.casa.ucl.ac.uk

Landscape Ecological Risk Assessment of the Shiyang River Basin

Xuebin Zhang[1,*], Peiji Shi[1,**], and Jun Luo[2]

[1] College of Geography and Environmental Science, Northwest Normal University, Lanzhou 730070, China
[2] College of Resources and Environmental Science, Gansu Agricultural University, Lanzhou 730070, China
{zhangxb428,journeylj6}@163.Com, shipj@nwnu.edu.cn

Abstract. It is significant to make a reasonable assessment of ecological risk to optimize the landscape pattern, establish the ecological risk alarm mechanisms, minimize the risk of ecological environment and maintain the ecological function in river basin. The study, based on the remote sensing data of 1987, 2000 and 2010, the study chooses the typical arid inland river basin – the Shiyang river basin as the subject to analyze the temporal-spatial distribution pattern of the ecological risk by constructing the ecological risk index with the help of the spatial analysis function of GIS. The results show that: (1)The main landscape types of the river basin are the grassland and unused land,but the predominance of the arable land, woodland and grassland decrease little by little during the study periods while the predominance of water and construction land increase gradually. (2)At the early stage of the study periods, the main ecological risk grades are extremely low and high. With the time going, three ecological risk models, extremely low, low, and extremely high, coexist in the study area, which mean that the threatens of the ecosystem is increasing. (3)The tendency of environmental deterioration is clear, which embodies in the space that extremely low ecological risk areas shrink to upstream, the low ecological risk areas spread to the upper and middle stream and the extremely high ecological risk areas expand to the downstream.

Keywords: landscape pattern, spatial analysis, transfer matrix, ecological risk index, Shiyang river basin.

1 Introduction

With the transformation of the environmental concept and management goals, the ecological risk assessment has become one of the hot issues of academic research both at home and abroad [1-3]. Regional ecological risk assessment is an important part of the ecological risk assessment, which is the description and evaluation of

* Foundation item: National Natural Science Foundation Project (40971078, 41271133). Author introduction: Xuebin Zhang, Male, doctoral student.
** Corresponding author.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 98–106, 2013.

likely damage of the ecosystem structure and function at regional scale [4]. River basin is a kind of complex natural geographical area, which is closely connected to certain areas of the integrated ecological system, provides a strong guarantee to biological survival and human activities [5]. With the continuous development of the society and the economy, natural and man-made risk source overlap within the basin, which make the watershed ecosystem suffer more and more from external stress [6]. Since the 1990's, scholars try to apply the ecological damage index method, the landscape ecological risk assessment model and the relative risk evaluation model [7-11] etc. to evaluate the water environment ecological risk [12], watershed ecological risk [13] and watershed integrated ecological risk [14]. Among the relevant researches, few are based on the landscape pattern to assess the ecological risk in arid inland river basin of China. Hunsaker et al. [15] published an article to illustrate how to apply ecological risk assessment to the regional landscape scales in the early 1990's. Therefore, the article is based on the landscape structure and land use information, chooses the Shiyang river basin as the subject to analyze the ecological risk of the study area by using the landscape ecology and spatial statistics analysis method and constructing ecological risk index to reveal the spatial and temporal variation characteristics of watershed ecological risk. The purpose is to maintain the ecosystem function of the Shiyang river basin, guarantee the ecological security of the Hexi corridor and provide quantitative and theoretical support to the scientific management of the ecosystem.

2 Study Areas

The Shiyang river basin is one of the three continental river basins in the Hexi corridor with the area of about 41,600 square kilometers of which the east is the Wushao mountain ridge, the west is Hexi corridor and the north is Qilian mountain ridge. The longitude is between 101°22'~104°16'N, and the latitude is between 36°29'~39°27'E, The terrain is higher in south and lower in north. The geomorphic unit can be divided into the Qilian Mountain, central plains corridor, low hilly land and desert. The water system of the Shiyang river basin originates in the Qilian Mountain and consists of eight rivers from west to east. The study area includes Wuwei city, Liangzhou district, Gulang County, Minqin County, Tianzhu County, Jinchuan district, Yongchang County and Su'nan County.

3 Methods

3.1 Data Sources and Processing

Data sources come from "the data center of the national natural science fund committee in western China", and the Landast TM images of 1987, 2000 and 2010. Referencing the current land use classification (GB/T21010-2007) and national remote sensing of land use/cover classification system and considering the characteristics of the Shiyang river basin, the land use types are divided into the farmland, woodland, grassland, construction land, water and unused land.

3.2 Risk Area

In order to illustrate the ecological risk, the article divides the study area into 20 × 20km risk area with the scope and workload in the study area in the consideration. The classified method is equally spaced sampling, which shape a total of 121 risk areas (Fig.1).

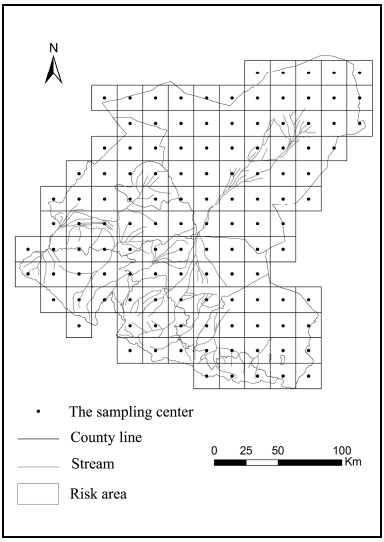


Fig. 1. The ecological risk evaluation cell on simple map of Shiyang river basin

3.3 Landscape Ecological Risk Index

Proceeding from the structure of landscape ecological system, the landscape disturbance degree index, the fragile index and the loss degree index are used to build the integrated ecological risk index (*ERI*) in the study to the analysis of watershed landscape ecological risk grade and changes. The formula is:

$$ERI_i = \sum_{i=1}^{N} \frac{A_{ki}}{A_k}(E_i \times F_i) \tag{1}$$

Where, ERI_i is the ith risk area's ecological risk index, A_{ki} is the ith landscape's area in the kth region. A_k is area of kth region, E_i is the landscape disturbance degree index of ith landscape, F_i is the fragile index of ith landscape. The formula of E_i is:

$$E_i = aC_i + bN_i + cD_i \tag{2}$$

Among which, $C_i = \frac{n_i}{A_i}$, $N_i = \frac{A}{2A_i}\sqrt{\frac{n_i}{A}}$, $D_i = \frac{(Q_i + M_i)}{4} + \frac{L_i}{2}$, Where, a, b, c represent the weights of landscape fragmentation, landscape isolation and landscape dominance index respectively, and a + b + c = 1. The indices reflect the interference of landscape,

which represents the influence of the ecological environment. According to the analysis, we consider that fragmentation index is the most important one, followed by separation index and dominance index, and the three indicators assign to the value of 0.5, 0.3 and 0.2 respectively, while to the unused land three indicators respectively assign to the value of 0.2, 0.3 and 0.5. n_i is the Number of patches ith landscape, A_i is the area of ith landscape, A is the area of all the landscape, Q_i is the sample number of ith pattern / total number of samples, M_i is the number of ith patches / total number of patches. L_i is the area of ith patches/ area of samples.

Landscape fragile degrees (F_i) represent the vulnerabilities of different internal structure of the ecological system, which can reflect the resistance ability of the external interference in different landscape types. When the resistance ability is smaller, the risk of the ecological system is greater. Reference to other studies [16], taking the actual situation of the study area into consideration and adopting Delphi to measure weights of the landscapes into six degrees from high to low, i.e. unused land, water, arable land, grassland, forest land, construction land. The fragile degrees (F_i) are acquired in the process of normalization.

3.4 Space Analysis Method

The regional ecological risk index, as a kind of typical regionalized variables, can be analyzed by the geostatistics method to show the law of heterogeneity. The calculation formula is:

$$\gamma(h)=\frac{1}{2N(h)}\sum_{i=1}^{N(h)}[Z(x_i)-Z(x_i+h)]^2,\quad (i=1,2,\cdots,N(h)) \tag{3}$$

Where, $\gamma(h)$ is the variation function, h is the step length, the matching of sampling space distance, $N(h)$ is the interval distance of the sample logarithm, $Z(x_i)$ and $Z(x_i+h)$ is the observed value of ecological risk index on the spatial location of x_i and x_i+h.

4 Results Analysis

4.1 Changes of Landscape Index

From 1987 to 2010, great changes have taken place in the Shiyang river basin (Fig.2), the farmland, forestland and grassland area have reduced 2.46 $\times 10^4$ hm^2, 4.77$\times 10^4$ hm^2 and 12.73$\times 10^4$ hm^2 respectively; Water and unused land area have increased by 2.58$\times 10^4$ hm^2 and 15.59$\times 10^4$$hm^2$. From 2000 to 2010, construction land has increased by 1.56$\times 10^4$ hm^2, accounting for 87.16% of the increment. Though the area is changing, the unused land and grassland, affected by the natural factors, are the main landscape types in the study area. Compared with other land types, the advantage is not obvious.

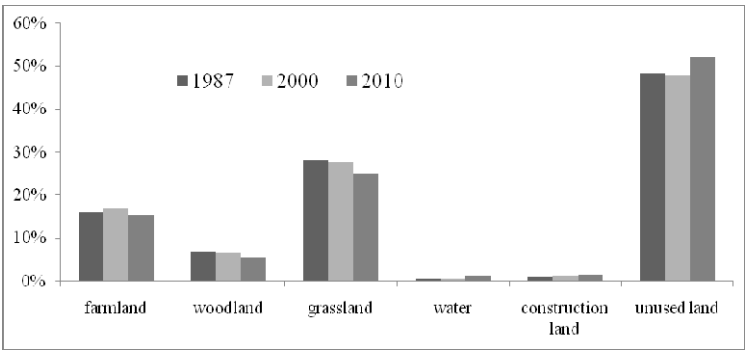


Fig. 2. Proportion of land use areas in the Shiyang river basin

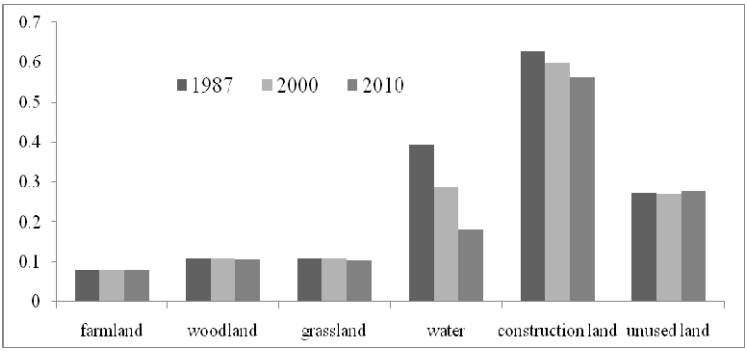


Fig. 3. Landscape disturbance degree index of Shiyang river basin

Separation index of the cultivated land, forest land, and grassland is increasing, which shows that the distribution of regional landscape pattern is more and more dispersed, the randomly scattered distribution phenomenon is rising; dominance index has been decreasing continuously, which shows that landscape advantage reduce little by little; separation index of water, construction land is decreasing, the dominance index increases gradually, which shows that landscape type focuses on regional distribution more and more, and with the speeding up of the urbanization, the dominance of residential areas increases, at the same time, other green ecosystem function is declining. Unused area is increasing, which means that the fragmentation and isolation index decreases continuously, the dominance of unused land is increasing, which indicates that the centralized tendency of unused land has been increased and the ecological environment get worse. By overlying fragmentation, isolation and dominance index, the landscape disturbance index can get (Fig.3),among which watershed is very obvious and the index drops from 0.6259 down to 0.5607 from 1987 to 2010. The Interference index of construction land is the largest, which is mainly caused by the higher fragmentation and isolation of construction land.

From the viewpoint of time, there's a little change in the landscape pattern indices from 1987 to 2000, but big changes happens in the phase of 2000 to 2010.

The cultivated land, forest land, grassland fragmentation and isolation index increase significantly, which indicates that the advancement of industrialization and urbanization, the phenomena of construction taking up land have become the important factors that affect the landscape pattern changes.

4.2 Time and Space Analysis of the Ecological Risk

According to the formula (1), the landscape ecological risk index (*ERI*) is calculated for each risk area: in 1987, the index value is between 0.1222 and 0.3499, in 2000 the value is between 0.1222 and 0.3471, and in 2010, the value is between 0.1255 and 0.3789. According to the range of the value, the ecological risk can be divided into five grades by using natural break point method as well as with the help of ArcGIS,: If $0.12 \leq ERI \leq 0.17$, then the ecological risk grade is extremely low; and if $0.17 < ERI \leq 0.22$, the ecological risk grade is low; and if $0.22 < ERI \leq 0.27$, then the ecological risk grade is medium; and if $0.27 < ERI \leq 0.32$, then the ecological risk grade is high, or the ecological risk grade is extremely high. Using the spatial analysis methods of ArcGIS, the experimental half variant function value is calculated, which is based on the ecological risk area, and the spherical model is found to be the best by comparing with different models. We can get the ecological risk grades figure through kriging interpolation method, and the area is showing in the table 1.

Table 1. The area of ecological risk grade ($10^4 hm^2$,%)

ecological risk grade	1987		2000		2010		changes of 1987-2010
	area	proportion	area	proportion	area	proportion	
extremely low	124.88	30.77	124.66	30.72	92.99	22.91	-31.89
low	62.35	15.37	64.39	15.87	91.65	22.58	29.30
medium	81.61	20.12	64.64	15.93	60.93	15.02	-20.68
high	132.02	32.53	144.21	35.54	96.60	23.81	-35.42
extremely high	4.93	1.21	7.89	1.94	63.62	15.68	58.69

From 1987 to 2010, the high ecological risk area expands unceasingly, and the area increases by 58.69×10^4 hm^2, an annual increase of 4.51×10^4 hm^2, and the increase speed is 7.69%, the proportion of high ecological risk area rises from 1.21% in 1987 to 15.68% in 2010. The low ecological risk area is decreasing from 124.88×10^4 hm^2 in 1987 to 92.99×10^4 hm^2 in 2010, reduces by 31.89×10^4 hm^2, the proportion reduces from 30.78% to 22.92%. Overall, the area change is different in the basin, in 1987 and 2000, the main risk grades are extremely low and high, and the extremely high ecological risk area is small. In 2010, extremely low ecological risk area significantly reduces, but low ecological risk area largely increases; high ecological risk area reduces significantly, but the extremely high ecological risk area increases largely, extremely low, low and extremely high ecological risk coexist in the region.

The ecological risk area changes mainly concentrate in the phase of 2000 to 2010, from 1987 to 2000, the area of extremely high ecological risk only increases 2.96×10^4 hm^2, which increases 55.73×10^4 hm^2 from 2000 to 2010, accounted for 95% of the total increments; All the phenomena show that the ecological environment in the Shiyang river basin changes greatly after 2000, which is mainly caused by the social economy development and the acceleration of urbanization, a large number of agricultural land, especially farmland are occupied by construction, water use in social and economic development crowd out that in natural ecology, ecological system balance is under strong interference. All these lead to the ecological environment becoming worse and worse and the chance of ecological risk is relatively high (Fig.4).

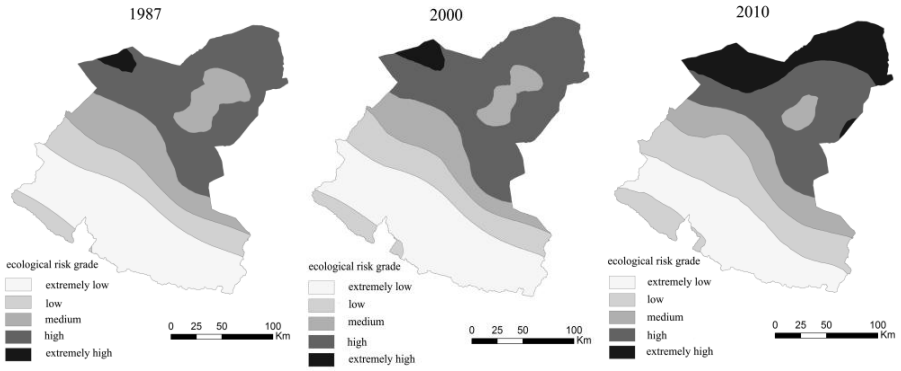


Fig. 4. Ecological risk grade map of Shiyang river basin

The spatial distributions of ecological risk show that: from southwest to northeast the grade is higher and higher, the ecological risk increases gradually, namely from the upstream to the downstream of the basin, the threaten to ecological environment becomes severe, which is mainly due to the widely distributed woodland and grassland in the upstream, a low ecological risk grade, a large number of unused land distributed in the downstream, and with little biomass, the food chain structure is simple, ecological system is fragile, which lead to the high ecological risk degree. With time going, extremely low ecological risk areas shrink to upstream, the low ecological risk areas spread to the upper and middle stream, and the extremely high ecological risk areas expand to the downstream, which indicate that with the growth and concentration of population in oasis, human activities have become frequent, which impact on ecological environment, impel the ecological risk degree increase.

Overall, the Shiyang river basin is part of the continental temperate arid climate, with strong solar radiation, evaporation, less rainfall, but dry air, which is a large decisive factor to the fragile and sensitive ecological environment of the basin, and with relatively low carrying capacity of water resources. Since the 1980s, population increase rapidly in the river basin, so the demand for the cultivated land sharply increase, the soil and land resource are excessively exploited and utilized, human activities, as a superposition of external force on the natural factors, exacerbate the deterioration of ecological environment.

5 Conclusion and Discussion

Based on the remote sensing image data, with the support of GIS technology, the ecological risk is evaluated in the Shiyang river basin, the conclusions are as follows:

(1) The main landscape types of the basin are grassland and unused land, great changes have taken place in the study period, that is, the farmland, forestland and grassland area reduce, while the construction land, water and unused land area increase.

(2) The dominance of the farmland, forestland and grassland gradually reduced while the water and construction land increase gradually, all the changes mainly occur in the phase of 2000 to 2010.

(3) During 1987 and 2000, the ecological risk grade is given priority to extremely low and high, while in 2010, extremely low, low and high ecological risk coexist, the threaten to the ecological environment is gradually increasing.

(4) On space, the extremely low ecological risk areas shrink to upstream, and the low ecological risk areas spread to the upper and middle stream, the extremely high ecological risk areas expand to the downstream; the ecological environment is obvious deteriorated.

The adoption of the interference degree index, fragile degree index and loss degree index is to build the ecological risk index objectively and to reflect the status of ecological risk in the study area. Meanwhile, with the help of the geostatistics of ArcGIS, the spatial and temporal variations of ecological risk reveal in the river basin, which is in conformity with the condition of ecological environment of the basin. In a relatively short time scale, landscape pattern can reflect the human influence on the ecological environment, so it is feasible and reasonable to study the ecological risk change in the arid watershed scale from the perspective of the landscape pattern. We can optimize land use ways and provide the basis for the ecological construction through judgment and recognition of risk grades of different risk areas.

Acknowledgments. This work was financially supported by National Natural Science Foundation Project (40971078, 41271133).

References

1. Xu, Y., Ma, M., Gao, J.: The evaluation method of ecological risk assessment based on watershed scale—Take the Taihu watershed as example. China Environmental Science 32(9), 1693–1701 (2012)
2. Glenn, W., Suter II, G.W.: Endpoints for regional ecological risk assessment. Environmental Management 14(1), 9–23 (1990)
3. Cormier, S.M., Smith, M., Norton, S., et al.: Assessing ecological risk in watersheds: a case study of problem formulation in the Big Darby Creek Watershed, Ohio, USA. Environmental Toxicology and Chemistry 19(4), 1082–1096 (2000)
4. Fu, Z., Xu, X., Lin, H., et al.: Regional ecological risk assessment of in the Liaohe River Delta wetlands. Acta Ecologica Sinica 21(3), 365–373 (2001)

5. Cai, H., Zhang, X., Huang, H.: The Theories and Methods of Lake-watershed Land Ecological Management. Journal of Natural Resources 25(6), 1049–1058 (2010)
6. Song, C., Yang, G., Leng, S., et al.: Prospect and research of lake and watershed science. Journal of Lake Sciences 14(4), 289–300 (2002)
7. Lu, H., Zeng, G., Xie, G., et al.: The regional ecological risk assessment of the Dongting Lake watershed. Acta Ecologica Sinica 23(12), 2520–2530 (2003)
8. Landis, W.G., Kelly, L.: Regional scale ecological risk assessment: using the relative risk model. CRC Press, Boca Raton (2005)
9. Zimmermann, P., Tasser, E., Leitinger, G., et al.: Effects of land-use and land-cover pattern on landscape-scale biodiversity in the European Alps. Agriculture, Ecosystems and Environment 139(1-2), 13–22 (2010)
10. Echeverría, C., Newton, A., Nahuelhual, L., et al.: How landscapes change: integration of spatial patterns and human processes in temperate landscapes of southern Chile. Applied Geography 32(2), 822–831 (2012)
11. Colnar, A.M., Landis, W.G.: Conceptual model development for invasive species and a regional risk assessment case study: the European green crab, Carcinus maenas, at Cherry Point, Washington, USA. Human and Ecological Risk Assessment 13(1), 120–155 (2007)
12. Wang, X., Liu, J., Ma, M., et al.: Aquatic ecological risk assessment and management strategies in a watershed: An overview. Acta Scientiae Circumstantiae 30(2), 237–245 (2010)
13. Li, H., Cai, Y.: Ecological risk assessment of flood disaster in major cities in Taihu basin. Journal of Catastrophology 17(3), 91–96 (2002)
14. Chen, P., Pan, X.: Ecological risk analysis of regional landscape in inland river watershed of arid area—a case study of Sangong River Basin in Fukang. Chinese Journal of Ecology 22(4), 116–120 (2003)
15. Hunsaker, C.T., Graham, R.L., Suter II, G.W., et al.: Assessing Ecological Risk on a Regional Scale. Environmental Management 14(3), 325–332 (1990)
16. Gao, B., Li, X., Li, Z., et al.: Assessment of ecological risk of coastal economic developing zone in Jinzhou Bay based on landscape pattern. Acta Ecologica Sinica 31(12), 3441–3450 (2011)

Research on the Prediction of the Geological Spatial Information Using Gray GIS Modeling Method Based on the Borehole Data and the Geologic Map

Jun Tao, Xu Liu, Jinli Huang, and Bohu Yu

Northwest Institute of Nuclear Technology
Xi'an 710024, Shanxi, China
taojun100@sina.cn

Abstract. The geologic spatial information modeling is an important research area in the 3D-GIS field. The paper adopts the spatial raster data to represent the geologic spatial information, and the form of the information is the serial slices of plan raster matrix along the altitude direction. The paper regards the vertical borehole data and the plan geologic map as the basic geographical information, and utilizes the gray GIS theory to build the probabilistic modeling for the geologic spatial information, and applies the inverse distance weighted method to predict the geologic information, and establishes the plan geological section grid matrix for different altitude according to the maximum probability. The experimental results show that the gray GIS modeling method adapts to the complex conditions, and can construct more reasonable geological model with more basic geographical information.

Keywords: Geological Spatial Information, Borehole Data, Geologic Map, Gray GIS, Data Prediction.

1 Introduction

In the scientific research area of the 3D-GIS, the geological spatial information modeling is an important analysis technique combined with geoscience statistics, geological interpretation and spatial prediction[1]. Through the establishment of the geological spatial information model, the image of the geological formations can be demonstrated in front of the experts and the geological structure analysis can be more accurate and more intuitive. Currently, using various types of borehole data and geological maps for spatial information modeling is a research focus. Literature [2, 3] proposed to adopt an overall borehole data fitting method, and to characterize the geological spatial information by creation the DEM of different stratum interface. Literature [4, 5] proposed to generate customized geological section according to the geological information, and to establish the geological boundary representation model (B-Rep). Literature [6, 7, 8] proposed the use of borehole data to build 3D geological body directly through the geological solid model of the Tetrahedron Net (TEN), Triangular Prisms (TP), General Triangular Prisms (GTP), etc. Literature [9, 10]

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 107–115, 2013.

proposed the introduction of virtual borehole to improve spatial data density, and to achieve the capabilities for complex 3D geological modeling and analysis. However, the above geological spatial information modeling methods are still insufficient: Firstly, in the borehole scarce regions and geological unit boundaries, modeling results and the actual distribution of the strata may be big difference and difficult to modify. Secondly, as the stratum is both randomness and structural, there is lack of effective evaluation of the accuracy of the geological spatial information generated by data processing.

In this paper, a novel geological spatial information modeling method is proposed based on the gray GIS theory[11]: the paper characterizes the geological information by the spatial raster data, and quantifies the borehole data and geological maps as the basic geographic information, and uses the gray GIS theory to build probabilistic modeling of the geological attributes, and predicts the probability distributions by the gray information processing. Experiments show that the gray GIS modeling method adapt to the complex situation, and can effectively reflect the presumed accuracy of geological spatial information.

2 The Gray GIS Model of the Geological Spatial Information

2.1 The Basic Concept of the Gray GIS Model

In the 3D geological space, assume that the geological information of the study area actually exists. Then according to the obtained knowledge of the geological information within the study area, the GIS model can be divided into three categories: (1) the black GIS model: There is no known geological information within the study area (as shown in Figure 1(a)). (2) The gray GIS model: The geological information in parts of the study area is known, and the other geological information is some of the inference and prediction (as shown in Figure 1(b)). (3) The white GIS model: all of the geological information within the study area is known (as shown in Figure 1(c)).

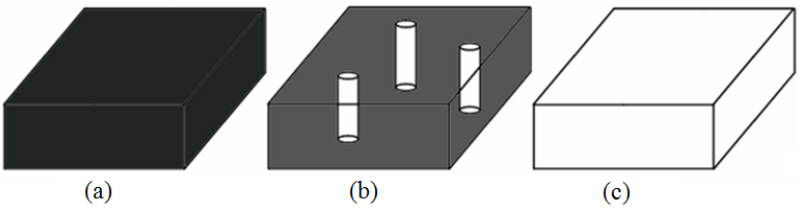


Fig. 1. The basic concept of the black, gray and white GIS model

Currently, widely used GIS is based on the white GIS model[11], and the gray GIS model is comparatively complex, and the basic concept of the gray GIS model is: For the 3D geological spatial entities, which the real world actually exists but not exactly

known, the model uses the non-deterministic way to describe the properties of geological information space, and dynamically corrects and updates the geological information to make the system close to the white GIS system. In this paper, the gray GIS model uses probabilistic description.

2.2 The Spatial Information Representation of the Gray GIS Model

The geological spatial information can be expressed as 3D-GIS spatial raster data. The raster data is the grid matrix sequence along the altitude direction, each of the plan grid $A_{l\times m}$ matrix reflects the plan geological section information for a determined altitude. The X-axis direction coordinate sequence is (x1, x2, x3,…, x*l*), which spacing value is Δx. The Y-axis direction coordinate sequence is (y1, y2, y3,…, y*m*), which spacing is value Δy. The geological spatial information Q is a group grid matrix: $Q=\{A_{z1}, A_{z2}, A_{z3},\ldots, A_{zn}\}$, where z1, z2, z3,…, z*n* is of n-pitch altitude sequence, which spacing value is Δz.

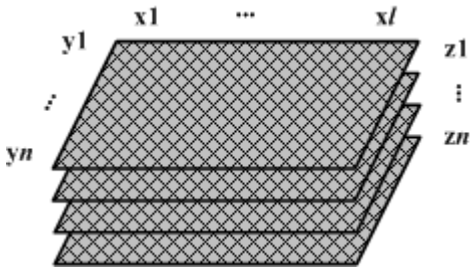


Fig. 2. The spatial raster data representation of the geological spatial information

The geological information can be divided into *s* class in accordance with the type of geological time. So the information matrix of the plan geological section A_k (k=z1, z2, z3,…, z*n*) is expressed as:

$$A_k = \begin{bmatrix} P_{1,1,k} & P_{2,1,k} & \ldots & P_{l,1,k} \\ P_{1,2,k} & P_{2,2,k} & \ldots & P_{l,2,k} \\ \ldots & \ldots & \ldots & \ldots \\ P_{1,m,k} & P_{2,m,k} & \ldots & P_{l,m,k} \end{bmatrix}_{l\times m} \tag{1}$$

In formula (1), $P_{i,j,k} = [P^1_{i,j,k}, P^2_{i,j,k}, \ldots, P^s_{i,j,k}, P^{N/A}_{i,j,k}]$ is a probability vector, and $P^t_{i,j,k}$ (t=1,2, ..., s) indicates the probability of the geological information for the *t* class geological type, $P^{N/A}_{i,j,k}$ represents a non-geological information (for example: above the terrain space). They satisfy the following formula:

$$P^1_{i,j,k} + P^2_{i,j,k} + \ldots + P^s_{i,j,k} + P^{N/A}_{i,j,k} = 1 \tag{2}$$

2.3 The Spatial Information Evaluation of the Gray GIS Model

From formula (1) and (2) can be seen that the probability of geological spatial information can be divided into three categories: white information, gray information and black information.

Define 1: If the geological information $P_{i,j,k} = [P^1_{i,j,k}, P^2_{i,j,k}, ..., P^s_{i,j,k}, P^{N/A}_{i,j,k}]$ exists $P^t_{i,j,k} = 1$, then the geological information $P_{i,j,k}$ is called the white information (WINF).

Define 2: If the geological information $P_{i,j,k} = [P^1_{i,j,k}, P^2_{i,j,k}, ..., P^s_{i,j,k}, P^{N/A}_{i,j,k}]$ does not exist $P^t_{i,j,k} = 1$, then the geological information $P_{i,j,k}$ is called the gray information (GINF).

Define 3: If the geological information $P_{i,j,k} = [P^1_{i,j,k}, P^2_{i,j,k}, ..., P^s_{i,j,k}, P^{N/A}_{i,j,k}]$ satisfies $P^1_{i,j,k} = P^2_{i,j,k} = ... = P^s_{i,j,k} = P^{N/A}_{i,j,k} = 1/(s+1)$, then the geological information $P_{i,j,k}$ is called the black information (BINF).

The WINF is known geological spatial information, such as: the borehole data and geological map within the the study area. The BINF is totally unknown geological spatial information. All possible geological type has the same probability. The GINF is part of the unknown, and between the BINF and the WINF.

Characteristic 1: The WINF remains unchanged.

Since the WINF is known geological information, the gray GIS will always keep it unchanged during the data prediction process. If the geological spatial information of the study area is all WINF, there is no need of data prediction.

Characteristic 2: The WINF affects the GINF.

Since the GINF is part of the unknown geological information, so as the study area increasing or changing the WINF, the GINF will also change. If the geological spatial information of the study area is all GINF, there is no way of data prediction.

According to Characteristic 1 and Characteristic 2, the evaluation mechanism for the gray GIS is established based on information entropy, the evaluation index $E_{i,j,k}$ of the geological information $P_{i,j,k}$ is:

$$E_{i,j,k} = \sum_{t=1}^{s} P^t_{i,j,k} \times \log(1/P^t_{i,j,k}) + P^{N/A}_{i,j,k} \times \log(1/P^{N/A}_{i,j,k}) \tag{3}$$

In the study area, the geological spatial information evaluation formula is:

$$E_Q = \frac{\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} E_{i,j,k}}{l \times m \times n} \tag{4}$$

3 The Borehole Data and the Geologic Map in the Gray GIS

3.1 The Initial Geological Spatial Information

Without considering the information of the borehole data and the geological map, the initial geological spatial information in the study area is completely unknowable. Then the initial geological spatial information Q_{ini} is full of the BINF, and for any geological information $P_{i,j,k}$ there is:

$$\mathbf{A}_{z1} = \mathbf{A}_{z2} = \ldots = \mathbf{A}_{zn} = \begin{bmatrix} \mathbf{BINF} & \mathbf{BINF} & \ldots & \mathbf{BINF} \\ \mathbf{BINF} & \mathbf{BINF} & \ldots & \mathbf{BINF} \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{BINF} & \mathbf{BINF} & \ldots & \mathbf{BINF} \end{bmatrix}_{l \times m} \tag{5}$$

Then, the initial geological spatial information Q_{ini} represents a completely black space. All possible geological types appear with equal probability in the data matrix. Using the formula (3) and (4), the initial geological spatial information's evaluation value is:

$$E_Q = \frac{\sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} \log(s+1)}{l \times m \times n} = \log(s+1) \tag{6}$$

3.2 The Geological Spatial Information Correction

Since the vertical borehole data may reflect the local area underground geological information, so for each borehole $B(X_B,Y_B,Z_B,H_B)$ there has:

$$\begin{cases} \mathbf{x_a} - \Delta x/2 \le \mathbf{X_B} \le \mathbf{x_a} + \Delta x/2 \\ \mathbf{y_b} - \Delta y/2 \le \mathbf{Y_B} \le \mathbf{y_b} + \Delta y/2 \end{cases} \tag{7}$$

$$\mathbf{Z_B} \ge \mathbf{z_c} \ge \mathbf{z_{c+1}} \ge \ldots \ge \mathbf{z_{c+d}} \ge \mathbf{Z_B} - \mathbf{H_B} \tag{8}$$

In formula (7) and (8), the (X_B,Y_B,Z_B) is the borehole location, and the H_B is the borehole depth. Then the geological information $P_{a,b,c}$, $P_{a,b,c+1}$,…, $P_{a,b,c+d}$ are corrected into the WINF.

Since the plan geological map may reflect the global geological information and elevation information, so for each location in the plan geological map $M(X_M,Y_M,Z_M)$ there has:

$$\begin{cases} \mathbf{x_a} - \Delta x/2 \le \mathbf{X_M} \le \mathbf{x_a} + \Delta x/2 \\ \mathbf{y_b} - \Delta y/2 \le \mathbf{Y_M} \le \mathbf{y_b} + \Delta y/2 \end{cases} \tag{9}$$

$$\mathbf{z_1} \ge \mathbf{z_2} \ge \ldots \ge \mathbf{z_{c-1}} \ge \mathbf{Z_M} \ge \mathbf{z_c} \tag{10}$$

In formula (9) and (10), the (X_M,Y_M) is the map coordinates, and the Z_M is the surface elevation. Then the geological information $P_{a,b,z1}$, $P_{a,b,z2}$,…$P_{a,b,c-2}$, $P_{a,b,c-1}$ are corrected into the WINF.

Through the geological spatial information correction with the borehole data and the geological map, the corrected geological spatial information's evaluation value is:

$$E_Q{}' = \frac{(l \times m \times n - N_{WINF}) \log(s+1)}{l \times m \times n} \tag{11}$$

In formula (11), the N_{WINF} is the number of the corrected geological information.

4 The Data Prediction Method in the Gray GIS

4.1 Define the Distance Formula

Generally, the geological information has the correlated relationship in space. So the paper defines the distance formula from (i1,j1,k1) to (i2,j2,k2) about the geological spatial information of the *t* class geological type:

$$DIS(P^t_{i1,j1,k1}, P^t_{i2,j2,k2}) = \begin{cases} \sqrt{(i1-i2)^2 + (j1-j2)^2 + (k1-k2)^2} & P^t_{i2,j2,k2} > 0 \\ \infty & P^t_{i2,j2,k2} = 0 \end{cases} \tag{12}$$

And the paper defines the distance formula from point (i,j,k) to borehole data $B(X_B,Y_B,Z_B,H_B)$ about the geological spatial information of the *t* class geological type:

$$DIS_B(P^t_{i,j,k}, B^t) = \min_{h=Z_B}^{H_B}(DIS(P^t_{i,j,k}, P^t_{X_B,Y_B,h})) \tag{13}$$

4.2 The Inverse Distance Weighted Data Prediction

The paper uses the inverse distance weighted method to predict the GINF in the geological space. Considering the borehole data, the prediction formula is:

$$PRE(P^t_{i,j,k}, B^t) = \begin{cases} 1/DIS(P^t_{i,j,k}, B^t) & DIS(P^t_{i,j,k}, B^t) \neq \infty \\ 0 & DIS(P^t_{i,j,k}, B^t) = \infty \end{cases} \tag{14}$$

Considering the geological map, when predicting the geological section of altitude Z_k, only using small region information of the geological section of altitude Z_{k-1}, the prediction formula is:

$$PRE(P^t_{i,j,k}, M^t) = 1/DIS(P^t_{i,j,k}, A^t_{k-1}) \qquad DIS(P^t_{i,j,k}, A^t_{k-1}) \leq D_{Region} \tag{15}$$

In formula (15), A^t_{k-1} is the information matrix of the probability of the *t* class geological type at altitude Z_{k-1}, D_{Regin} is the distance limitation for prediction.

Then combining the borehole data and the geological map, the final normalized prediction formula of the GINF is:

$$\begin{aligned} &P^{t}_{i,j,k}(PRE) = P^{t}_{i,j,k} + C_B \times PRE(P^{t}_{i,j,k}, B^{t}) + C_M \times PRE(P^{t}_{i,j,k}, M^{t}) \\ &\overline{P^{t}_{i,j,k}(PRE)} = P^{t}_{i,j,k}(PRE) \Big/ \sum_{d=1}^{s} P^{d}_{i,j,k}(PRE) \end{aligned} \tag{16}$$

In formula (15), C_B and C_M are the coefficients for the prediction of the borehole data and the geological map, and the geological types will be determined according to the maximum probability.

5 Experiments and Results

In the experiments, the study area is a 10000m×10000m×200m geological space, and has the same surface elevation. According to the gray GIS modeling method, the study area is built into a 200×200×40 spatial raster data set (Δx=Δy=50m, Δz=5m). The study area has three classes of geological types, and the vertical borehole data and the plan geological map is shown in Table 1 and Fig. 3.

Table 1. The information of the borehole data

Borehole NO.	X_B	Y_B	TYPE 1	TYPE 2	TYPE 3
1# Borehole	1000m	1000m	-	0m-200m	-
2# Borehole	4000m	8000m	0m-200m	-	-
3# Borehole	8000m	5000m	-	-	0m-200m
4# Borehole	2000m	2500m	0m-75m	75m-200m	-
5# Borehole	6000m	5000m	0m-100m	-	100m-200m
6# Borehole	6200m	1000m	0m-25m	-	25m-200m
7# Borehole	7500m	9000m	0m-30m	-	30m-200m
8# Borehole	3000m	1500m	0m-50m	50m-200m	-
9# Borehole	9000m	1000m	-	-	0m-200m
10# Borehole	5500m	1000m	0m-30m	-	30m-200m

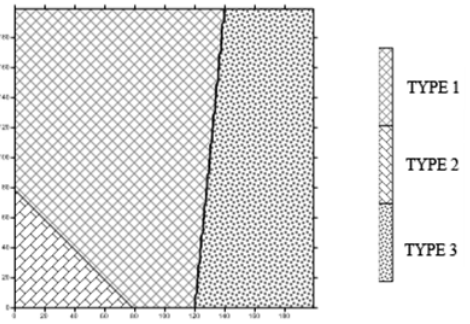


Fig. 3. The information of the geological map

Using the proposed data prediction method, set D_{Regin}=500m and C_B=C_M=0.5, the spatial raster data of the geological space can be constructed. The paper uses the VTK software[12] to volume-render the experimental results.

Experiment 1: Prediction base on the 1#-5# borehole and the geological map.

The predicted geological spatial information is shown in Fig. 4 (The geological body of TYPE 1 and TYPE 3), and the information's evaluation value is 0.4650.

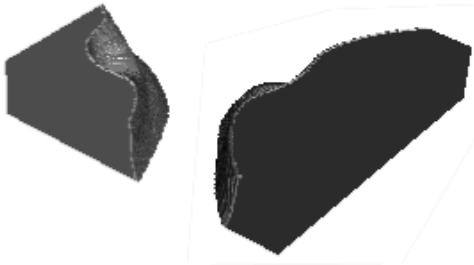


Fig. 4. The volume rendered result of experiment 1

Experiment 2: Prediction base on the 1#-7# borehole and the geological map.

The predicted geological spatial information is shown in Fig. 5 (The geological body of TYPE 1 and TYPE 3), and the information's evaluation value is 0.4521.

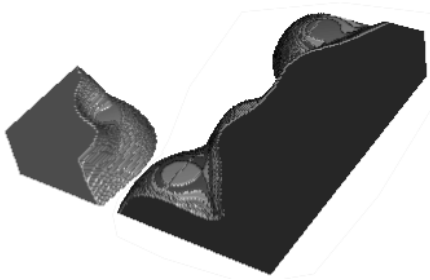


Fig. 5. The volume rendered result of experiment 2

Experiment 3: Prediction base on the 1#-10# borehole and the geological map.

The predicted geological spatial information is shown in Fig. 6 (The geological body of TYPE 1 and TYPE 3), and the information's evaluation value is 0.4448.

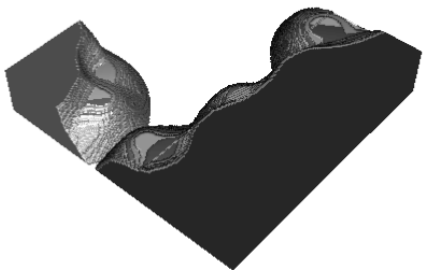


Fig. 6. The volume rendered result of experiment 3

From the results of the experiments, it's known that the proposed gray GIS model is practicable, and the prediction method can construct geological spatial information more reasonable when borehole data increasing.

6 Conclusion

In this paper, a gray GIS modeling, evaluation and prediction method is proposed, and the experimental results show that the geological spatial information can be more dependable using the proposed method when more WINF is given. But the proposed method depends very much on the distance between the WINF and the GINF, so the future work is to research a better data prediction method.

References

1. Zheng, G., Shen, Y.: 3D Analyses of Geological Characteristics and Status Research of 3D Geology Modeling. Advance in Earth Sciences 19(2), 218–223 (2004)
2. Cheng, P., Liu, S., Wang, W., Chen, H.: Study and application of a 3D geological model construction method. Journal of Jilin University (Earth Science Edition) 34(2), 309–313 (2004)
3. Zhang, Q., Xu, S., Yu, W., Cheng, H.: Application of particle swarm optimization to Kriging three dimensional geological modeling. Journal of Daqing Petroleum Institute 35(1), 85–89 (2011)
4. Tian, Y., Yuan, Y., Li, S., Wu, C.: Interpolation method for establishing 3d tectonic-stratigraphic basin framework. Journal of China University of Geosciences (Earth Science) 25(2), 191–194 (2000)
5. Lemon, A.M., Jones, N.L.: Building solid models from boreholes and user-defined cross-sections. Computer and Geosciences 29, 554–555 (2003)
6. Liu, Y., Song, Z., Niu, W.: 3D GIS Data Model Based on TEN and Its Generation Algorithm. Computer Application 24(7), 153–158 (2004)
7. Rui, X., Yang, Y., Xi, Y.: Study into Visualization of 3D Stratum Based on Triangular Prism. Journal of China University of Mining & Technology 33(5), 584–588 (2004)
8. Penggen, C., Jianya, G., Wenzhong, S.: Geological Object Modeling Based on Quasi Tri-prism Volume and Its Application. Geomatics and Information Science of Wuhan University 29(7), 602–607 (2004)
9. Wang, R., Li, Y., Liu, Y., Xiang, Z.: Import and determination methods for virtual borehole in geo-3d modeling. Geology and Prospecting 43(3), 102–107 (2007)
10. Tian, Y., Wu, C., Dong, Z.: Auto-joining of Strata and Geological Body Model Based on Introduced Virtual Borehole. Geological Science and Technology Information 29(5), 117–120 (2010)
11. Mao, S.: Gray Geographical Information System: The Theory and Technology of Correct Geological Spatial Data Dynamically. Universitatis Pekinensis 38(4), 556–562 (2002)
12. Visualization Toolkit (VTK), `http://www.vtk.org/`

Study of Ecological Security Changes in Dongjiang Watershed Based on Remote Sensing

Kaiwen Zhong, Caige Sun, and Kekui Ding

Key Laboratory of Guangdong for Utilization of Remote Sensing and Geographic Information System, Guangzhou Institute of Geography, Guangzhou, GD510070, People's Republic of China

Abstract. In this paper, the ecological security (eco-security) of Dongjiang watershed in 1988, 1998 and 2007 has been evaluated and the eco-security in 2016 was predicted. The Pressure-State-Response model was used to establish the eco-security index system. The temporal and spatial changes of the eco-security from 1988 to 2007 were detected. At the same time, the eco-security in 2016 was predicted temporally and spatially by using The Markov chain model and the Cellular Automata model (the CA_Markov model). The results show that the eco-security of Dongjiang watershed has obvious space differences. The innovation point is that the relative change rate and CA_Markov model were applied to the study of eco-security.

Keywords: Ecological security, Relative change rate, Markov chain, Cellular Automata model.

1 Introduction

The eco-security has a profound impact on economy, society and environment. Up to now, many scholars put their attentions to the eco-security. Their research mainly concentrates on the narration of the ecological security problems [1], the discussion of the ecological security concept [2], the index and standard selection of the ecological security evaluation [3] and so on. In the applications, many studies aim at the regions, such as western China [4] [5], ecotone between agriculture and animal husbandry [6], arid area oasis [7], loess hilly region [8] and Tibetan plateau [9]. While there are few concerning the ecological security research on small watershed in southern China. On the time scale, many studies research on the current ecological security, few concentrate the comparative study of long time interval so as to reflect the dynamic changes.

Based on remote sensing (RS) and geographical information system (GIS), the eco-security of Dongjiang watershed in 1988, 1998 and 2007 has been evaluated and the temporal and spatial changes were analyzed. At the same time, the change matrix was built to obtain the changing quantity and rate of each level. The eco-security in 2016 was predicted temporally and spatially by using the CA_Markov model.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 116–124, 2013.

2 Study Area

The area of this study is Dongjiang watershed (Fig.1), which located from 113°29′E to 115°41′E,22°23′N to 24°47′N. The area of Dongjiang watershed is 35340 km^2, among it, 90% in Guangdong province. It has the typical subtropical monsoon humid climate, average annual temperature 20℃-22℃, and the precipitation is 1500mm-2400mm.With the rapid economic development in the recent 20 years, ecosystems have been destroyed.

TM images of 1988, 1998 and 2007 were acquired for the study respectively. Moreover, the data of DEM, water quality assessment, statistical yearbook and meteorological record etc. were also been collected.

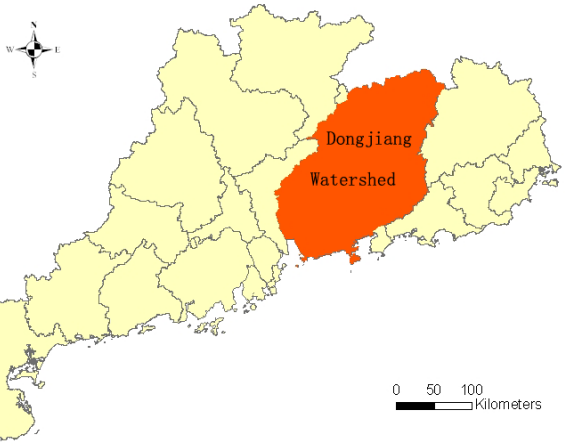


Fig. 1. The location of the study area

3 Research Method

The index system (Table 1) was established including three first-class indicators and eighteen second-class indicators by the Pressure-State-Response model [10]. The weight of each hierarchy is valued and calculated by the analytical hierarchy process considering the expert's weighted vectors.

A pixel was taken as the basic evaluation unit. Thematic maps were spatial overlay after standardized processing, using the weight sum method. The index of eco-security in each pixel and the comprehensive index in each country have been calculated by the models, in order to achieve a quantitative evaluation of the regional eco-security.

The models of eco-security index:

$$A_i = \sum_{k=1}^{n} W_k Y_k \tag{1}$$

$$C_i = \frac{\sum A_i \times S_i}{\sum S_i} \tag{2}$$

Table 1. The index system of the eco-security evaluation

Criterion layer (Weight)	Element layer (Weight)	Indicator layer(Weight)
Eco-security Pressure (0.35)	Resource pressure (0.2)	Farmland areas per person (0.33)
		land degradation index (0.67)
	Social pressure (0.3)	population density (0.20)
		Pressure from residential points (0.40)
		Pressure from traffic line (0.40)
	Environmental pressure (0.5)	Intensity of fertilizer application per farmland area (0.33)
		Influence degree of sand mining (0.53)
		River water quality (0.14)
Eco-security State (0.55)	natural conditions (0.5)	Elevation index (0.30)
		slope index (0.40)
		Annual average temperature (0.10)
		Annual rainfall (0.20)
	resources quantity (0.5)	vegetative cover index (0.38)
		Rivers density index (0.14)
		organism abundance index (0.28)
		soil index (0.20)
Eco-security Response (0.1)	social response (1)	per capita GDP (0.67)
		environmental protection (0.33)

A_i denotes the eco-security in the ith pixel; W_k denotes the weight of the kth index in the same pixel; Y_k denotes the quantitative value after standardized processing. C_i denotes the comprehensive eco-security of the ith country; S_i denotes the number of A_i in the same country. The result is a value between 0 ~ 10, and the larger the better.

For the purpose of comparison, and according to practical situation, the eco-security was divided into six levels From level Ⅰ to level Ⅵ, the eco-security is worse and worse. (Table 2).

Table 2. Corresponding table between eco-security index and eco-security level

	level Ⅰ	Level Ⅱ	Level Ⅲ	Level Ⅳ	Level Ⅴ	Level Ⅵ
value	>7	6.5-7	6-6.5	5.5-6	5-5.5	<5

Borrowed the model of land-use change [11], the relative change rate model has been built. It is a Effective way to reflect regional difference of changes.

$$R = (K_2/K_1)/(C_2/C_1) \tag{3}$$

K_1 and K_2 denote respectively the area of a particular level in the five regions at the beginning of the study period and at the end of the study period. C_1 and C_2 denote respectively the area of a particular level in the whole Dongjiang watershed at the beginning of the study period and at the end of the study period. If the result R>1, the change of the particular level in the region is higher than that in the whole watershed.

If the result R<1, it is lower.

In this paper, the CA_Markov model was developed for modeling and predicting the eco-security change in this area, which was based on a rule according to the effect of comparison between core and outer- cellular. The Markov chain model has shown the capabilities of descriptive power and simple trend projection of the eco-security change, regardless of whether or not the trend actually persists. The analysis can serve as an indicator of the direction and magnitude of change in the future as well as a quantitative description of change in the past. Compared with Markov model, CA model has the capability to show spatial information of the eco-security changes. It is obvious that the Markov chain model is the key process and footstone of CA model. So the CA_Markov model can provide a way to predict temporal and spatial change of the eco-security change. Here, the changes images and matrixes from 1998 to 2007 were taken to determine the modeling rules and probability, which were used to predict the eco-security in 2016.In the CA model, the pixel is the cellular, and the six types of the eco-security, from level Ⅰ to levelⅥ, are the cellular state. A suitable filter and transition suitability image will be collected to define the evolution rules of CA-Markov model.

4 Results and Discussions

4.1 Overall Evaluation

Table 3. Area statistical of different eco-security level in DongJiang Watershed (Unit: km^2)

	level Ⅰ	Level Ⅱ	Level Ⅲ	Level Ⅳ	Level Ⅴ	Level Ⅵ
1988	8528.14	11069.48	7608.73	2834.36	694.59	105.44
1998	9208.01	10545.90	6913.43	2887.61	843.11	442.69
2007	3022.63	9997.47	9905.70	4349.79	2252.09	1313.01

The area statistical of different eco-security level in DongJiang Watershed was shown in Table 3. The results clearly indicated that the level Ⅰ area in Dongjiang watershed decreased 5505.51 km2 in past 20 years from 8528.14 km^2 in 1988 to 3022.63 km2 in 2007, and its loss rate reached 64.56%. It is noteworthy that the level Ⅰ area increased 679.87km^2 from 1988 to 1998. Its trend was “first increase and then decrease”. The level Ⅱ area vanished 1072.01km^2 from 1988 to 2007, with the loss rate of 4.73% and 5.20% respectively in the first 10 years and the last 9 years. Its trend was keeping losing. The level Ⅲ area vanished 695.3 km^2 from 1988 to 1998, while increased 2992.27km^2 from 1998 to 2007, with the loss rate and the increasing rate of 9.14% and 43.28%. Its trend was “first decrease and then increase”. The level Ⅳ area

increased 53.25 km^2 and 1462.18 km^2 respectively from 1988 to 1998 and 1998 to 2007, with the high increasing rate of 53.47%. Its trend was keeping increasing. The trend of level Ⅴ and Level Ⅵ were the same as level Ⅳ. The former increased 1557.5 km^2 totally from only 694.59 km^2 in 1988 to 2252.09 km^2 in 2007, with the increasing rate of 224.23%. The latter increased 337.25 km^2 from only 105.44 km^2 in 1988 to 442.69 km^2 in 1998, and continue increased 870.32 km^2 to 2007, with the increasing rate of 319.85% and 196.6% respectively.

4.2 Region Difference

In order to analysis the region difference, Dongjiang watershed is divided into five regions; they are upstream, middle reaches, downstream, xinfengjiang and dongjiang delta. The eco-security evaluation images are shown in Fig.2. The area proportion of different levels in different regions in 1988, 1998 and 2007 are shown respectively in Fig.3, Fig.4 and Fig.5. The relative change rate from 1988 to 1998 and 1998 to 2007 has been calculated respectively by the model, which was shown in Table4. The results show that the eco-security of Dongjiang watershed has obvious space differences. And the eco-security was better in Xinfengjiang, upstream,and middle reaches than that in dongjiang delta and downstream.

Upstream: The eco-security value was 6.62, 6.75 and 6.47 respectively in 1988, 1998 and 2007. The level Ⅱ and the level Ⅲ area accounted for the largest share. There was only a little change in the past 20 years. The relative change rate of level Ⅰ and level Ⅱ from 1988 to 1998 in upstream was lower than the whole watershed. The phenomenon was opposite from 1998 to 2007.

Middle reaches: The eco-security value was 6.77, 6.8 and 6.57 respectively in 1988, 1998 and 2007. It has a more or less uniform spatial distribution. There were some slight changes in the northern and southwestern regions, Yuancheng and central urban area of Longchuan. In 1988 and 1998, the level Ⅰ and the level Ⅱ area

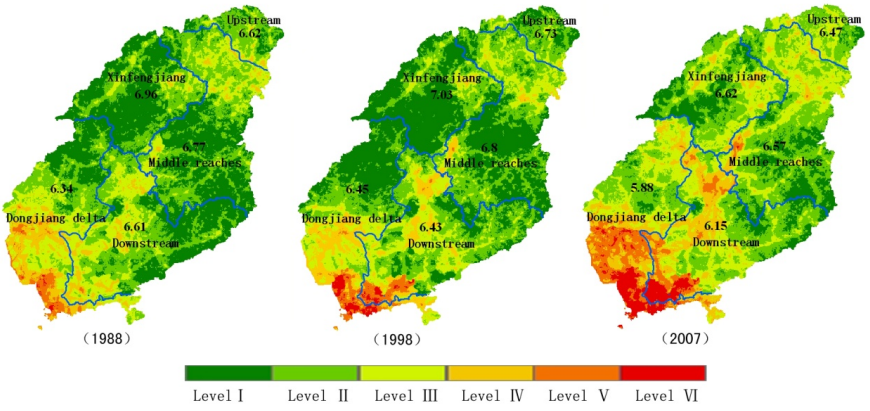


Fig. 2. Eco-security evaluation result in different regions

accounted for the largest share, while were level Ⅱ and level Ⅲ in 2007. The proportion of the level Ⅱ, Ⅳ, Ⅴ and Ⅵ area had no big changes. The level Ⅰ area vanished 17.97% and the level Ⅲ increased 11.89%. All the relative change rate in middle reaches was higher than or close to the whole watershed, in especial , the relative change rate of level Ⅵ from 1988 to1998 and 1998 to 2007 reached 5.92 and 7.40 respectively.

Downstream: The eco-security value was 6.61, 6.43 and 6.15 respectively in 1988, 1998 and 2007. It was declined year by year. The region had large changes mainly concentrated in the central and southern. From 1988 to 1998, the level Ⅱ, Ⅲ and Ⅳ area had no big changes. The level Ⅰ decreased 9.66% and the level Ⅴ increased 4.17%, and the level Ⅵ appeared. From 1998 to 2007, the level Ⅰ, Ⅱ, and Ⅲ area all had different degree of reduced, other levels opposite. The level Ⅴ and Ⅵ increased 4.92% and 3.95% respectively. The relative change rate of level Ⅴ and level Ⅵ in downstream from 1998 to 2007 was much higher than others, which reached 8.11 and 228.67.

Xinfengjiang: The eco-security value was 6.96, 7.03 and 6.62 respectively in 1988, 1998 and 2007. It was the best in the five regions. The worse eco-security were only scattered sporadically in opencast mining area. There were little change from 1988 to 1998, and some changes from 1998 to 2007. But overall, The level Ⅰ, Ⅱ and Ⅲ area accounted for the largest share. Most relative change rate in Xinfengjiang was higher than the whole watershed, except level Ⅴ from 1988 to 1998 and level Ⅱ from 1998 to 2007.

Dongjiang delta: The eco-security value was 6.34, 6.45 and 5.58 respectively in 1988, 1998 and 2007. It was worst in the five regions. Large changes had taken place both on the distribution and the area. The level Ⅰ area decreased 13.99% from 15.6% in 1988 to only 1.61%, while The level Ⅴ and Ⅵ increased 10.22% and 10.27% respectively. Most relative change rate in Dongjiang delta was lower than the whole watershed, except level Ⅴ from 1988 to 1998 and level Ⅰ, level Ⅱ from 1998 to 2007.

Table 4. The relative change rate of DongJiang watershed in different parts

		Upstream	Middle reaches	Downstream	Xinfengjiang	Dongjiang delta
From 1988 to 1998	Level Ⅰ	0.64	1.29	1.68	1.05	0.18
	Level Ⅱ	0.75	1.12	0.79	1.39	0.84
	Level Ⅲ	1.62	1.11	0.62	1.96	0.93
	Level Ⅳ	2.09	1.12	1.16	2.72	0.73
	Level Ⅴ	1.58	1.77	0.77	0.42	1.17
	Level Ⅵ	/	5.92	1.12	/	0.94
From 1998 to 2007	Level Ⅰ	1.57	0.96	0.54	1.07	1.57
	Level Ⅱ	1.06	1.04	1.05	0.83	1.02
	Level Ⅲ	0.90	1.06	1.16	1.10	0.76
	Level Ⅳ	0.28	0.99	1.19	1.11	0.93
	Level Ⅴ	0.04	1.35	8.11	1.30	0.58
	Level Ⅵ	/	7.40	228.67	/	0.69

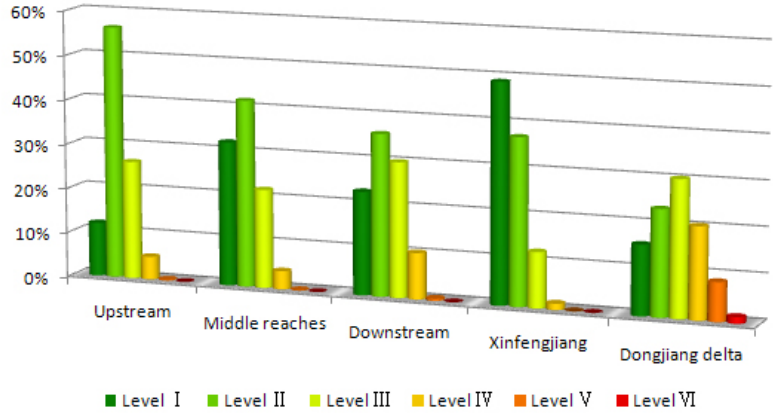


Fig. 3. Area proportion of different levels in different regions in 1988

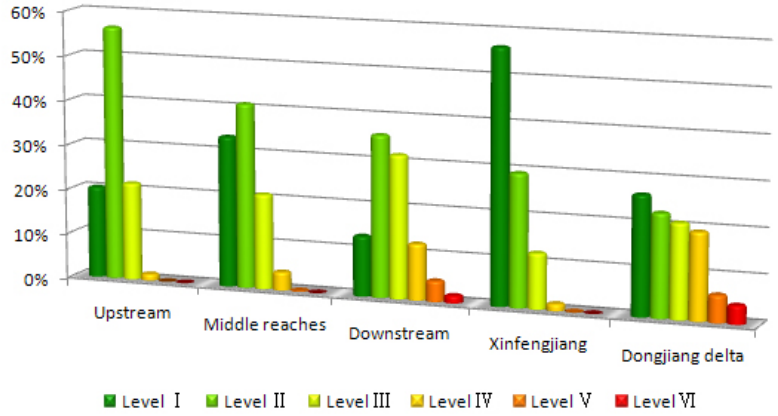


Fig. 4. Area proportion of different levels in different regions in 1998

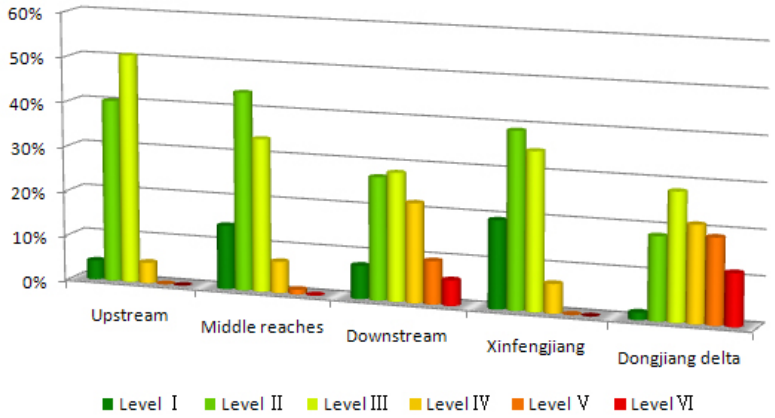


Fig. 5. Area proportion of different levels in different regions in 2007

4.3 CA_Markov Model

Here, the changes from 1998 to 2007 were taken as modeling rules to predict the eco-security in 2016. The transition matrix of the eco-security was shown as Table 5. While in the CA model, the pixel is the cellular, and the six types of the eco-security, from level Ⅰ to level Ⅵ, are the cellular state. The 5*5 filter was used to define neighbors. The eco-security of a cellular next time depends on the eco-security of this cellular and its neighbors this moment. Transition suitability image was collected to define the evolution rules of CA-Markov model. And the cycle number was set 9, which was the same of 9-year interval from 1998 to 2007. The predicted image in 2016 is shown in Fig.6. From level Ⅰ to level Ⅵ, the area will be 2031.51 km^2, 6391.82 km^2, 10993.31 km^2, 6376.62 km^2, 2895.77 km^2 and 2073.97 km^2 respectively in 2016.

Table 5. The transition matrix of the eco-security from 1998 to 2007

(Unit: km^2)

1998 / 2007	Level Ⅰ	Level Ⅱ	Level Ⅲ	Level Ⅳ	Level Ⅴ	Level Ⅵ
Level Ⅰ	2888.21	123.51	10.91	0.00	0.00	0.00
Level Ⅱ	4445.12	4992.48	545.67	14.20	0.00	0.00
Level Ⅲ	1676.06	4460.10	3671.85	94.09	3.60	0.00
Level Ⅳ	196.01	709.79	1927.46	1503.59	12.94	0.00
Level Ⅴ	2.60	260.02	705.34	969.26	311.48	3.39
Level Ⅵ	0.00	0.00	52.15	306.48	515.08	439.30

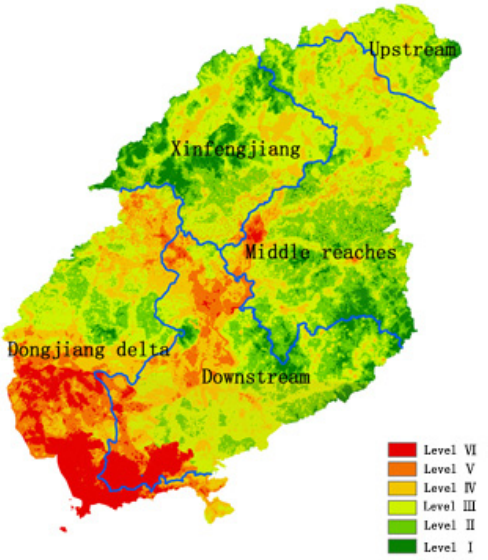


Fig. 6. The predicted image in2016

Acknowledgments. This work was supported by Guangdong NSF (10151007003000002), Comprehensive strategic cooperation project of the Chinese academy of sciences and Guangdong province (2011B090300048) and Fund of Guangdong modern information service industry development (GDIID2009IS001).

References

1. Norman, M.: Ultimate Security—the Environmental Basis of political Stability. W.W.Norton, New York (1993)
2. Public Meeting—Science for ecological security, http://www.cseindia.org
3. Li, Y., Liu, Y., Yan, X.: A DPSIR-Based Indicator System for Ecological Security Assessment at the Basin Scale. Acta Scientiarum Naturalium Universitatis Pekinensis 48, 971–981 (2012)
4. Yang, Y., Ren, Z.-Y.: Assessment and Analysis on the Ecological Security in the Middle and Lower Reaches of the Jinghe River Basin. Arid Zone Research 26, 441–446 (2009)
5. Du, Q., Xu, X., Li, H., Peng, H.: Analysis on Ecological Security Changes of the Oases in the Middle and Lower Heiher River. Acta Scientiarum Naturalium Universitatis Pekinensis 41, 173–181 (2005)
6. Tang, J., Zhu, Y.-F., Li, Z.-Y., Si, A., Cui, J.: Evaluation on Ecological Security of Land Resources in Ecotone Between Farming and Animal Raising in Northeastern China: A Case Study of Zhenlai County. Journal of Arid Land Resources and Environment 20, 119–124 (2006)
7. Zhang, J.: The ecological safety and its assessment principle in arid: A case of Xinjiang. Ecology and Environment 16, 1328–1332 (2007)
8. Zhang, H.-B., Liu, L.-M., Zhang, J.-L., Zhu, Z.-Q.: A Dynamic Assessment of Ecological Security of Land Resources in Loess Hills Region. Resources Science 29, 193–200 (2007)
9. Shang, Z.-H., Long, R.-J., Ma, Y.-S.: Review on environmental problems in the headwater areas of Yangtze and Yellow Rivers in Qinghai-Tibetan Plateau. Pratacultural Science 24, 1–7 (2007)
10. Sun, C., Zhong, K., Liu, X.: Remote Sensing Analysis on Ecological Security Changes in the Recent 20 years of Dongjiang Watershed. Territory & Natural Resources Study 2, 51–52 (2012)
11. Wang, X., Bao, Y.: Study on the methods of land use dynamic change research. Progress in Geography 18, 81–87 (1999)

US Experience Will Advance Gulf Ecosystem Research

Nabil Abdel-Jabbar[1,*], António M. Baptista[2], Tuomas Karna[2], Paul Turner[2], and Gautam Sen[1]

[1] American University of Sharjah, Sharjah, PO Box 26666, UAE
nabdeljabbar@aus.edu
[2] NSF Science & Technology Center for Coastal Margin Observation & Prediction (CMOP), Oregon Health & Science University, Beaverton, OR 97006

Abstract. This paper addresses the vision and early steps of the cooperation between the Center for Coastal Margin Observation and Prediction (CMOP) in Oregon, United States and the emerging Gulf Ecosystems Research Center (GERC) at the American University of Sharjah, UAE. The cooperation focuses on a better understanding and ability to predict the Arabian Gulf as a complex ecosystem, and involves science, technology and training components. An ultimate goal is the development for the Gulf of a "collaboratory" inspired on the concepts of integration of observations, simulations and stakeholder needs developed by CMOP for the Columbia River coastal margin, in the Eastern North Pacific. An early phase of the cooperation addresses the development of a 3D numerical model for the Arabian Gulf water circulation. A very preliminary forecasting system has been developed at CMOP, and its skill will be systematically assessed and improved by GERC and CMOP over the next several years, with the progressive deployment of a targeted observation network. Preliminary products include the visualization of the salinity fields associated with various river plumes. The model used was SELFE (a Semi-implicit Eulerian–Lagrangian Finite-Element model for cross-scale ocean circulation), the same that is being used for the Gulf predictions. Exploratory simulations were made to assess the ability of simple grid refinement strategies and/or use of higher order numerical schemes in improving the representation of the complex dynamics of plumes, filaments (eddies) and upwelling in the continental shelf of the Eastern North Pacific, off the Columbia River. Results suggested the need for automated grid optimization strategies, which are currently in progress.

Keywords: Finite Element, Mesh Refinement, EulerianLagrangian method, River Plumes, Filaments, Eddies, Model Skill Assessment, Arabian Gulf, Columbia River.

* Corresponding author.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 125–140, 2013.

1 Introduction

The Arabian Gulf is a shallow (average depth 35 m) inland sea in which the water circulation is essentially closed with the exception of small opening through the Strait of Hormuz to the Gulf of Oman. It supports a rich ecosystem of coral reefs, dugongs, fisheries, and other coastal population. However, pollution resulting from desalination, wars, extensive shipping activity, oil spills, and external factors such as climate change induced sea-level rise have created major environmental stress in the Gulf, threatening its rich biodiversity [1].

Evidence of stress on the Gulf's environment comes from recent bleaching of corals and the harmful algal bloom ('red tide') that occurred in 2008-2009 [2]. Such blooms can have a potentially devastating impact; red tide may result in mass killing of fish, birds, and other marine animals, thus threatening the fisheries industry, both wild and aquacultured. It also poses a potentially catastrophic threat to the drinking water supplies of the region due to shutting down desalination plants. The Gulf region heavily depends on desalination plants that convert seawater into drinking water. During the 2008-2009 red tide, many desalination plants in the region were affected, with one plant closed for 55 days during the bloom [1].

Accurate modeling of physical circulation and transport is foundational to understanding the Arabian Gulf coastal ecosystem in its biogeochemical complexity, and to addressing a broad range of increasingly essential operational and management issues: from support to navigation and response to oil spills, to the management of desalination plants and the mitigation of red tides).

Many investigators (e.g, [3], [4]) have employed or developed hydrodynamic models to study the circulation in the Arabian Gulf. Such models predict the salinity and temperature distribution fields, fresh water intrusion from the Strait of Hormuz and river inflows (Shatt-al-Arab), and the generally counterclockwise circulation especially in the southern half of the Gulf [3]. Also, tidal and winds effects can be incorporated in the models. Typically, the surface temperature of the Gulf is about 33 ^{o}C in summer and it varies to about 22 ^{o}C near the Strait of Hormuz and about 16 ^{o}C in the up north of the Gulf. The salinity is almost constant throughout the year, but varies spatially from about 36 psu near the Strait of Hormuz to about 41 psu off the Saudi Arabian shores. The high salinity is more apparent in the southern and southwestern coasts of the Gulf [3].

Motivated by the above considerations, a center called the Gulf Ecosystem Research Center (GERC) was established at the American University of Sharjah, UAE, to address all challenges pertinent to the Gulf ecosystem. Part of the mission of GERC will be to monitor and conduct research on ecosystems. GERC aims to offer advice on long-term solutions to a number of ecological problems. In particular, the development of Arabian Gulf circulation models would help with prediction of red tides in an attempt to prevent damage to the water processing plants filtering membranes that filter the salt water.

Collaboration between the Center for Coastal Margin Observation and Prediction (CMOP) in the United States and the emerging Gulf Ecosystems Research Center (GERC) at the American University of Sharjah, UAE, was initiated in 2011. This cooperation involves mutual exchange visits of the scholars from both institutions. The extensive experience of CMOP in modeling and monitoring the complex ecosystem of the Columbia River (CR) estuary and its continental shelf was seen as a potential starting point of such cooperation. In preparation for the cooperative maintenance and improvement of that forecast system, the first author spent a sabbatical period at CMOP, where he became familiar with SELFE and explored ways to improve the modeling of circulation of the continental shelf of the Eastern North Pacific.

2 Columbia River (CR)

The CR is the largest river entering the Eastern North Pacific Ocean (Figure 1), extending from the British Columbia coast in Canada to the Washington, Oregon, and California coasts of the United States. The CR plume exports dissolved and particulate matter hundreds of kilometers along and across the continental shelf [5, 6], which in turn influences the shelf ecosystems. Hickey and Banas[7] and Hickey et al. [8] suggested that the seaward front of CR plume likely provides a barrier to the onshore transport of harmful algal bloom in summer and early fall [9].

The CMOP research addresses biological hotspots in the CR estuary, from an integrated perspective across disciplines, observations and simulations. Of immediate interest to GERC, the CMOP studies include the characterization of the river-to-shelf circulation at various scales of spatial and temporal variability, including tidal and wind-driven baroclinic dynamics [9, 10]. Advanced modeling tools have been developed to support these studies, constituting what the “Virtual Columbia River” [11].

The Virtual Columbia River simulate the 3D baroclinic circulation in the river-to-shelf system, both retrospectively and in near real-time forecasting mode. The simulations of circulation are conducted in unstructured grids, typically with the finite element code SELFE [12] and historically also with the finite volume code ELCIRC [13]. All simulations are skill assessed ([9], [10]) against an extensive observation network maintained by CMOP and supported by strong cyber-infrastructure capabilities..

Transferring the CMOP modeling capabilities to GERC, in a form customized to the Arabian Gulf, is a priority objective, and a key step towards achieving the global vision of GERC. The transfer requires a process of workforce development and training. As a part of that process, the first author conducted an exploratory research with SELFE, in a complex yet well understood benchmark. The scientific goal was to assess the ability of simple grid refinement strategies and/or use of higher order numerical schemes to improve the representation of the complex dynamics of upwelling and circulation in the continental shelf of the Eastern North Pacific, off the Columbia River.

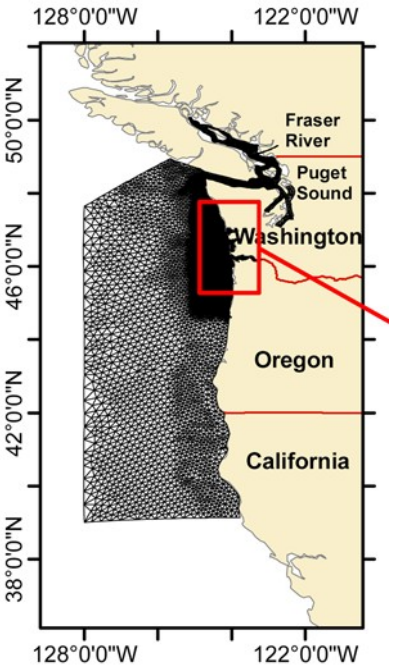


Fig. 1. Columbia River Estuary-Plume-Shelf System (inside the square zoom)

3 The SELFE Model

SELFE (Semi-implicit Eulerian–Lagrangian Finite Element; Baptista et al. 2008) is a finite element code that anchors an open-source, community-supported, interdisciplinary modeling system. It uses unstructured grids, and is designed for the effective simulation of 3D baroclinic circulation, transport and support of ecological modeling across river-to-ocean gradients and scales.

While originally developed to meet specific modeling challenges for the Columbia River, SELFE has been extensively applied to study coastal margins around the world. SELFE uses a semi-implicit finite-element/volume Eulerian-Lagrangian scheme to solve the primitive form of the shallow water equations (in either hydrostatic or non-hydrostatic form). The numerical algorithm is robust, stable and computationally efficient. It also naturally incorporates wetting and drying of tidal flats.SELFE v3.1d available at CMOP [14] was used for all simulations in this study. Parallel computations were carried out in a cluster using standard Message Passing Interface (MPI).

4 Results and Discussion

One of the operational products of the Virtual Columbia River are multi-year simulation databases of circulation. One specific such database ("DB29") was used as reference in this study, and we focused specifically on its results for April-May 2012. We modified that simulation database in various ways, each leading to a separate simulation. First, the continental shelf domain was re-discretized for a base case (run02). Refinements of the based mesh were made, leading to two additional simulations (run04 and run05). A sample mesh grid of the continental for run04 is shown in Figure 2. This grid was generated using the Surface Water Modeling System (SMS) software [15]. Table 1 summarizes the size of the grid used for the base run (run02) and the other two runs (run04 & run05). For the baseline run run02, a total of 37,146 triangles elements was used in the horizontal grid, with a higher resolution (<500 m) concentrated in the near-plume and shelf regions. The bathymetry of CR shelf was loaded onto the program.

Figure 3 shows one sample plot of surface salinity for the CR shelf from SELFE for the baseline run (run02), and runs run04 and run05, where significant refinement of the grid was applied. Furthermore, Figure 4 shows close-up comparison of the surface salinity near the plume for runs 04 and run05. Evaluation of model skill is shown in Figure 5 as a comparison of salinity for all runs with observed salinity data from SATURN-02 station at depth of 1 m. Figure 6 depicts the model skills for the temperature profile for NH10 station at a depth of 2 m.

Clearly, mesh refinement did not affect the simulation results drastically in terms of detailing plume dynamic features (upwelling, filaments, eddies, etc.). Furthermore, the SELFE model did not capture well the salinity data at the depth of 1 m and the simulation runs exhibited almost same deviation with a maximum error of -20 psu was observed. However, as the grid is more refined, it can be noticed that sharper salinity gradients become more apparent.Nevertheless, the model skills for temperature profile were more accurate as shown in Figure 6.

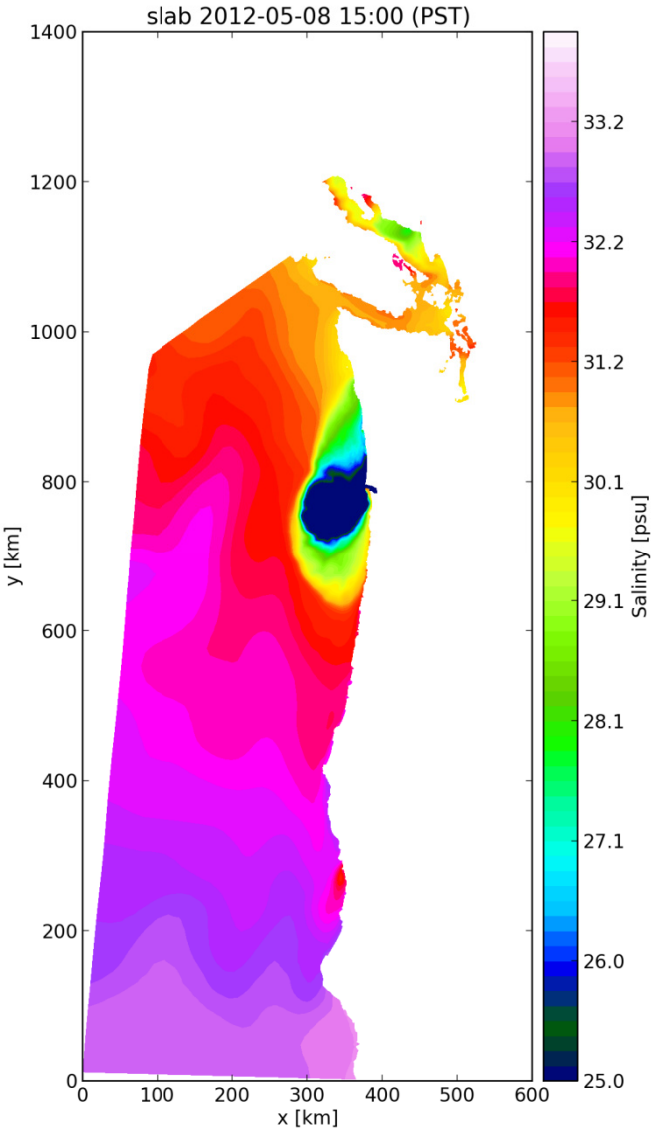
Analysis of the above simulation results triggered the need to try different numerical scheme to explore the complex dynamics of the CR shelf system. Total Variation Diminishing (TVD) transport scheme was employed in run06 and the surface salinity at one instance is shown in Figure 7. The zoomed surface salinity is shown in Figure 8. Obviously, run06 with TVD provided sharper salinity gradients in the plume area, yet eddies were not detected with this numerical scheme.

The experience acquired with this sensitivity study for the Columbia River has permitted the first author to become familiar with SELFE, and to begin exploring the application of SELFE to the Arabian Gulf. A Gulf grid was prepared as shown in Figure 9 on SMS. Then, the gulf grid along with bathymetry was used in SELFE simulations. Figure 10 shows the preliminary surface salinity of the Gulf simulated by SELFE. The early stages of development of the river plumes from Shatt-Al-Arab up north and other river plumes were captured as seen in Figure 10, whereas more work is still needed to include the observed discharge data and to spin-up the model for long enough time to obtain realistic initial conditions.

Fig. 2. SMS Grid for run04. Grids are hybrid and resolution is finest in the shelf.

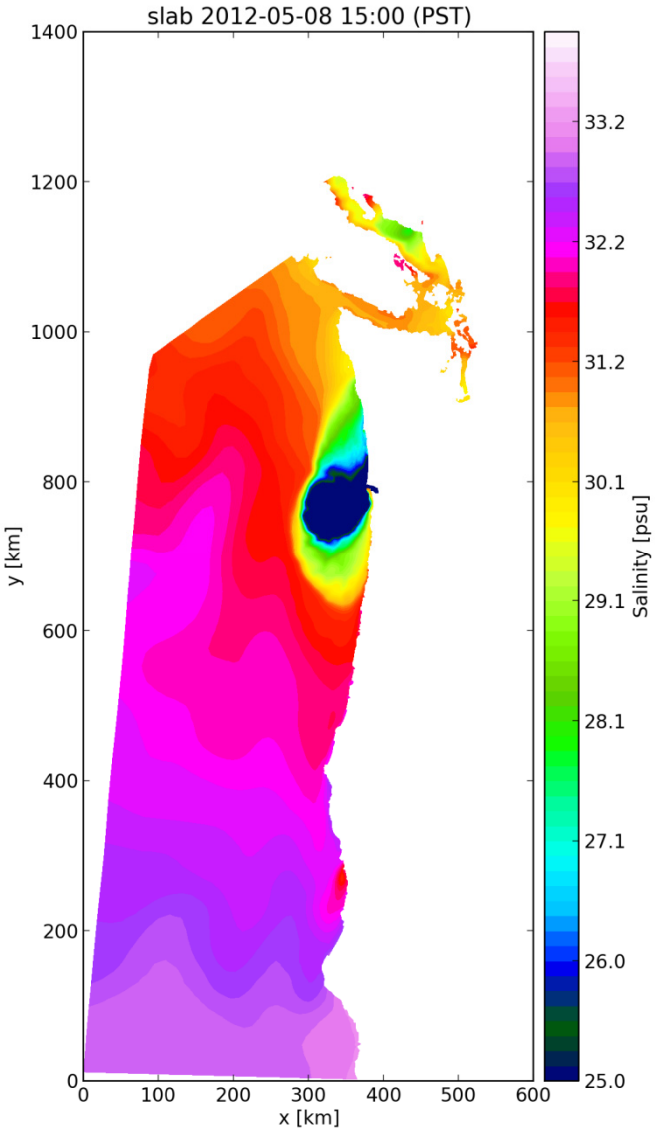
Table 1. Grid resolution and size for SELFE runs

Run	# of Elements	Resolution (Max size)
02	361938	500 m
04	3179179	200 m
05	3324056	100 m



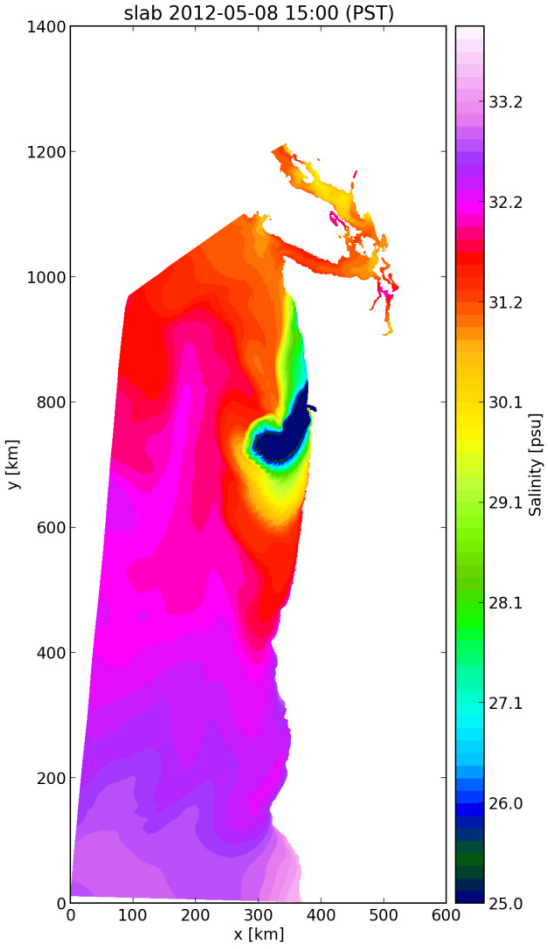
(a)

Fig. 3. Surface salinity for the three runs. (a) run02, (b) run04, (c) run05.



(b)

Fig. 3. (*Continued*)



(c)

Fig. 3. (*Continued*)

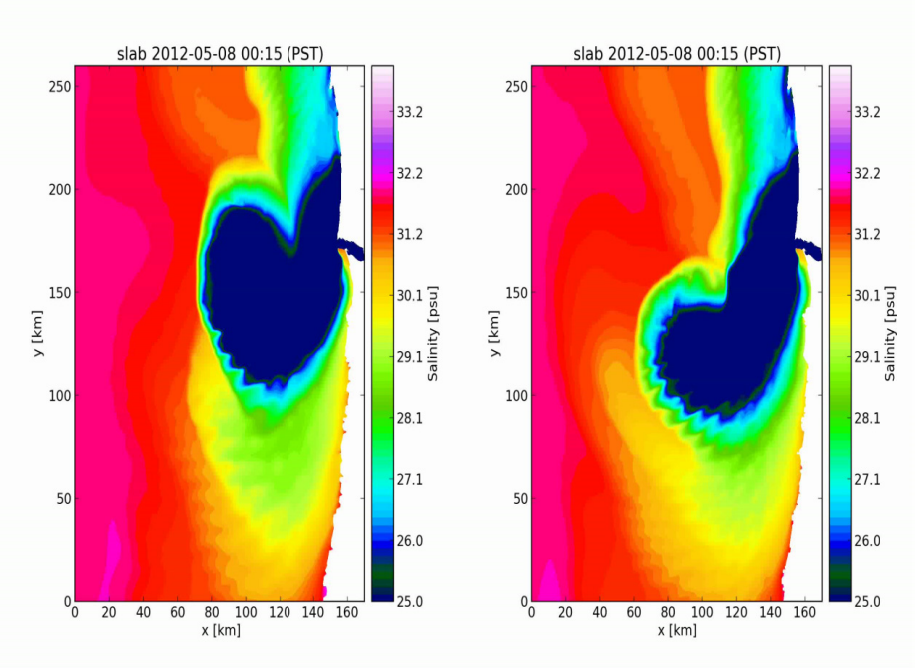


Fig. 4. Zoomed surface salinity for run04 and run05

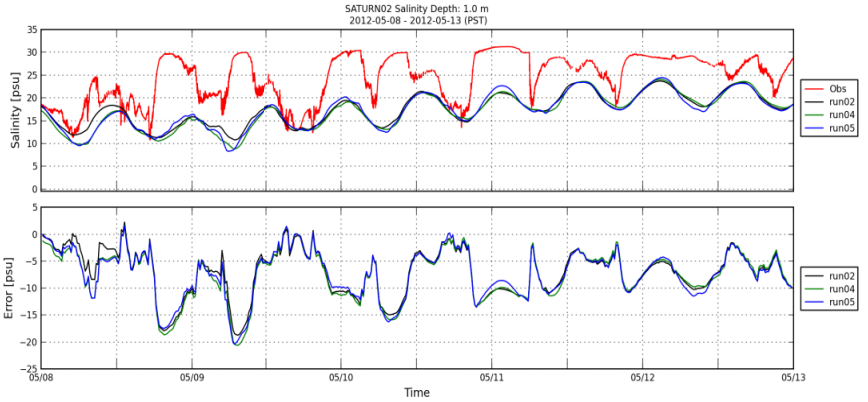


Fig. 5. Comparison of observed and simulated surface salinity at Saturn-02 station

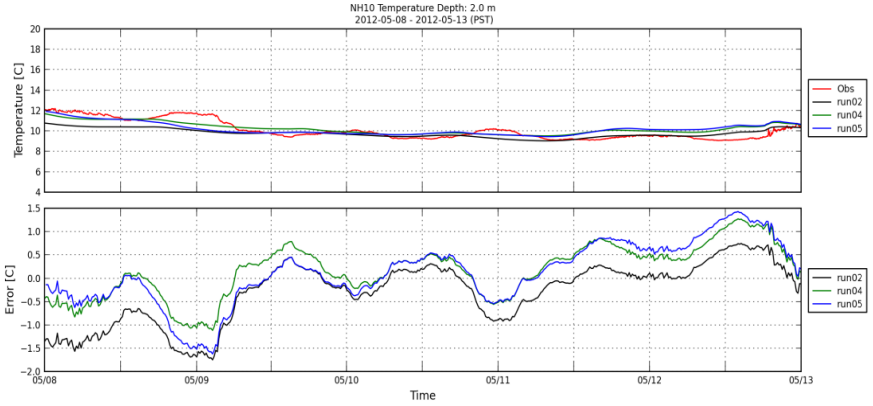


Fig. 6. Comparison of observed and simulated surface temperature at NH10 station

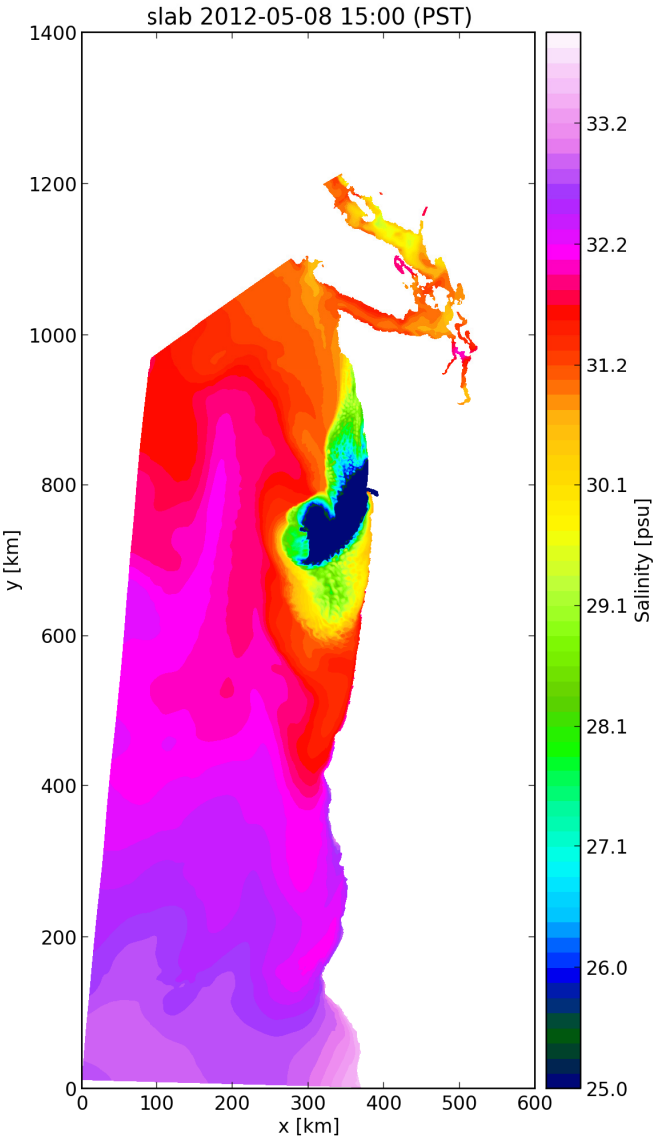


Fig. 7. Surface salinity for run06

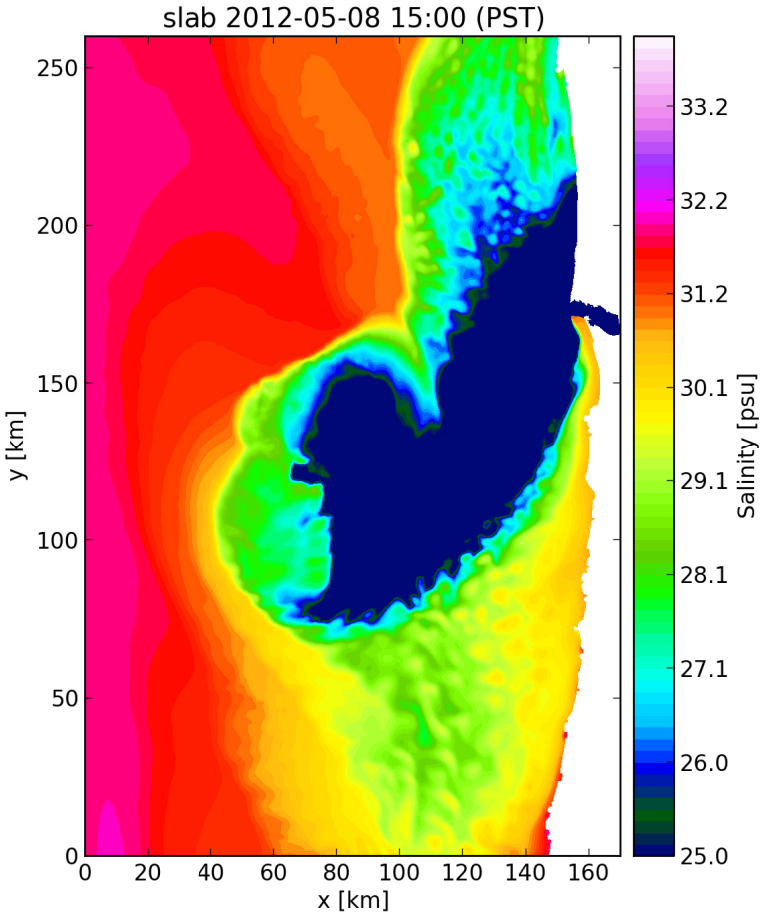


Fig. 8. Zoomed surface salinity for run06

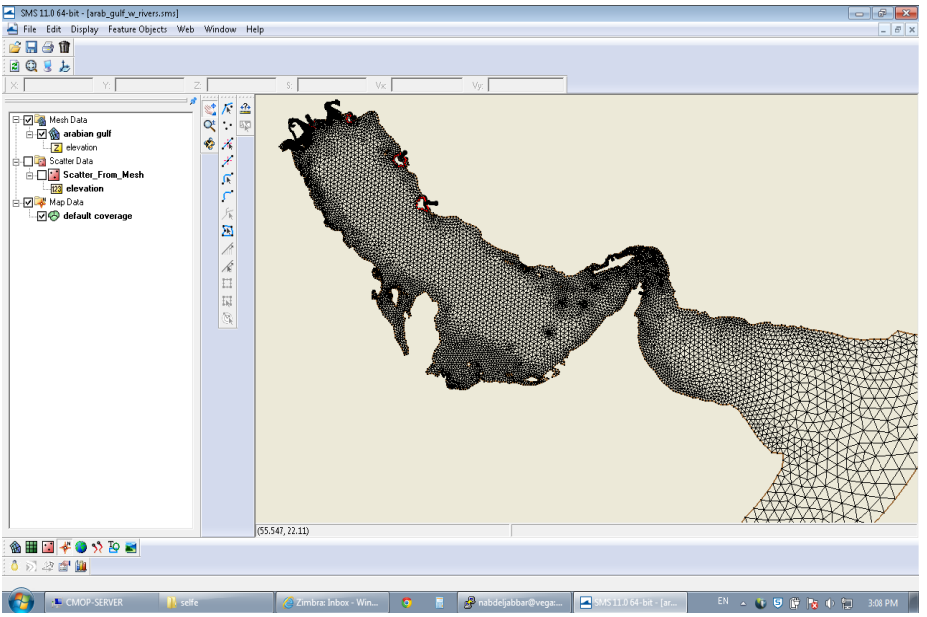


Fig. 9. A mesh for the Arabian Gulf

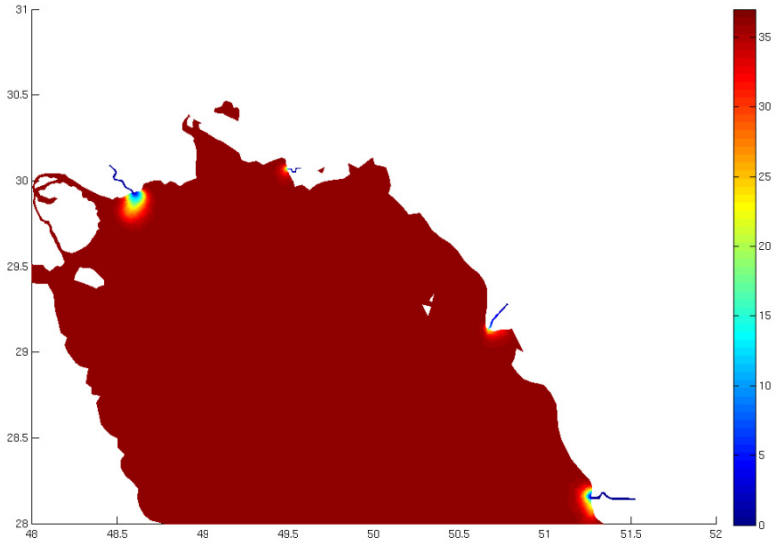


Fig. 10. Arabian Gulf surface salinity as simulated by SELFE

5 Conclusion

This paper emphasizes the vision and early steps of the cooperation between the Center for Coastal Margin Observation and Prediction (CMOP) in the United States and the emerging Gulf Ecosystems Research Center (GERC) at the American University of Sharjah, UAE.

An early phase on the cooperation focuses on the development of a 3D numerical model for the circulation in the Arabian Gulf. A very preliminary forecasting system has been developed at CMOP, and its skill will be systematically assessed and improved by GERC and CMOP over the next several years, with the progressive deployment of a targeted observation network.

Exploratory simulations with SELFE were made to assess the ability of simple grid refinement strategies and/or use of higher order numerical schemes in improving the representation of the complex dynamics of plumes, filaments (eddies) and upwelling in the continental shelf of the Eastern North Pacific, off the Columbia River estuary. Results suggested the need for automated grid optimization strategies, which are currently in progress.

Acknowledgments. The first author would like to acknowledge the financial support from AUS for his sabbatical leave and CMOP for hosting him during his sabbatical leave and for use of CMOP computer facilities.

References

1. Sen, G.: Long Term Monitoring of the Coastal Environment of the Arabian Gulf-Implications for Harmful Algal Blooms (Red Tides), Oil Spills and Other Catastrophic Events, A Proposal submitted to the Ministry of Environment & Water (October 23, 2011)
2. Richlen, M., Mortonb, S., Ebrahim, A., Rajand, A., Anderson, D.: The catastrophic 2008–2009 red tide in the Arabian gulf region, with observations on the identification and phylogeny of the fish-killing dinoflagellate Cochlodinium polykrikoides. Harmful Algae 9, 163–172 (2010)
3. Chao, S.-Y., Kao, T., Al-Hajri, K.: A Numerical investigation of circulation in the Arabian Gulf. Journal of Geophysical Research 97(C7), 11,219–11,236 (1992)
4. Kampf, J., Sadrinasab, M.: The circulation of the Persian Gulf: a numerical study. Ocean Science Discussions 2, 129–164 (2005)
5. Barnes, C.A., Duxbory, A.C., Morse, B.A.: Circulation and selected properties of the Columbia River effluent at sea. In: Alverson, D.L., Pruter, A.T. (eds.) The Columbia River Estuary and Adjacent Ocean Regions: Bioenvironmental Studies, pp. 41–80. University of Washington Press, Seattle (1972)
6. Grimes, C.B., Kingsford, M.J.: How do riverine plumes of different sizes influence fish larvae: do they enhance recruitment? Marine & Freshwater Research 47, 191–208 (1996)
7. Hickey, B.M., Banas, N.S.: Oceanography of the U.S. Pacific Northwest coastal ocean and estuaries with application to coastal ecology. Estuaries 26(4), 1010–1031 (2003)
8. Hickey, B.M., Geier, S., Kachel, N., MacFadyen, A.: A bidirectional river plume: The Columbia in summer. Cont. Shelf Res. 25, 1631–1656 (2005)

9. Burla, M., Baptista, A.M., Zhang, Y., Frolov, S.: Seasonal and interannual variability of the Columbia River plume: A perspective enabled by multiyear simulation databases. Journal of Geophysical Research 115, C00B16 (2010)
10. Baptista, A.M., Zhang, Y., Chawla, A., Zulauf, M., Seaton, C., Myers, E., Kindle, J., Wilkin, M., Burla, M., Turner, P.: A cross-scale model for 3D baroclinic circulation in estuary–plume–shelf systems: II. Application to the Columbia River. Continental Shelf Research 25, 935–972 (2005)
11. Baptista, A., Howe, B., Freire, J., Maier, D., Silva, C.: Scientific Exploration in the Era of Ocean Observatories. Computing in Science & Engineering 10(3), 53–58 (2008)
12. Zhang, Y., Baptista, A.M.: SELFE: A semi-implicit Eulerian-Lagrangian finite-element model for cross-scale ocean circulation. Ocean Modelling 21, 71–96 (2008)
13. Zhang, Y., Baptista, A.M., Myers, E.: A cross-scale model for 3D baroclinic circulation in estuary–plume–shelf systems: I. Formulation and skill assessment. Continental Shelf Research 24, 2187–2214 (2004)
14. SELFE at CMOP, `http://www.ccalmr.ogi.edu/CORIE/modeling/selfe/`
15. Surface Water Modeling System (SMS), `http://www.aquaveo.com`

The Application of an Improved Multi-surface Function Based on Earth Gravity Field Model in GPS Leveling Fitting

Bohu Yu, Zhen Guan, Xiaofeng Xu, and Liu Ou Yang

Northwest Institute of Nuclear Technology, Xi'an 710024, China
yubohu_2001@126.com

Abstract. This article mainly improved the Multi-surface function based on earth gravity field model EGM2008 by considering the shortcomings of Multi-surface function in theory. Improved Multi-surface function optimized the core function and reduced the difficulty in actual use. Compared with traditional Multi-surface function, improved Multi-surface function improved the precision and reduced workload.The result shows that improved fitting function can significantly improve the fitting precision of height anomaly and it also has high practical value.

Keywords: GPS leveling, fitting model, Multi-surface function, core function, earth gravity field model.

1 Introduction

GPS measurements can accurately determine the ground point geodetic height, and leveling can accurately obtain the normal hight he of the ground point accurately, comprehensive utilization of GPS and leveling can get accurate geoid. In the actual project, the way which use of GPS and leveling results by numerical fitting to determine the geoid is called GPS leveling. Therefore, the research of GPS leveling fitting model has important practical significance. By the accuracy and resolution of earth gravity field model improved and GPS technology improved the accuracy of coordinates in space. Using GPS measurement data and height anomaly data combine with leveling data can't only improve the accuracy of GPS Leveling fitting model greatly[1] but also make up the low accuracy of single fitting model. This article will combine Multi-surface function and Earth gravity field model organically. The author improved the Multi-surface function by introducing of the Earth gravity field model. Improved Multi-surface function improved the accuracy of fitting model and more perfect in theory.

2 Multi-surface Function Model

Multi-surface function was proposed by Hardy(American) in 1971, then this function was applied to geodetic survey, gravity anomaly fitting, geoidal height fitting, etc in

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 141–148, 2013.

1976 and crustal deformation fitting in 1987 in America. The basic idea of Multi-surface function is any one of a continuous regular or irregular surface can be determined by approximation of a few simple surface(or Single mathematical surface). This means to build a surface on each data point, then the surface will be stack into a continuous whole surface at certain percentage in the direction and make sure the the whole surface through each data point strictly. [2, 3]

Multi-surface function is an excellent interpolation which solved a mathematical surface adjustment problem by the data of points from the geometrical view. Suppose the height anomaly of a point can be expressed as

$$\xi = \sum_{j}^{n} A_j Q(X, Y, X_j, Y_j) \tag{1}$$

In formula, A_j is parameter to be determined, $Q(X,Y,X_j,Y_j)$is core function, (X_j,Y_j) is coordinate of center point, n is numbers of center point. There are many forms of core functions, such as

(1) cone

$$Q(X, Y, X_i, Y_j) = C + [(X - X_j)^2 + (Y - Y_j)^2]^{\frac{1}{2}} \tag{2}$$

(2) hyperboloid

$$Q(X, Y, X_i, Y_j) = [(X - X_j)^2 + (Y - Y_j)^2 + \delta^2]^{\frac{1}{2}} \tag{3}$$

(3) reciprocal hyperboloid

$$Q(X, Y, X_i, Y_j) = [(X - X_j)^2 + (Y - Y_j)^2 + \delta^2]^{-\frac{1}{2}} \tag{4}$$

In formula, δ is called smooth factor, usually a non-zero positive number, C is a constant, $(X\text{-}X_j)^2+(Y\text{-}Y_j)^2$is square of horizontal distance which from the reference point to the interpolation point.

if the number of known height anomaly is $m \geq n$, n points may be chosen as the center points of core function, its coordinates can be expressed as (X_i, Y_i), make

$$Q_{ij} = Q(X_i, Y_i, X_j, Y_j) \tag{5}$$

each point of height anomaly should fit

$$\xi_i = \sum_{j}^{n} A_j Q_{ij}, \; i = 1,2 \cdots m, \; j = 1,2 \cdots n \tag{6}$$

Observation equation can be listed as

$$V = QA - \xi \tag{7}$$

Solving the normal equation

$$A = (Q^T Q)^{-1} Q^T \xi \tag{8}$$

Then the height anomaly ξ_k ($k > n$) of any point X_k,Y_k) can be expressed as

$$\xi_K = Q_K^T (Q^T Q)^{-1} Q^T \xi \tag{9}$$

In formula(9)

$$Q_K = (Q_{K1} Q_{K2} \cdots Q_{Kn})^T \tag{10}$$

$$Q_{Kj} = Q(X_k, Y_k, X_j, Y_j) \tag{11}$$

If take whole height anomaly points as the the center points of the core function ($m = n$), then

$$A = Q^{-1} \xi \tag{12}$$

$$\xi_K = Q_K^T Q^{-1} \xi \tag{13}$$

suppose

$$P^T = Q_K^T Q^{-1} \quad QP = Q_K \tag{14}$$

formula （13） can be expressed as

$$\xi_K = P^T \xi \tag{15}$$

Combine formula (14) and formula (15)

$$\xi_K = \sum_{i=1}^{n} P_i \xi_i \tag{16}$$

In formula (16), P_i is the weight coefficient which can be obtained by solving equation (14). weights P_i determines the ξ_k, which is the average value of weights. The value of ξi is relate to the form of core function so the fitting accuracy of ξ_k has close relationship with core function. Optimizing the selection of core function can greatly improve the fitting accuracy.

3 Improvement of Multi-surface Function Model

3.1 Analysis of Multi-surface Function Model

Multi-surface function model is an excellent interpolation, but it has many problems in practical applications. First, the smooth factor is difficult to determine, it needs

constant trial to fit. Mass of trial work increase the difficulty of function using; Secondly, the nature of the core function is the horizontal distance function which between two points on the plane. The value of weight coefficient for each data point is only determined by horizontal and distribution of center point and point of certainty, ignoring the different regions have different height anomaly variations, from the change in elevation between the two relevant characteristics and regional considerations, the core functions which formula (2), (3) and (4) were showed are flawed in theory. Therefore, improving Multi-surface function by introducing the changing characteristics of the height anomaly into core function can not only improve the fitting accuracy in theory but also reduce the use of the model difficult.

3.2 Improve Core Function

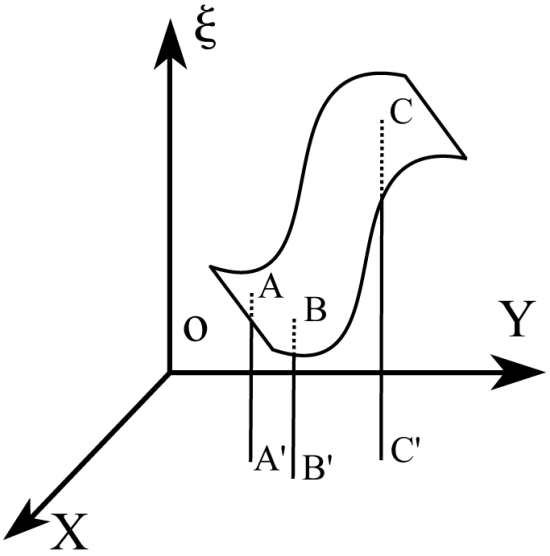


Fig. 1. Definition of new core function

As showing in Figure 1, assuming line *AA*', *BB*', *CC*' perpendicular to the plane *OXY*, according to common core function, the horizontal distance between points A and point B is less than point A and point C, so the weight of P_C is much smaller than the weights of P_B, it's clearly incompatible with the facts. In this paper, actual calculation found that in a small project area, the variety of height anomaly is a very tiny amount compare to variety of horizontal distance. Improved the core function by mass of calculation then introduced the character of height anomaly variety into core function in the form of implicit function, increased the height anomaly influence to the entire core function. improved core function expands horizontal distance of cone core function into the spatial distance. Improved core function as shown in equation17:

$$Q(X,Y,X_i,Y_j)=\sqrt{(X-X_j)^2+(Y-Y_j)^2}+(\xi-\xi_j)^2 \tag{17}$$

3.3 Earth's Gravity Field Model Introduced

According to multi-surface function model improvement in§3.2, when resolve the fitting parameters from the public point of A_i (i = 1,2,3 ... n) and calculate the height anomaly of unknown point, you need to seek an initial height anomaly and calculated into the final height anomaly. In previous research of improving Multi-surface function, literature [4] calculates height anomaly ξ_0 based on traditional Multi-surface function, and make ξ_0 the initial iteration to calculates height anomaly of point of certainty. Although this method achieved a high fitting accuracy, however, that this method is more complex, and lack of reliable theoretical basis. This paper found in experiments, normal height which obtained by leveling and earth gravity field model are vary in a small range (Table 1), normal height which was calculated by using earth gravity field model has more representation in region varies. Therefore, using the normal height which was obtained by the earth gravity field model as an initial value fed into normal fitting model, and calculate height anomaly of unknown point, not only reduces the amount of calculation but also has more persuasiveness in theory.

Table 1. Height anomaly comparison of leveling and EGM2008

point	height anomaly form leveling	height anomaly formEGM2008
MLD	-42.423	-56.573
GA1	-42.556	-56.743
GA2	-41.834	-55.919
P56	-43.500	-57.963
P52	-42.680	-56.897
P53	-43.229	-57.636
P58	-43.780	-58.026
P66	-43.849	-58.145

There are many available gravity field model currently, such as higher-order gravity field model GM-10C in 1978 model (expanded to 180 bands, model resolution 111*km*), OSU91A model in 1992 (expanded to 360 order model resolution 55*km*), EGM96 model in 1996 (expanded to 360 bands, model resolution 55*km*) [5] etc. These models lack of gravity data in china, so the accuracy of height anomaly is not high when using them. EGM2008 gravity field model is a high-precision gravity field model which was completed by NASA Gotha Flight Center (NASA / GSFC), the U.S. National Imagery and Mapping Agency (NMA), the U.S. Department of Defense (DOD) and the Ohio State University in 2008. EGM2008 gravity field model was built by ground gravity data (mainly gravity anomaly data), satellite tracking data and

satellite altimetry data, it's order was expanded to 2159, resolution of model in spatial is about 9km. Therefore, EGM2008 gravity field model greatly improved in both precision and resolution [6-8]. Previous research show that the accuracy of height anomaly which obtained by EGM2008 gravity field model is more accurate than EGM96 gravity field model, so this paper take normal height which obtained by EGM2008 gravity field model as the initial value to calculate height anomaly of unknown point.

4 Examples of Application

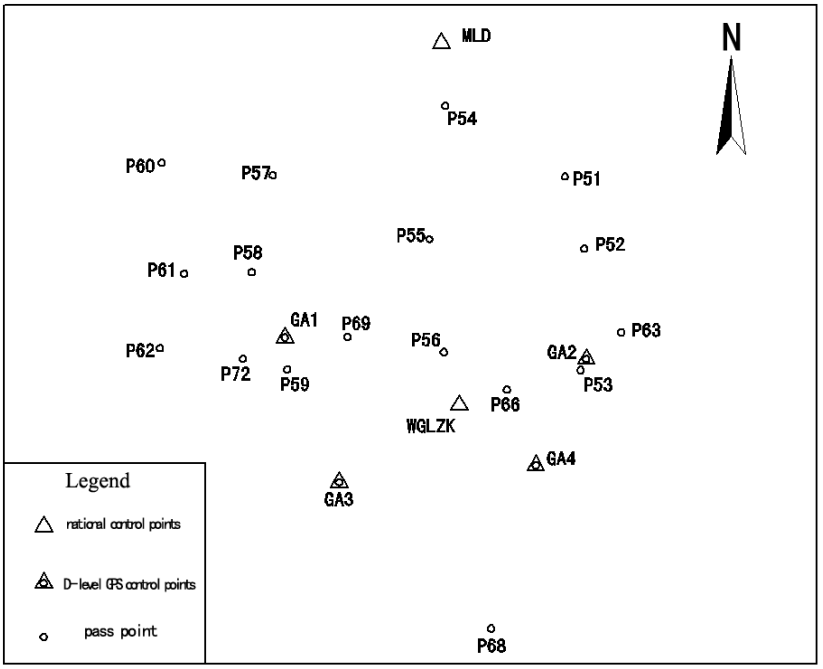


Fig. 2. Distribution of GPS points

Figure 2 shows the distribution of GPS points in a survey area ,the area is144km^2 in the low mountains. Minimum elevation is 1750m, maximum elevation is 3360m, elevation is 1600m. In the survey area, WGLZK and MLD are two national control points, GA1, GA2, GA3, GA4 are D-level GPS control points and other 17 pass point. The points in the survey area have both WGS84 coordinates and 1954 Beijing coordinates, height system is Huanghai vertical datum 1956.

we selected 15 points as the known point for fitting and the remaining 8 points for check. The improved Multi-surface function model fit and function with the use of inverted hyperbolic function method for Multi-faceted model fitting results. The comparison of calculation results which are used improved Multi-surface function and traditional Multi-surface function are shown in Table 1.

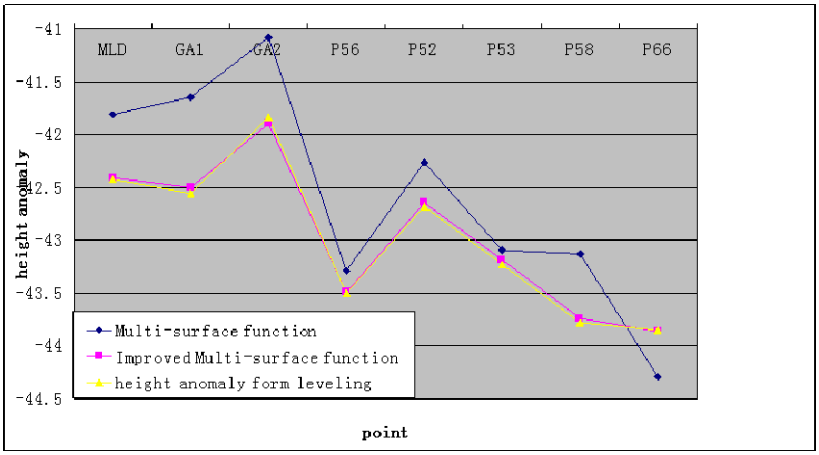


Fig. 3. Comparison of fitting precision between Multi-surface function and improved Multi-surface function

From fig 3 we can see that Fitting precision of improved Multi-surface function is better than traditional Multi-surface function model and the results are more closer to the true value.

5 Result

This paper improved multi-surface function model based on analysis of shortcomings in practical use, optimized form of the core function by introducing the Earth gravity field model.The improved multi-surface function overcomes the shortages that single fitting model is not reliable,and accuracy is not enough. the improved function not only simplifies the calculation process, bu also make the core function more perfect in theory, the final fitting precision has also been greatly improved. The fitting function can significantly improve the fitting precision of height anomaly and it also has high practical value.

References

1. Luo, Z., Ning, J., Xu, J.: Unified regional vertical datum. Science of Survey and Mapping (2), 13–15 (2004)
2. Liu, D., Tao, B.: Practical Methods for Surveying Data Processing. Surveying and Mapping Press, Beijing (2000)
3. Li, Z., Zhu, Q.: Digital Elevation Model. Wuhan University of Technology Press, Wuhang (2000)
4. Huang, Z., Feng, W.: Study of An Improved Method for Calculating the Normal Height by GPS. Bulletin of Surveying and Mapping (8), 36–39 (2007)

5. Hu, M.: The High Precision and High Resolution of the Earth Gravity Field Model EGM96. China Academic Journal (2010)
6. Pavlis, N.K., Holmes, H.A., Kenyon, S.C.: An Earth Gravitational Model to Degree 2160: EGM2008. Presented at the 2008 General Assembly of the European Geosciences Union, Vienna, Austria, pp. 13–14 (April 2008)
7. Earth Gravitational Model EGM 2008 (2008), http://earth-info.nga.Mil/GandG/wgs84/gravitymod/egm2008/index.html
8. Liu, X., Deng, Y., Ye, X.: Precision Comparison of the Earth Gravity Field of EGM96 and EGM2008. Hydrographic Surveying and Charting 30(2), 55–57 (2010)

An Emergency-Response Timing Constraint Workflow Model

Hai Liu, Jian Chen, and Shilong Ma

Department of computer science and technology, Beihang university, 37 xueyuan road, 100083 Beijing, China
{Liuhai,Chenjian,slma}@nlsde.buaa.edu.cn

Abstract. In view of the requirements for high reliability, high time accuracy and complicated time constraints in emergency response workflow systems, a time constraints modeling method is proposed to meet these requirements. The method decomposes the time constraints model into three views, namely the resource view, the control view and the time view. In the resource view, time constraints prediction of resources capability is introduced to improve the accuracy of time prediction. In the control view, the Equivalent Mechanism and Exception Handling Mechanism are introduced to resolve time constraint conflicts and improve the success rate of the process execution. Finally, in the time view, a time flow diagram is constructed from the source view and the control view according to the time flow graph inference rule. In the time flow diagram, the runtime collision detection and resolution are preceded for both definition-time phase and runtime phase. Through the relevant comparison and experiment evaluation, the model is proved to be able to support the automatic detection and resolution for runtime time constraints conflict and improve the time accuracy and success rate of the process execution.

Keywords: emergency response workflow, view, time constraints, conflict detection, conflict resolution.

1 Introduction

With the rapid development of the Internet of Things (IOT), and the growing demands of data management, data mining, data analysis in the application layer of IOT, the business data in both intra-industry and inter-industry need to be integrated. The integration of databases in the various distributed and heterogeneous subsystems has become a hot research topic. As the technical basis of data integration in IOT, Data middleware is the key technique to the integration and management of the distributed and heterogeneous data. Therefore, the study on data middleware has important significance for data integration.

In recent years, with the rapid development of Internet and information technology, more and more disaster emergency-response disposal system applied in the field of disaster warning and disposal. Emergency-response workflow model is the key technology of disaster emergency-response disposal system. Traditional disposal of

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 149–159, 2013.

writing has the following deficiencies: existing text description is not clear, the responsibility divided not clear, fuzzy execution flow, etc. Workflow model has the advantages of systematization, visualization and responsibility trenchant. Time management is vital for emergency disposal of workflow model, it is the key to an effective modeling and validation, which is one of the most challenging research subject. For example, in international earthquake relief disposal system, when the earthquake happens, the process of fast automatic data collection, analysis, reporting and feedback will be carried out. Emergency response workflow model helps to carry out rapid and accurate rescue plan for the system.

The complexity of emergency response workflow modeling and instantiate makes general workflow modeling technology does not apply to emergency disposal process modeling. Compared with other workflow models, complexity of the time property in emergency response workflow model is manifests in following points: 1.Time constraints of the activities in emergency response workflow model include latency, duration, the earliest start time, latest end time, etc. How to achieve these time constraint modeling is an important issue in the study of emergency disposal of the workflow. 2.Emergency response disposal belong to time sensitive business, the time control of the process must be accurate. Meanwhile, the emergency response workflow requires for high reliability. Reasonable handling runtime conflicts of time constraints, to ensure the process will not die easily, improve the success rate of the process execution is an important issue in the study of emergency disposal of the workflow..

Current research on workflow with time restriction[1]-[5] has no clear description of resource capability and analysis of time constraints. By introducing resources capability time, emergency response workflows carries on reasonable time constraint conflict detection and process adjustment, so as to improve the time accuracy and success rate of the entire business process. In the mean time, current research on time constraint verification and conflict resolution[6]-[7] focused primarily on time constraint verification, activities schedulability and performance analysis. Reasonable automatic conflict resolution mechanism are not introduced to deal with runtime conflicts. Automatic conflict resolution mechanism will greatly improve the success rate of the workflow as well as the reliability of business processes in emergency response workflow management system.

According to the above problem, this paper propose a time constraints modeling method to meet these requirements. The method decomposes the time constraints model into three views, namely the resource view, the control view and the time view. In the resource view, time constraints prediction of resources capability is introduced to improve the accuracy of time prediction. In the control view, the equivalent mechanism and exception handling mechanism are introduced to resolve time constraint conflicts and improve the success rate of the process execution. In the time view, the time constraint is divided into two kinds. One is the earliest finish time, limiting task cannot complete within a period of time when the target task is finished. The other one is the latest finish time, limiting task must be complete in a period of time after the target task is finished. By converting time constraints into unified time flow graph, to detect and handle time constraints conflict statically and dynamically.

The rest part of the paper is constructed as follows: Section 2 gives a detailed instruction of Emergency-Response Timing Constraint Workflow Net. Section 3 construct time flow graph. Section 4 introduce the time constraint conflict detection algorithm. Section 5 introduce the time constraint conflict resolution mechanism. Section 6 conducts experiments and evaluation of ETCWM. Section 7 concludes the paper.

2 Emergency-Response Timing Constraint Workflow Model

Definition 1. Emergency-Response Timing Constraint Workflow Net(ETCWN)

ETCWM is defined as seven tuple (P, T, F, M, TC, TA, FC). It satisfies:

1) (P,T,F) is a model inside based on Petri net work. P express library set, T express the change set, F express the arc with direction, $F \subseteq P \times T \cup T \times P$。

$x \in P \cup T$ is one element of WN, $\cdot x = \{y|(y \in P \cup T) \wedge (y,x) \in F\}$ is a pre-set for x ; $x \cdot = \{y|(y \in P \cup T) \wedge (x,y) \in F\}$ is the rear set for x。

A Petri net is a working inside when and only when meet the following two conditions：1. There are two special library I and o, library i is source library: $\cdot i = \emptyset$, library o is Tam library：$o \cdot = \emptyset$; 2.If a change t^* is add to connect o and i, makes $\cdot t^* = \{o\}$ and $t^* \cdot = \{i\}$ to get PN^*, then PN^* is of strongly connected.

2) M is a sequence of state: $M = (M_0 M_1 M_2 \ldots M_n)$, Mi is a m dimensional vector: $M_i = (tp_0, tp_1, tp_2, \ldots, tp_m)$, M is the number of elements in P, tp_i is the token in library p_i, M is the execution of a workflow instance, M_i is execution state of a workflow instance.

3) TC is a description of the explicit time constraints demand of emergency disposal, defined as a mapping from the changes to a time interval $[\alpha, \beta]$, $TC: (T_1, T_2) \to [\alpha, \beta], T_2 \in T, T_1 \in T \cup \{T_\emptyset\}, 0 \le \alpha \le \beta < \infty$, recorded as $T_1 \prec_\alpha^\beta T_2$, α is the earliest triggering time of T_2 when $T_1$is finished, β is the latest triggering time of T_2 when $T_1$is finished. If $T_1 = T_\emptyset$,then α is the earliest triggering time of T_2 when T_1 began, β is the latest triggering time of T_2 when T_1 began.

4) TR is a description of explicit requirements of resource constraints in emergency disposal, a mapping from changes to a resource model. $TA: T \to \{RS\}$. RS is a resource model, including staff, supplies, etc.

5) FC is a mapping from arc to transfer conditions Condition, $FC: F \to Condition$; Condition is a Boolean expression, represent token transfer conditions of arc.

6) Changes can not only be the process main body, but also be a subject in logic relationships, such as conditions determine and branch type judgment, etc.

7) Change t can be activation , if and only if each library in its pre-set has at least a token and meet the token transfer conditions. when an activated t is triggered, it will take one token in very library in its pre-set and generate one token and pass to all library which meet the transfer conditions in its rear set.

In the modeling process of an emergency disposal, we use change represent an operation, library represent the operation state after it is finished, the arc from change to library represent a task to complete into the corresponding processing, use the arc from

library to change represent engine in response to the current state and to do next operation, use the function on the arc to represent different branches of execution condition.

Based on the emergency disposal time constraint workflow model, this paper proposed the concept of three views: control view, time view, resource view. Time View: workflow execution time information and time constraints, Namely the start and end time of each activity, time constraints between the various activities and the activities of the time limit to complete; Control View: the schedule and go to state of workflow instance, such as where is the execution, what is the next to perform. Resource View: resources assignment in workflow model.

In the process of time constraint conflict detection, firstly set up rules by time constraint flow graph, construct time constraint flow diagram using the logical relationship and time constraint resource competence model in Resource View. Then to detect time constraint conflict by time constraint conflict detection algorithm. After found time constraint violation, Time View feedback the information of violations to Control View, who let the process to carry on the back. After process of equivalent substitution and exception handling, to continue a new round of conflict detection time constraints. Through iterative adaptive adjustment to realize time constraint conflict resolution. As shown in Fig.1.

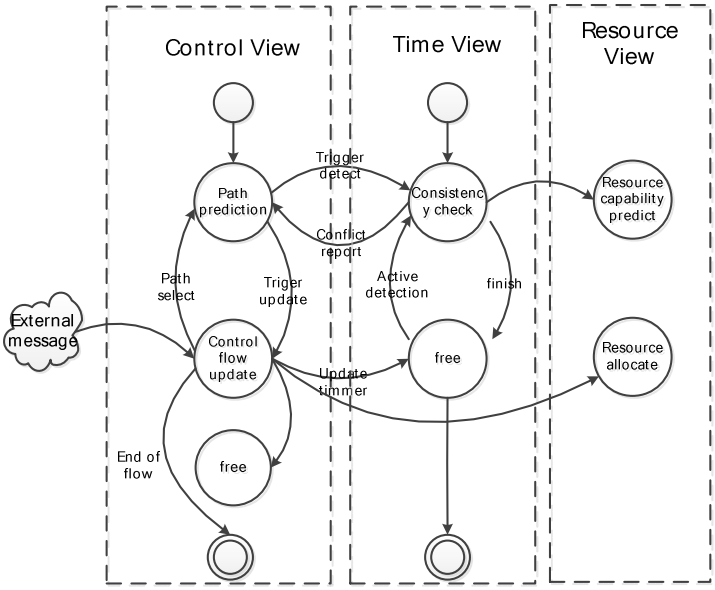


Fig. 1. State diagram of time constraints conflict detection and exception handle

3 Structure and Deduction of Time Flow Graph

Time view represents time constraints information of all activities in workflow instance, including the relations between a sequence of activities, explicitly tasks earliest and latest finishing time, capability time. Time view calculation is done in the

Time Flow Graph. All kinds of time constraints information is organized in Time Flow Graph to detect time constraint conflict by time constraint conflict detection algorithm.

Definition 2 Time Flow Graph. Time Flow Graph is constitute by a triple:$\Sigma = (P, V, E)$, P is a node t_i which is corresponding to the change. V is the directed edge $v_{i,j} \to (t_i, t_j)$ between the changes. E is the time constraint $e_{i,j} \to t_i \prec_{\alpha}^{\beta} t_j$ which changes activated.

Definition 3. Time constraint consistency: If and only if any constraint $e_{i,j} \to t_i \prec_{\alpha}^{\beta} t_j$ who meet the requirement $\alpha < \beta$ is in time constraints consistency, if not, it is time constraint conflict.

The execution time of the work is the determinants of whether time constraints of preset process can be satisfied. Traditionally when a work is pushed to the execution of the work list, the workflow management system may have no idea how long it will take the work to finish. By introducing the concept of capability time, Time View can establish complete Time Flow Graph based on the resource time constraint to detect and deal with time constraint conflict, which improve the accuracy of the time constraints analysis.

Definition 4. Capability Time[1]. In change t, according to the resource constraint, we forecast the shortest execution time recorded as TET(t) and the longest execution time recorded as TLT(t), TET(t) and TLT(t) are capability time of change t. Resources capability time meet the requirement of implicit time constraints in emergency disposal workflows.

Rule 1. structure rule of time flow graph

1) Dealing with changes: copy the collection of changes in ETCWN model and paste to Σ, as $P = T$. Then add $T_{\emptyset}$ as a starting node to P, as $P = P + \{T_{\emptyset}\}$.

2) Dealing with explicit time constraints: To any changes t_i, t_j in ETCWN model, if $\exists TC = (< t_i, t_j >, [\alpha, \beta]) \in TC$, build directed edge $V' = (t_i, t_j)$ between t_i and t_j ,namely $V = V + V'$, Then add time constraint $E' = t_i \prec_{\alpha}^{\beta} t_j$ to E, namely $E = E + E'$.

3) Dealing with resource capability time: To any changes t_i, t_j, if there is library p, $p \in t_i \cdot \wedge p \in \cdot t_j$, build directed edge $V' = (t_i, t_j)$ between t_i and t_j,namely $V = V + V'$; according to the resource capability time, we know that after t_i is finished, t_j must be finished TET(t_j) later but in TLT(t_j), namely $E' = t_i \prec_{TET(t_j)}^{TLT(t_j)} t_j$, add E' to E, namely $E = E + E'$.

Rule 2. Extended rule of time constraints derivation

1) Serial structure: Two serial time constraints $t_i \prec_{\alpha_j}^{\beta_j} t_j$, $t_j \prec_{\alpha_k}^{\beta_k} t_k$, it can be derived as $t_i \prec_{\alpha_j + \alpha_k}^{\beta_j + \beta_k} t_k$

2) Parallel structure: Two parallel time constraints $t_i \prec_{\alpha_1}^{\beta_1} t_j$, $t_i \prec_{\alpha_2}^{\beta_2} t_j$, if the relationship is "with", then $t_i \prec_{\max(\alpha_1, \alpha_2)}^{\min(\beta_1, \beta_2)} t_j$; if the relationship is "or", then $t_i \prec_{\min(\alpha_1, \alpha_2)}^{\max(\beta_1, \beta_2)} t_j$.

[1] Resources waiting time has been included in the task execution time.

4 Time Constraint Conflict Detection Algorithm

Time view represents time constraints information of all activities in workflow instance, including the relations between a sequence of activities, explicitly tasks earliest and latest finishing time, capability time. Time view calculation is done in the Time Flow Graph. All kinds of time constraints information is organized in Time Flow Graph to detect time constraint conflict by time constraint conflict detection algorithm.

In a workflow with the time constraint, explicit time constraints, process temporal logic and resource capability time may lead to conflict. For instance, a work will be finished in 10 minutes, however, we concluded by temporal logic and resource capability time that the work can be finished after 20 minutes, the earliest. In order to guarantee the consistency of all kinds of time constraints, in this section we provide time constraint conflict detection algorithm in both definition time and runtime to realize real time constraints conflict detection.

Algorithm 1. Conflict detection algorithm in definition time.

Input: ETCWN=(P,T,F, M,TC,TA,FC)

Output: Boolean, whether the process meet the requirements of time constraints consistency or not

1)Establish time flow graph in accordance with rule 1.

2)Number the nodes in Σ and initializes the queue Q. Put the starting node t in Q, number t as 1. Cycle out of the queue when Q is not empty, t'=DEQ(Q), put subsequent of t' into the queue in turn , number++. When Q is empty, end the cycle.

3)Check the consistency of time constraint. Mark the starting time of the $T_{\emptyset}$ as 0, check all the nodes t_j(i=2,…,{V}) in turn: to every precursor $t_i \in \cdot t_j$ of t_j, calculate the cumulative time constraint $e_{i,j} = t_i \prec_{\alpha}^{\beta} t_j$ by parallel rule 2, calculate every time constraint $e_{i,j} = t_i \prec_{\alpha}^{\beta} t_j$ by serial rule 2.If $\alpha > \beta$, there is conflict ,end the algorithm; if not, continue to calculate until all the nodes are checked. if there is no conflict, end the algorithm and the time constraint in ETCWN is consistent.

Let use a detailed example to show the execution of algorithm 1, including construction, extension, and detection of time flow graph. As shown in Fig 2, $t_3$must be finished earliest 2 seconds after $t_1$is finished, and latest in 3seconds after $t_1$is finished, namely: $t_1 \prec_2^3 t_3$, t_5 must be finished in 5 minutes since the workflow begin. t_4 must be finished in 4 minutes since the workflow begin, namely: $t_{\emptyset} \prec_0^5 t_5$, $t_{\emptyset} \prec_0^4 t_4$

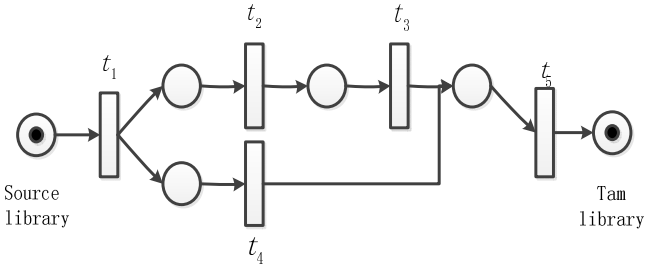


Fig. 2. The flow chart

Table 1 shows resource capability time of every work.

Table 1. Resource capability time constraint

Change	Time interval
t_1	[1,2]
t_2	[1,3]
t_3	[2,3]
t_4	[3,4]
t_5	[3,5]

According to rule 1(3), time constraint set is get: $\{t_\emptyset \prec_1^2 t_1, t_1 \prec_1^3 t_2, t_2 \prec_2^3 t_3, t_1 \prec_3^4 t_4, t_3 \prec_3^5 t_5, t_4 \prec_3^5 t_5\}$, combined with initial constraints set $\{t_1 \prec_2^3 t_3, t_\emptyset \prec_0^5 t_5, t_\emptyset \prec_0^4 t_4\}$, constitutes the final time constraints set of time flow graph. The time flow graph is shown in figure 3

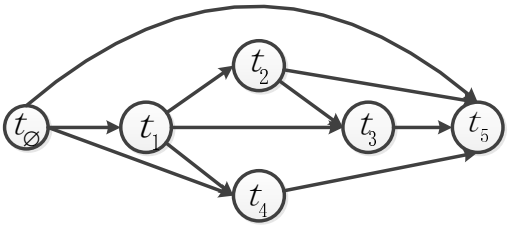


Fig. 3. Time flow graph

On the basis of the time flow graph in Fig.3, we do time constraint derivation following rule 2, do constraint conflict detection using algorithm 1. Derivation process is shown in table 2:

Table 2. Font sizes of headings. Table captions should always be positioned *above* the tables.

No	constraint	premise	rules
1	$\boldsymbol{t_\emptyset \prec_0^4 t_4}$	Time constraint definition	rule1(2)
2	$\boldsymbol{t_\emptyset \prec_0^6 t_5}$	Time constraint definition	rule1(2)
3	$\boldsymbol{t_1 \prec_2^3 t_3}$	Time constraint definition	rule1(2)
4	$\boldsymbol{t_\emptyset \prec_1^2 t_1}$	Resource capability time	rule1(3)
5	$\boldsymbol{t_1 \prec_1^3 t_2}$	Resource capability time	rule1(3)
6	$\boldsymbol{t_2 \prec_2^3 t_3}$	Resource capability time	rule1(3)
7	$\boldsymbol{t_1 \prec_3^4 t_4}$	Resource capability time	rule1(3)
8	$\boldsymbol{t_3 \prec_3^5 t_5}$	Resource capability time	rule1(3)
9	$\boldsymbol{t_4 \prec_3^5 t_5}$	Resource capability time	rule1(3)
10	$\boldsymbol{t_\emptyset \prec_2^5 t_2}$	conclusion 4、5	rule2 serial
11	$\boldsymbol{t_1 \prec_3^6 t_3}$	conclusion 5、6	rule2 serial
12	$\boldsymbol{t_1 \prec_3^3 t_3}$	conclusion 3、11	rule2 parallel with
13	$\boldsymbol{t_\emptyset \prec_4^5 t_3}$	conclusion 4、12	rule2 serial
14	$\boldsymbol{t_\emptyset \prec_4^6 t_4}$	conclusion 4、7	rule2 serial

Table 2. (*Continued*)

15	$\boldsymbol{t_{\emptyset} \prec_4^4 t_4}$	conclusion 1、14	rule2 parallel with
16	$\boldsymbol{t_{\emptyset} \prec_7^{10} t_5}$	conclusion 8、13	rule2 serial
17	$\boldsymbol{t_{\emptyset} \prec_7^9 t_5}$	conclusion 9、15	rule2 serial
18	$t_{\emptyset} \prec_7^6 t_5$	conclusion 16、17、2	rule2 parallel with

We can learn from the form 2 that conclusion 18 $t_{\emptyset} \prec_7^6 t_5$ is in violation of the time constraints. According to definition 3, this process does not meet the time constraints, feedback conflict information to control view to made adjustment to the workflow.

Algorithm 2. Conflict detection algorithm in runtime.

Input: time flow graph in algorithm1 and the triggering time of all the changes already triggered.

Output: Boolean, whether the process meet the requirements of time constraints consistency or not

1)Update the time constraints. If both the pre-set and rear set of changes of constraint are triggered, remove the constraint from the collection. If neither the pre-set nor rear set of changes of constraint are triggered, remain the constraint unchanged. If pre-set of changes are triggered and rear set of changes are not, to $e_{i,j} \rightarrow t_i \prec_{\alpha}^{\beta} t_j$, calculate the time which the pre-set of changes take(current time minus pre-set changes trigger time), $e_{\emptyset,j} \rightarrow t_{\emptyset} \prec_{\alpha-T}^{\beta-T} t_j$. For the time constraints generated by rule1(3), recalculate the resource capability time and update.

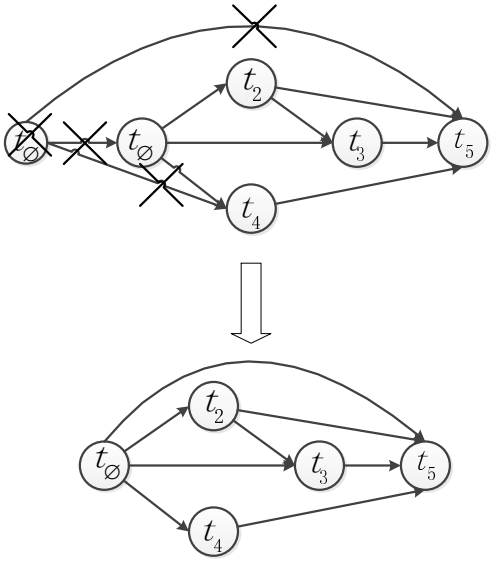


Fig. 4. Dynamic update process in time flow graph

2) Same steps with algorithm 1.

Assuming that in the instance in Fig.3, t_1 is triggered, $t_\emptyset$'s trigger time is 0, t_1's trigger time is 2, current time is 2, then the original time constraints $\{t_1 \prec_2^3 t_3,\ t_\emptyset \prec_0^5 t_5,\ t_\emptyset \prec_0^4 t_4\}$of workflow turn to new constraints $\{t_\emptyset \prec_2^3 t_3,\ t_\emptyset \prec_0^3 t_5,\ t_\emptyset \prec_0^2 t_4\}$, the update process of time flow graph are shown in Fig.4.

5 Automatic Conflict Resolution Mechanism

The setup of both the hardware and software environment for the test are listed as follows:

To support time constraints automatic conflict resolution, this section propose a constraint violation handling mechanism based on equivalent replacement. The concept of equivalent structure and exception handling structure are introduced to realize automatic conflict resolution. equivalent activity can replace the original activity when the original activity failed to perform or conflict. Exception handling activities are based on the equivalent activity to record abnormal warning, if equivalent activity still failed to perform or conflict, then execute exception handling activities.

In the workflow process, priority to perform normal activities, if normal activity conflict, go back from this node and perform the equivalent activity, if the equivalent activity still conflict, check exception handling activity, if exception handling activity conflict, continue to go back until cannot go back and alarm the error.

We add equivalent activity and exception handling activity to the instance in section 3. The time flow graph is shown in Fig.5. Dotted line represent the normal activities in work flow.

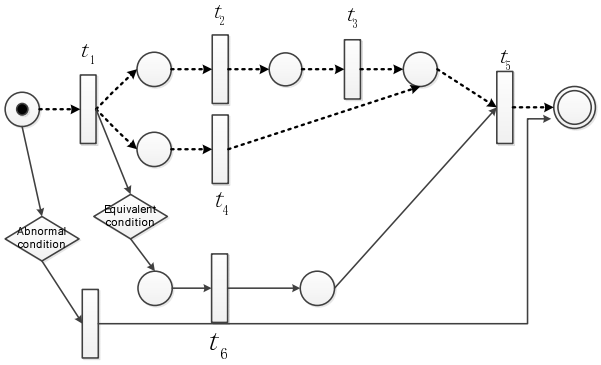


Fig. 5. The flow chart with equivalent structure and exception handle structure

If it come across time constraint conflict at t_5, go back to t_1 and perform the equivalent activity, the process is shown in Fig.6. Dotted line represent the check and perform of the equivalent activity.

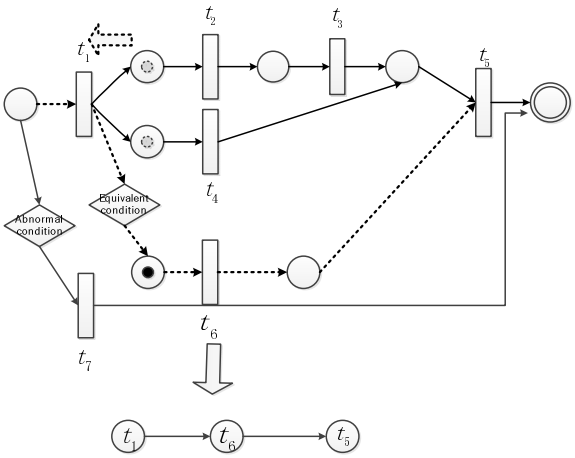


Fig. 6. The flow chart of go back and select the new path

The path of performing equivalent activity meet the time constraint of the workflow, through automatic adjustment to avoid the abortion caused by failure of normal activities in the workflow. Equivalent activity and exception handling activity provides effective means for normal workflow execution, greatly improve the success rate of execution

6 Time Constraint Conflict Detection and Resolution Experiment

Firstly, we randomly selected workflow that have no conflict resolution, it meet the time constraints in define time but time constraint conflict could occur at run time. Run the workflow for 1000 times, record number of successful executed workflow as N. secondly, we add conflict resolution mechanism in section 4 and 5. Run the workflow for 1000 times again, record successful executed workflow number as M, the equivalent substitution routing number as M1, exception handling path selection number as M2. Finally, repeat the experiment for 100 times to get the average result. The result are shown as follows.

Table 3. Confict resolution experiment results

Experiment	Scucess	failure	Sucess rate
With conflict resolution	979	21	97.9%
Without conflict resolution	863	137	86.3%

In the workflow with conflict resolution mechanism, equivalent activities perform 83 times while exception handling activities perform 33 times. It can be concluded from the experimental results that the equivalent replacement and the exception

handling mechanism, as a supplement for possible runtime time constraints, have significantly improved the success rate of workflow execution.

The experimental results shows that ETCWM can dynamically evaluate the runtime time constraint conflict in workflow, provide automatic conflict resolution which greatly improve the success rate of emergency-response workflow.

7 Conclusions

This paper proposed a time constraints modeling method to meet the requirements for high reliability, high time accuracy and complicated time constraints in emergency response workflow systems. On the basis of this model, the time constraints flow diagram is built, conflict detection and resolution algorithm is derived.

The method decomposes the time constraints model into three views, namely the resource view, the control view and the time view. In the resource view, time constraints prediction of resources capability is introduced to improve the accuracy of time prediction. In the control view, the Equivalent Mechanism and Exception Handling Mechanism are introduced to resolve time constraint conflicts and improve the success rate of the process execution. Finally, in the time view, a time flow diagram is constructed from the source view and the control view according to the time flow graph inference rule. In the time flow diagram, the runtime collision detection and resolution are preceded for both definition-time phase and runtime phase. Through the relevant comparison and experiment evaluation, the model is proved to be able to support the automatic detection and resolution for runtime time constraints conflict and improve the time accuracy and success rate of the process execution, which provides strong support for emergency disposal.

References

1. Tan, G., Xiao, R.: Dynamic Verification and Predication of Temporal Constraints for Workflow Based on Petri. Computer Measurement & Control 15(12), 1801–1803 (2007)
2. Li, H., Fan, Y.: Overview on Managing Time in Workflow Systems. Journal of Software 13(8), 1552–1558 (2002)
3. Li, T., Zhong, S.: Time Constraints Analysis of Task Scheduling for Colored Timed Workflow Net. Computer Engineering and Applications (8), 16–19 (2006)
4. Pang, H., Fang, Z., Zhao, Y.: Simplification analysis and schedulability verification of timing constraint workflow model. Computer Integrated Manufacturing System 14(11), 2217–2223 (2008)
5. Deng, H., Zhou, C., Gao, Y., Liu, L.: A Temporal Constraints Workflow Model Based on Petri. Journal of Changchun University of Technology 30(1), 59–63 (2009)
6. Li, H., Fan, Y.: Workflow Model Analysis Based on Time Constraint Petri Nets. Journal of Software 15(1), 17–26 (2004)
7. Zeng, Q., Duan, H.: Modeling and time mining for workflow involving time factors. Computer Integrated Manufacturing System 11(6), 855–860 (2005)

Study on the Application of GIS in Comprehensive Risk Assessment of Hazardous Chemical Plants

Shaobo Zhong[1,2,*], Yi Liu[1,2], Fei Wang[3], and QuanYi Huang[1,2]

[1] Department of Engineering Physics, Tsinghua University, 100084 Beijing, China
[2] Institute for Public Safety Research, Tsinghua University, 100084 Beijing, China
[3] Shenzhen Graduate School, Tsinghua University, Shenzhen, China
zhongshaobo@tsinghua.edu.cn

Abstract. Comprehensive risk assessment is very important work for effective prevention and control of chemical accidents in hazardous chemical procedure industries. This paper first proposes a perspective and workflow for comprehensive risk assessment of hazardous chemical plants. Several key components: case base, vulnerability analysis model, indicator system based risk assessment model are included in the workflow. As a kind of professional tool for geographical information manipulation, GIS can provide strong support for analysis of case data, especially some spatial location relevant data, which are in the majority among all case information. Next, GIS can provide support for implementation and run of vulnerability analysis models. Further, with strong spatial analysis capability of GIS, one can carry out Impact evaluation of a chemical accident. Thus, GIS can provide versatile support in comprehensive risk assessment of hazardous chemical plants. In this paper, we will affirm: Spatial analysis and visualization of GIS can facilitate the management of case base and provide fundamental support for vulnerability analysis and risk assessment models.

Keywords: Comprehensive risk assessment, Hazardous chemical plant, GIS, Case base, Vulnerability analysis.

1 Introduction

In chemical procedure industries, comprehensive risk assessment is universally implemented program for prevention and control of potential accidents. There are at least two approaches to comprehensive risk assessment: a technological source-oriented one and a receptor-oriented one [1]. In recent decades, many methodologies have been developed to undertake a risk analysis on an industrial plant. Tixer et al identified more than 60 risk analysis methodologies, which can include up to three phases (identification, evaluation and hierarchisation) [2]. According to their technological characteristics, These methods can be ranked in six classes [1], based on the combination of four usual criteria (qualitative, quantitative, deterministic and probabilistic). Though these methods have been widely used for risk analysis of industrial plants, there are some limitations and difficulties with these methods [3].

* Corresponding author.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 160–170, 2013.

In theory, comprehensive risk assessment aims at combining source and receptor. For technological risks, this includes both plant and external safety considerations; aspects of regulatory assessment, planning, training, and emergency management, but also risk communication and the integration of physical and chemical criteria with environmental, human health, and socioeconomic considerations. In addition, it specifically aims at integrating the power of dynamic simulations models as a major tool of risk assessment.

Different from some engineering research, dangerous chemicals incidents occur at random and dynamically. It is also difficult and often impossible for us to make experiment of accidents. So collection of historical cases is almost the only way for data of chemical accidents. On the other hand, we also believe that objective laws of incidents are hidden under abundant cases. With increasingly developed methods and technologies, especially information technologies and data mining methods, the substantial principle dominating the accidents can be disclosed. Obviously the collection and analysis of cases need to be studied deeply.

For cases, in addition that some general data need be obtained, some spatial and temporal data, and some environment-related data also should be collected for better analysis of them. In these additional data, most are geo-referenced which can be well handled with GIS (Geographical Information System). For example, the occurrence location of a case can be parsed from the natural language form of it with the geocoding function from GIS, some elements of a case can be represented with feature editing or symbol drawing functions of GIS, and GIS can also provide support for input data retrieval and evaluation model running and so on.

The structure of this paper is organized as follows. Firstly, the core thoughts centered on case study of chemical accidents are explained and the analysis flow based on case study is illustrated. As a result, several essential components including case management system, hazardous chemical base and dynamic model are presented. Next, the necessity and possibility of the application of GIS in the integrated risk analysis process is explored. Lastly, the paper concludes with the promising application of GIS in construction of comprehensive risk assessment system of hazardous chemical plants and several aspects worth further improvement.

2 Perspective and Workflow

We think the comprehensive risk of a chemical plant depends on not only the internal status and characteristics but its ambient environments including natural and man-made ones. Furthermore, management and response capabilities also affect the result of risk assessment. As risk is a combination of probability and consequence, the stronger the management capability is, the lower the occurrence probability of accidents caused by a chemical plant, and the more efficient and effective response to an accident is, the less serious the consequence caused by an accident is.

There are two types of classic approaches to risk assessment: one is indicator system based method and the other is dynamic model method. As mentioned above, there are some factors which dominate the final risk. These factors can be organized

hierarchically into a tree structure. For example, in the first hierarchy, we can consider physical and chemical properties of a chemical plant, natural and man-made environments around the plants and the emergency management capability are three kinds of factors that affect the resultant risk. In the second hierarchy, the emergency management capability can consist of preparedness, mitigation, monitoring, response and recovery etc. sub-factors. On the other hand, risk can be analyzed through most likely accident identification and corresponding consequence calculation. Generally, the consequence can be calculated with specific dynamic model.

It is natural to integrate the two approaches in comprehensive risk assessment. In scientific research, data can provide evidences while a mechanism model describes a dynamic process of a potential accident. Therefore, our perspective is that an integrated approach which can combine both of them will be a kind of more promising one. According to this thread, we propose a flowchart of comprehensive risk assessment of chemical plants as shown in Figure 1 [11].

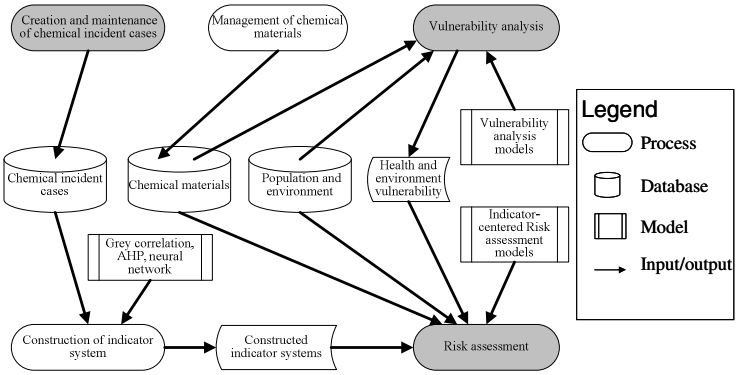


Fig. 1. The flowchart of comprehensive risk assessment system of chemical plants based on cases

The main steps of the analysis are as follows.

(1) Creation and maintenance of chemical incident cases

In this operation, the system reshapes chemical incident cases obtained from all kinds of sources and stored them with structured form through interaction with human. The stored cases can be updated if necessary. Furthermore, the system can import cases in batch from external files. These files are created by manual processing or exported by other relevant systems. In case base, the case is stored with formal structure. In addition to exporting textual form of electronic document or paper, the case can be presented dynamically. Query, locating, etc. operations can also be executed on case base.

(2) Management of chemicals

This work maintains a list of chemicals. The information of a chemical includes physical and chemical properties of it, environment, health and treatment actions etc.

(3) Vulnerability analysis

Vulnerability analysis will choose a simulation model to predict the effect of toxic, flammable or explosive chemicals on population health or environment. In this analysis, the population or environment data are required.

(4) Construction of indicator system

The construction of indicator system is a complex process. The core tasks in this process include indicator choice, weight calculation etc. some mathematical methods for indicator system analysis are needed such as grey correlation, fuzzy logic, artificial neural network etc.

(5) Risk assessment

This work is to implement some risk assessment models which are based on indicator system. In process 4), some alternative indicator systems are proposed. And in this process, the models will choose an indicator system and request user input or extract data from database to calculate risk of chemical places. Then, given a certain rating rule, the risk rate of chemical places will be identified.

3 The Role Played by GIS

Geographical information system is a comprehensive software package for manipulation of spatial location relevant information. Its core functions include spatial data retrieval, spatial data processing, spatial data storage, spatial analysis, spatial visualization etc. In process of comprehensive risk assessment of hazardous chemical plants, plenty of data, such as population distribution, environmental coverage, rescue station, emergency materials and health resource and so on are spatial location relevant. GIS provides a strong substantial tool for the manipulation of these data. The geo-database of GIS is an extended database management system which is built up based on a proprietary database management system such as Microsoft SQL Server and Oracle and can be used for storing and maintaining these data.

Given the requirements of analysis activities in fig. 1 and the capability of GIS, the possibility of the application of GIS in comprehensive risk assessment of hazardous chemical plants are presented as follows.

3.1 Retrieval of the Geographical Properties of a Case

Originally, a case is presented with papery or electronic document where some geo-referenced information is described in form of natural language. For example, the location of a case is expressed as some street, some road. This kind of expression of geographical features is difficult to be processed and analyzed with computer and graphically cognized by man. Some functions of GIS can be utilized in retrieval of the geographical properties of a case.

First of all, geocoding function in GIS can be applied to interpret the natural language description of the occurrence location of a case. Geocoding is an essential function in GIS which works with spatial database and natural language interpretation techniques. It can retrieve accurate occurrence location of a case with latitude-longitude form. Table 1 presents the general workflow of geocoding of the occurrence of a case with GIS.

Table 1. The general workflow of geocoding with GIS

No.	Step	Description
1	Original address is presented	Geocoding starts with a specified address in a natural language form.
2	Address with natural language form is parsed	When an address locator parses an address, it dissects the address into its address elements based on the style of the address locator.
3	Address locator searched	Once the address locator has parsed the address and generated a list of values for each address component, it finds features in the locator that match the elements of the input address based on a set of search criteria.
4	Score of each potential match established	When the address locator has generated a set of potential location candidates, it scores each candidate to determine how closely each one matches the address.
5	List of candidates filtered	Once each candidate is scored, the address locator generates a set of candidates that are potential matches for the address based on the minimum candidate score set for the address locator.
6	Best candidate matched	The address locator ranks the candidates by score and an address is matched automatically to the candidate with the highest score.
7	Matched feature indicated	The feature that corresponds to the best candidate is used for generating the location for the matched address.

Next, the spatial feature editing can be used to mark out the affected zone by a case and its evacuation zone, if any. The affected zone of a case is the domain of influence caused by an accident happening in a hazardous chemical plant while the evacuation zone is a temporal assembly area for protecting the affected people from harming and providing basic life support for them. Both of the affected zone and the evacuation zone can be bordered according to some textual description of the accident, which is easy given a detailed record of the case. There are several basic functions of spatial feature editing that can be used to draw the zones including draw circle, draw ellipse, draw rectangle, draw polygon. Illustrative symbol plotting is especially useful for representing situation, tendency, tactics etc. on a map. This tool derived from GIS can be used to render some essential elements of a case such as disaster scenario, evolvement and response and so on. Figure 2 illustrates an example of the representation of basic elements with some spatial feature editing tools and normalized symbols which are able to be plotted through tools integrated in a proprietary GIS software package or implemented through second development techniques provided by it.

Furthermore, one can use spatial query function to retrieve important information around the location of the accident. The information includes risk sources, protection objects, population, economics etc. Fig. is an illustration of geo-referenced information retrieval of a case with GIS. The process is as follows.

Step 1: a query feature is drawn manually or automatically;

Step 2: In background, GIS fetches the objects meeting certain conditions according to the marked area and object types;

Step 3: the fetched data is exchanged into input form for case information maintenance;

Step 4: If necessary, the user edits the fetched data;

Step 5: the user confirms and submits the case information into case base.

Through interaction between case information maintenance application and GIS, the geo-referenced information of a case can be retrieved and saved to case base finally. This process will largely increase the efficiency of case maintenance.

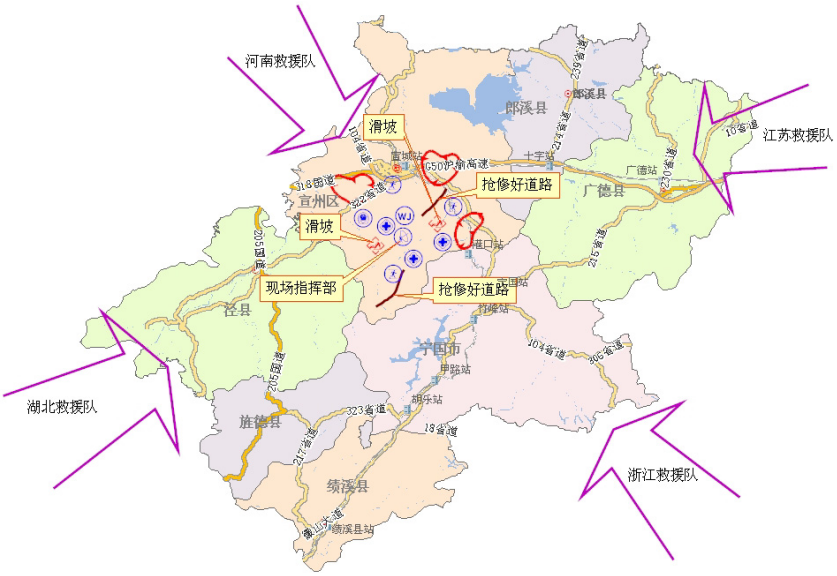


Fig. 2. An illustration of case representation with feature editing and symbol drawing from GIS

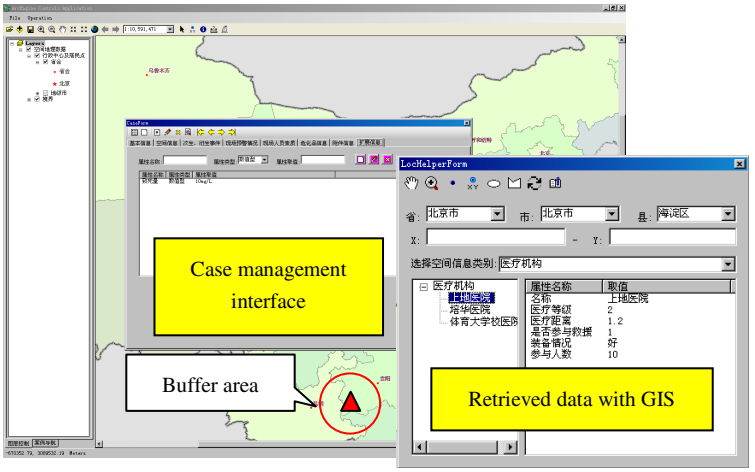


Fig. 3. A screenshot of case data retrieval function based on spatial query function of GIS

3.2 Support for Implementation and Run of Models

As seen, vulnerability analysis models and indicator-centered risk assessment models are going to be required in the course of comprehensive risk assessment of hazardous chemical plants. Firstly, for vulnerability analysis, some models, no matter hazardous material leakage models or explosion effect analysis models, need geo-referenced parameters as input. Secondly, GIS can provide support for the run of the model, especially in case that the processing of the model involves some spatial analysis functions. Finally, the output of a model is more effective and vivid if they can be presented on map directly. Fig. illustrates the input interface and the analysis result of a leakage analysis model presented on a map.

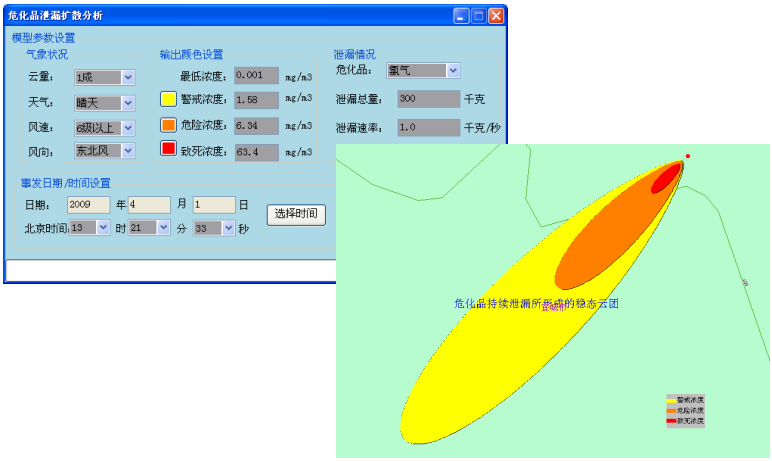


Fig. 4. Input interface and analysis result presentation of vulnerability analysis models with GIS

3.3 Impact Evaluation of a Chemical Incident

Following the prediction of the analysis models, the question what is the consequence of the incident is going to be asked. The answer can be provided easily leaning upon spatial analysis functions of GIS. With overlay analysis of GIS, one can create an integrated analysis by applying a common scale of values to diverse and dissimilar input. On one hand, analysis models obtain the impact range of the incident which is generally presented with polygon features. On the other hand, some demographic data, economic data and environmental data are collected and held by GIS in a geo-database they often are presented in form of certain feature types (mostly point or polygon types). With overlaying demographic, economic or environmental data with the impact range of the incident, an overall risk can be quantified.

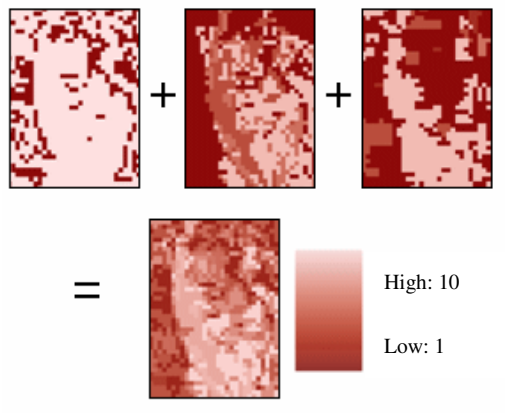


Fig. 5. Impact analysis of a chemical incident with GIS overlay analysis function

3.4 Input Data Retrieval of Risk Assessment Model

As mentioned above, there are two types of models which are necessary for the retrieval of risk assessment results: the dynamic ones and the hierarchy based ones. For the former, three most commonly used models include: fire spread model, explosion impact analysis model and hazardous chemical leakage analysis model. These models require some data as input before and during execution. And some procedural data may be produced and need be stored temporarily. The following table lists some key parameters whose data need be provided for execution of the models.

Table 2. Three most commonly used dynamical models in risk assessment of hazardous chemical plants

Model name	Purpose	Main parameters	Purpose and parameter description
Explosion impact analysis	Analyze impact area for a specified explosion	Input parameters	The central location of the explosion, explosion type, explosion material type and explosion material mass, volume, pressure etc.
		Output results	The spatial distribution of the hierarchical damage areas
Hazardous chemical leakage analysis	Analyze and simulate the dynamic spread process of leakage materials	Input parameters	The leakage location, leakage mass, leakage velocity, leakage pattern, occurrence time, chemical type, wind conditions, cloud condition, weather etc.
		Output results	Space-time distribution of chemical material coverage and thickness.
Fire spread analysis	Analyze and simulate the spread process for a specified fire event	Input parameters	DEM, fire location, wind conditions, fire load etc.
		Output results	The dynamic of fire spread and space-time distribution of fire zone and strength.

In order to carry out the analysis with the indicator based model, one must provide the input data corresponding to the hierarchical indicator system. The following figure shows a typical hierarchical structure of indicator system for the risk of a hazardous chemical plant. As seen, this indicator system is built in view of three phases of emergency management of chemical incidents: preparedness, mitigation and response. Thus, some key factors are filtered out naturally.

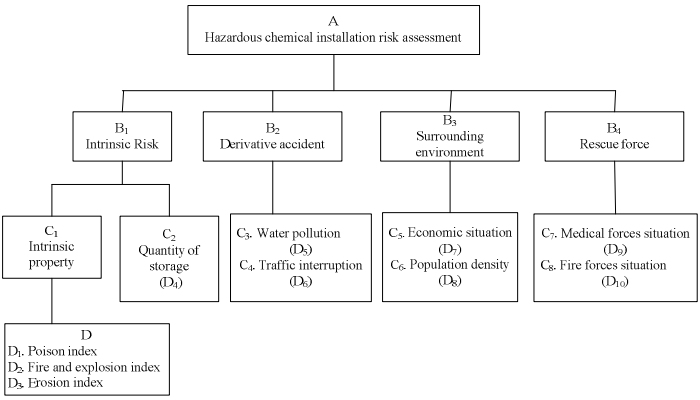


Fig. 6. The AHP model for the risk assessment of chemical installation

With some spatial analysis functions provided by GIS, the input data for many of these indicators can be retrieved from spatial database. The following table lists some indicators and the corresponding retrieval approaches with GIS.

Table 3. The corresponding retrieval approaches of indicator data with GIS

Factor	Functions or Tools	Raw data
Poison index D_1	Attribute analysis of hazardous chemical installations	Attribute data of hazardous chemical installations
Fire and explosion index D_2	Attribute analysis of hazardous chemical installations	Attribute data of hazardous chemical installations
Erosion index D_3	Attribute analysis of hazardous chemical installations	Attribute data of hazardous chemical installations
Quantity of storage D_4	Attribute analysis of hazardous chemical installations	Attribute data of hazardous chemical installations
Water pollution D_5	overlapping analysis (polygon and polyline)	Water bodies with polyline or polygon features
Traffic interruption D_6	overlapping analysis (polygon and polyline)	Traffic roads with polyline features
Economic situation D_7	Zonal statistics	Areal economic data
Population density D_8	Zonal statistics	Areal population data
Medical forces situation D_9	Buffer analysis (around hazardous chemical plants)	Medical force distribution thematic data
Fire forces situation D_{10}	Buffer analysis (around hazardous chemical plants)	Fire force distribution thematic data

4 Conclusions and Discussion

Comprehensive risk assessment is very important work for effective prevention and control of chemical incidents. Chemical cases are the systematical archives of the historical chemical incidents and underlying law of incident evolvement is hidden under abundant case information. Cases are useful data source from where risk assessment can be implemented with some data-mining methods. As a necessary component of comprehensive risk assessment, vulnerability analysis need take physical, chemical characteristics of chemicals as input parameters. Therefore, a chemical base is needed when make vulnerability analysis of health or environment. Furthermore, as a kind of professional tool for geographical information manipulation, GIS can provide strong support for analysis of case data, especially some spatial location relevant data, which are in the majority among all case information.

For some reasons, experiments for accidents are not available. Thus, historical accidents (cases) are the valuable experimental data for analysis of incident objective laws which are hidden under abundant cases. On the other hand, traditional cases have some inherent shortcomings when used for an comprehensive risk assessment of dangerous chemicals. From this paper, we conclude: Case study can provide useful service for risk assessment. Comprehensive and structured cases are good materials for data-mining; spatial analysis and visualization of GIS can facilitate the management of case base and provide fundamental support for vulnerability analysis and risk assessment models.

Acknowledgements. The authors would like to thank the support of the National Natural Science Foundation of China (Key Scientific Problems and Integrated Research Platform for Scenario Response Based National Emergency Platform System, Grant No. 91024032), the National Natural Science Foundation of China (Study on Emergency Decision-making Based on Generalized Knowledge Coupling, Grant No. 70901047), the National Natural Science Foundation of China (Study on Pre-qualification Theory and Method for Influences of Disastrous Meteorological Events, Grant No. 91224004) and the National Natural Science Foundation of China (Study on Systematical Integration Theory and Method of Emergency Technologies for Unconventional Incidents, Grant No. 91024024).

References

1. Fedra, K.: Comprehensive risk assessment and management; overview and state of the art. Journal of Hazardous Materials 61(1/3), 18 (1998)
2. Tixier, J., Dusserre, G., et al.: Review of 62 risk analysis methodologies of industrial plants. Journal of Loss Prevention in the Process Industries 15(4), 13 (2002)
3. Khan, F.I., Abbasi, S.A.: Risk analysis of a typical chemical industry using ORA procedure. Journal of Loss Prevention in the Process Industries 14, 43–59 (2001)

4. Contini, S., Bellezza, F., Christou, M.D., et al.: The use of geographic information systems in major accident risk assessment and management. Journal of Hazardous Materials 78, 223–245 (2000)
5. Suddle, S., Ale, B.: The third spatial dimension risk approach for individual risk and group risk in multiple use of space. Journal of Hazardous Materials A123, 35–53 (2005)
6. Bień, J.D., ter Meer, J., Rulkens, W.H., Rijnaarts, H.H.M.: A GIS-based approach for the long-term prediction of human health risks at contaminated sites. Environmental Modeling and Assessment 9, 221–226 (2004)
7. Papazoglou, I.A., Ale, B.J.M.: A logical model for quantification of occupational risk. Reliability Engineering and System Safety 92, 785–803 (2007)
8. Fabiano, B., Currò, F., Palazzi, E., et al.: A framework for risk assessment and decision-making strategies in dangerous good transportation. Journal of Hazardous Materials 93, 1–15 (2002)
9. Ale, B.J.M.: Risk assessment practices in the Netherlands. Safety Science 40, 105–126 (2002)
10. Khan, F.I., Abbasi, S.A.: Techniques and methodologies for risk analysis in chemical process industries. Journal of Loss Prevention in the Process Industries 11, 261–277 (1998)
11. Zhong, S.B., He, C.H., Wang, Z.B., et al.: Design and implementation of an integrated risk assessment system for chemical places. In: 6th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2009, vol. 5, pp. 533–537 (2009)

A Preliminary Research on Incident Chain Modeling and Analysis

Shaobo Zhong[1,2,*], GuoFeng Su[1,2], Fei Wang[3], Jianguo Chen[1,2], Fushen Zhang[1,2], Chao Huang[1,2], QuanYi Huang[1,2], and Hongyong Yuan[1,2]

[1] Department of Engineering Physics, Tsinghua University, Beijing, China
[2] Institute for Public Safety Research, Tsinghua University, Beijing, China
[3] Shenzhen Graduate School, Tsinghua University, Beijing, China
zhongshaobo@tsinghua.edu.cn

Abstract. In the real world, an incident is often followed by some secondary or derivative incidents which may further trigger some new incidents. The process goes so repeatedly. How to present this kind of chain relationship and formalize cascading effect of these incidents is very important for prevention and control of the complicated incident scenarios like this. This paper preliminarily explores incident chain modeling including the mechanism of incident chain and its construction and presentation. In mechanism, a term meta-force is coined for theoretical research of the incident chain modeling. A general theoretical framework including incident, meta-force and disaster receptor is proposed which can be used for incident chain research. Following the research on construction and presentation of incident chain modeling, the paper further investigates the analysis approach to incident chain based on Bayesian network. As a kind of natural presentation of incident chain, Bayesian network techniques have inherent advantages when used in inference and prediction of a complicated disaster scenario. Concrete applications include risk analysis and decision optimization. The work can be consulted in case of prevention and control of complicated disaster circumstances.

Keywords: incident chain, Bayesian network, GIS, chain breaking & disaster reduction.

1 Introduction

An incident is often followed by some secondary or derivative incidents. For instance, when an earthquake occurs, fire, landslide, hazardous material leakage may be triggered. If not prevented or controlled effectively and efficiently, the consequence caused by these resultant incidents will probably leads to more serious loss than the earthquake itself does. This kind of cascading and interactive effect of incidents is called an incident chain. The incident chain makes the prevention and control of incidents more complicated and difficult. Nevertheless, there exist internal principles that dominate the formation of incident chain. Incident chain expresses the causal relationships between incidents and it depends on three elements, i.e. incidents, disaster

* Corresponding author.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 171–180, 2013.

receptor and emergency management and is subject to exchanging process of substance, energy and information among the three elements [1]. In existing researches, the research on the incident chains caused by earthquake, typhoon etc. natural disasters have drawn much interest [2-5]. Some typical examples include the disaster analysis that an earthquake causes hazardous chemical leakage and fire, and the analysis that an explosion breaks a container and then causes hazardous chemical leakage. [6-9]. There is also some disaster network related research work. Literature [10] defines the conception of disaster network, semi-quantitatively analyzes the occurrence probability of the network nodes, and gets response tactics in view of the whole minimum loss of the disaster network. Based on the vulnerability of disaster-causing factor, disaster receptor itself and the space-time environmental characteristics around the disaster receptor, exploring the triggering and transforming law of incidents is the basis and core of incident chain analysis [11]. In order to get insight into the law of incident chain, it is necessary to build up decision rule database of interchanging process between incidents through coupling the mechanism research and plenty of incident case data and construct the responding reasoning mechanism. With existing science knowledge and experience, preliminary incident model can be made, which we call prior information while abundant incident cases can be looked on as fact information or experiment data. Both can be integrated and augmented into new recognition of incident chains. Bayesian network is a kind of causal reasoning technique that utilizes the classical Bayesian theorem. It is widely applied to natural language understanding, pattern recognition, artificial intelligence etc. fields. It is a kind of DAG (Directed Acyclic Graph) where the events are nodes and the causal relationship between events are directional edge [12]. It is imaginable that the incident chain can be formed when defining the incident as node and the link strength between edges as triggering probability of an incident (starting) to another incident (end). Thus, we can use Bayesian network analysis to carry out analysis of incident chain.

The paper firstly proposes an approach to incident chain modeling and then, a kind of incident chain analysis technique based on Bayesian network. With simulated example, the calculation process is explained. The work can be consulted in case of prevention and control of complicated disaster circumstances.

2 Incident Chain Model

2.1 Overview of Incident Chain Technology

The occurrence of incidents is not isolated. They are related to each another and have some causal relationship between them. When a certain type of incident occurs, it is likely to induce another kind of incident due to the appearance of disaster-causing factors produced by the release and transfer of energy, substance and information etc. Unexceptionally the induced incident may further cause another incident, and further ⋯. Investigation on plenty of cases of incidents has shown that there exist underlying laws that govern the occurrence and evolvement of these incidents. There are complicated connections between different types of incidents. Incident chain is a kind of innovative research approach of complicated disaster-causing phenomena where a disaster will lead to secondary ones iteratively. It presents a chain-like graph which describes potential correlation between the resultant disasters and their causal disasters [13-14].

In order to explain the incident chain technology, we first put forward two basic concepts: incident chain class and incident chain instance. The incident chain class is the abstract description of the relationship among potential disasters while the incident chain instance depicts the concrete presentation of a complicated disaster scene. The incident chain instance is a specific instantiation of the incident chain class. The incident chain class expresses the underlying principle of complicated disaster phenomena from the cognitive aspect. It is the concentrated reflection of commonsense, expertise, physical law etc. the incident chain instance depicts the real scene of a past occurrence or a predicted one of a probable occurrence.

Figure 1 illustrates an incident chain class of typhoon. As seen, typhoon incident may induce rainstorm and gale. Rainstorm will further cause torrent and inundation. Gale will cause storm surge and billow. Torrent will cause mud-rock slide and landslide. Incident chain of a certain type of incident is not invariant. We can look on incident chain as an express of a priori knowledge which reflects the logic of relationship between different types of incidents. This kind of knowledge is not same for different degree of cognition and different space-time environments. For example, in incident chain of typhoon in coastal area, billow, storm surge are noticed more, while gale and rainstorm are paid more attention in land area. Even, for the same area, the same type of incident has different incident chains in different years, different seasons of each year. In details, the structure of chain-like graphs and the probability of transition between two incidents in the graph are different. The probability of transfer between two incidents can be gained from plenty of historical cases through data mining, case reasoning etc. techniques and can also be studied based on physical laws (e.g. logical causal relationship). Furthermore, rule, model based reasoning techniques are also supposed to enhance the accuracy of the probability. It has been shown that considering space-time factors are quite necessary and essential in the research on pattern and application of incident chain.

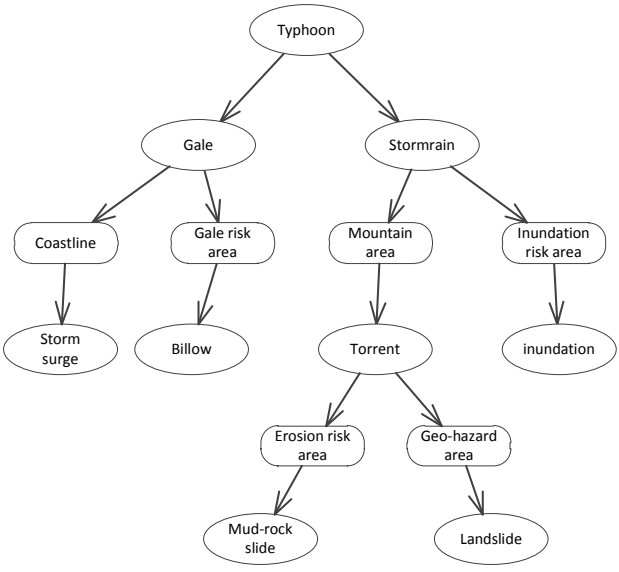


Fig. 1. An example of incident chain class of typhoon

2.2 Formation of Cascading Effect of Incident Chain

According to some academicians' research, an original incident triggers secondary incident through three types of factors inclusively: material, energy and information. More specifically, these three types of factors can be divided into some general impact forms which are called meta-force. Pressure, impulsive force, erosion etc. are some examples of meta-force.

There are relatively intrinsic mapping between meta-force and disaster receptor (including human, thing and social system). Thus, if we can find out two classification system and their mapping, an incident chain expression model can be constructed. Figure 2 illustrates the forming process of incident chain due to cascading effect from one incident to another incident. Obviously, the construction of meta-force set and disaster receptor set, further their mapping relationship is a substantial work. The mapping between incident and meta-force, and meta-force and disaster receptor are many-to-many.

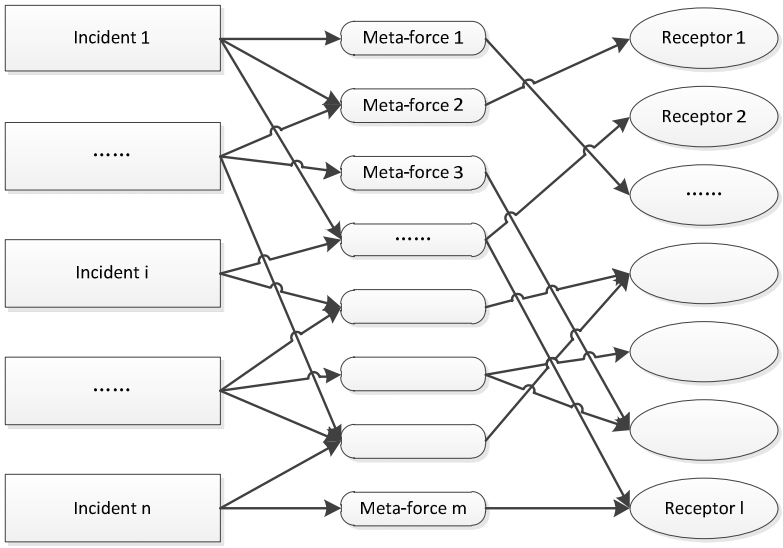


Fig. 2. Illustration of incident, meta-force and disaster receptor

From Figure 2, we can easily understand the possibility of transfer between incidents is dependent on certain exterior conditions. The relevant targets (disaster receptor) are one of the most conditions. For instance, when typhoon occurs, the landslide and debris flow may be induced. However, how much the possibility is depends on whether some relevant risk targets exist within the affected area by the original incident, such as landslide risk regions, flood risk regions etc. If only some incident is configured in the incident chain and the relevant targets exist within the affected area, the relevant secondary and derivative incidents should be concerned. Incident chain can be stored and expressed with the data structure of adjacent table in computer [11].

3 The Analysis of Incident Chain with Bayesian Network

3.1 Introduction to Bayesian Network

Conditional probability is an important concept. While the sum and product rules of probability theory can cope with this factor of conditionality, in many cases such calculations are NP-hard. For example, the prospect of managing a scenario with less than 10 discrete random variables might be tractable. However, an emergency decision support system for inferring likely consequences caused by an incident with 32 random variables resulting in a joint distribution of over 2^{32} parameters would get intractable.

Bayesian network is a simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions [15]. It consists of a set of nodes, one per variable. The links between nodes denotes the causal relation of variable pairs, i.e. if one variable is influenced directly by another variable, there is a link between two variables. The influence variable is called parent variable, and the influenced variable is called child variable. All nodes and links between them constitute a directed, acyclic graph.

Consider a domain U of n variables, x_1, x_2, …, x_n, Each variable may be discrete having a finite or countable number of states, or continuous. The joint probability of all these n variables can be calculated with the following formula according to chain rules of probability theory,

$$p(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} p(x_i \mid x_1, x_2, \cdots, x_{i-1}) \tag{1}$$

For every variable x_i, if there is some subset $\prod_i \subseteq \{x_1, x_2, \cdots, x_{i-1}\}$, which makes x_i be independent of $\{x_1, x_2, \cdots, x_{i-1}\} \setminus \prod_i$, then

$$p(x_i \mid x_1, x_2, \cdots, x_{i-1}) = p(x_i \mid \prod_i) \tag{2}$$

Put x_1, x_2, …, x_i in a certain order, determine a set of variables meeting eq. (2). Then take $\prod_i$ as parent nodes of x_i. and add a direct link between every variable in $\prod_i$ and x_i. Thus, a directed, acyclic graph representing causal relations among variables is generated. Namely, the structure of Bayesian network is determined. Following the construction of Bayesian network, a network learning algorithm will be implemented according to sample data. The goal of network learning is to quantify the probability distribution of the network according to observations and update the prior distribution.

Assume the joint distribution of variable set x={x_1, x_2, …, x_n} can be encoded into the network structure [16]:

$$p(x \mid \theta_s, s^h) = \prod_{i=1}^{n} p(x_i \mid \Pi_i, \theta_i, s^h) \tag{3}$$

Where θ_i is the vector of parameters of the distribution $p(x_i|\Pi_i, \theta_i, s^h)$, θ_s is the vector of parameter group $(\theta_1, \theta_2, \ldots, \theta_n)$, and s^h denotes the assumption that physical joint distribution can be decomposed according to S.

Following network learning from sample data, a verification step is required to take to test if the resultant network is significant and applicable.

3.2 Incident Chain Diagnosis and Prediction with Bayesian Network

In the simplest case, conditional distribution is represented as a conditional probability table (CPT) giving the distribution over x_i for each combination of parent values.

Let incidents be the nodes of Bayesian network and variable set $\{x_1, x_2, \ldots, x_n\}$ are used to represent the random binary variables and 0 is nonoccurrence and 1 is occurrence. Construct the Bayesian network following the steps below.

```
(1) Choose an ordering of variables x1, … , xn
(2) For i = 1 to n
    add xi to the network
    select parents from x1, … ,xi-1 such that
        P(xi | Parents(xi)) = P(xi | x1, ..., xi-1)
```

With Bayesian network, we can do two types of things. One is diagnosis, which is a bottom-up process. This term is from disease treatment where a doctor will look for causes according to the patients' symptom. The other is prediction which is a top-down process. Prediction will look for the possibility of potential effects. For incident chain analysis, the former will be able to find out the most possible factors for current consequence or infer the key nodes for existing incident risk, and the latter will help us predict the possible consequence given a specified incident scenario.

4 The Application of Incident Chain in Emergency Prevention and Response

4.1 Construction of Bayesian Network of Incident Chain

Suppose an incident chain class in figure 1. In order to implement area specific analysis, one need to instantiate the incident chain class in terms of space-time information such as spatial objects and seasonal factors. This process can be done under the support of GIS [11]. First of all, according to the condition factors configured in the data structure of incident chain, GIS can filter out potential (worth concerning) secondary incidents by use of spatial query functions. The analysis steps are listed as follows:

Step 1: The path and range of the presumptive original incident (typhoon) is identified. This result may come from professional prediction of meteorological bureau.

Step 2: potential secondary incidents are filtered out. For every incident in the configuration of secondary incidents in the incident chain class, GIS searches for associated targets (condition factors). If the associated condition factors are present, then a node of the secondary incident is added into the corresponding incident chain instance.
Step 3: The influence areas for every identified secondary incident is calculated according to proper influence analysis models. These kinds of models are generally from some professional departments such as environment, hydrology and geology and so on.
Step 4: Repeat Step 2 – Step 3, until the whole study area is covered or the expected level of secondary incidents.
According the above analysis, the incident chain structure (or the structure of Bayesian network) is determined. Then the probability distribution (In particular, the inducing probability of incident pairs in the chain graph) of the network needs to be calculated. In literature [17], a calculation framework is proposed through which inducing probabilities for incident pairs in the chain graph are able to be calculated.

Figure 3 is an area specific incident chain for some area located within a coastal city of China, where typhoon strikes it in routine every year. After going through some calculations mentioned above, the following incident chain instance is identified.

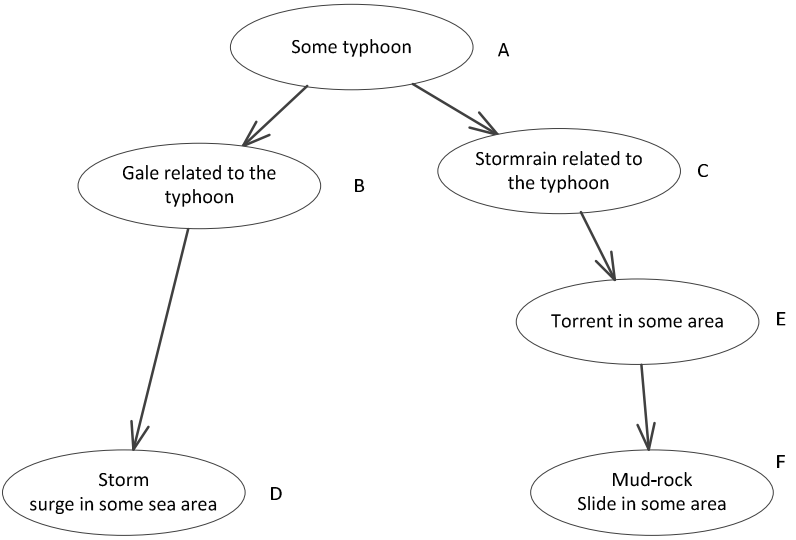


Fig. 3. Incident chain instance derived from some incident chain class

4.2 Risk Analysis

In essence, Risk analysis for incident chain is an inference procedure based on Bayesian network. From the incident chain instance of figure 2, According to probability chain rule and conditional independence, the joint probability of variables of incident chain can be written as follows:

$$P(\mathrm{A,B,C,D,E,F})=P(\mathrm{A})P(\mathrm{B|A})P(\mathrm{C|A})\times P(\mathrm{D|B})P(\mathrm{E|C})P(\mathrm{F|E}) \tag{4}$$

Thus, one can calculate the occurrence probability of storm surge triggered by the original typhoon incident:

$$P(\mathrm{D=1\,|\,A=1}) = \frac{P(\mathrm{D=1,A=1})}{P(\mathrm{A=1})} \tag{5}$$

Where,

$$P(\mathrm{D=1,A=1}) = \sum_{\mathrm{B,C,E,F}} P(\mathrm{A=1,B=b,C=c,D=1,E=e,F=f}) \tag{6}$$

$$P(\mathrm{A=1}) = \sum_{\mathrm{D}} P(\mathrm{A=1,D=d}) \tag{7}$$

In this means, the occurrence probabilities for all potential secondary incidents can be calculated according to the joint probability of all incidents.

Then, given the consequence or loss of every potential incident, the risk can be calculated. For example, if the loss for each node is:

$$\mathrm{C}=\{\mathrm{c_B},\mathrm{c_C},\mathrm{c_D},\mathrm{c_E},\mathrm{c_F}\} \tag{8}$$

Then the expected risk caused by the original incident is:

$$\mathrm{c_{total}}=P(\mathrm{B|T})\mathrm{c_B}+P(\mathrm{C|T})\mathrm{c_C}+P(\mathrm{D|T})\mathrm{c_D}+P(\mathrm{E|T})\mathrm{c_E}+P(\mathrm{F|T})\mathrm{c_F} \tag{9}$$

4.3 Decision Optimization

Based on Bayesian network of the incident chain, several approaches to decision optimization can be made. Firstly, suppose some chain(s) are broken in an incident chain, the resulting average risk can be recalculated according to Equation (9). Thus, through comparing the resulting risk values, one can easily determine which tactics is the best one and then should be employed to implement in emergency response. Secondly, on condition that there are limited resources which can be used in emergency response activity, how to allocate these resources optimally is a key issue to promote efficiency and effectiveness of emergency response.

Suppose the expected risk of an incident chain is R_0. Now there are n tactics t_1, t_2, …, t_n which can be employed to break chain. The cost for the i^{th} tactics is q_i and the resulting average risk is R_i then the unit cost can be calculated according to the following equation:

$$q_i^0 = \frac{q_i}{R_0 - R_i} \tag{10}$$

After simple ordering of the resultant unit cost, the best tactics with minimum unit cost can be chosen as the final solution to emergency response. With limited emergency resources, more allocation will be made for the prevention and control of the crucial node.

5 Conclusions and Prospects

Complicated prediction is very important to achieve pre-warning and timely treatment of incidents especially in situation where an original incident will trigger secondary or derivative incidents. The principle of incident chain is one of the most fruits in public safety science. It originates from disaster chain. The study of the underlying mechanism and evolvement rule is extremely helpful for prediction and control of incidents.

This paper preliminarily explores incident chain model including mechanism of incident chain and its construction and presentation. In mechanism, we illustrate the Formation of cascading effect of incident chain. We think both of meta-force and receptor attribute are generalized factors in generation of incident chain. This provides a general theoretical framework for incident chain research. Following the research on construction and presentation of incident chain modeling, the paper further investigates analysis approach to incident chain based on Bayesian network. As a kind of natural presentation of incident chain, Bayesian network techniques have inherent advantages when used in inference and prediction of complicated disaster scenario. Concrete applications include risk analysis and decision optimization.

Currently, the application of incident chain in public safety is at its initiative phase. In order to meet the construction of emergency platform, it is urgent that some relevant theories and techniques need to be studied more deeply. The breakthrough and solution of these problems such as risk analysis incorporating incident chain, the application of advanced spatial analysis of GIS, the study on patterns of chain-break and disaster reduction, incident chain based intelligent decision etc. are quite meaningful for prevention and control of incidents.

Acknowledgement. The authors would like to thank the support of the National Natural Science Foundation of China (Study on Pre-qualification Theory and Method for Influences of Disastrous Meteorological Events, Grant No. 91224004), the National Natural Science Foundation of China (Key Scientific Problems and Integrated Research Platform for Scenario Response Based National Emergency Platform System, Grant No. 91024032), the National Natural Science Foundation of China (Study on Systematical Integration Theory and Method of Emergency Technologies for Unconventional Incidents, Grant No. 91024024) and the National Natural Science Foundation of China (Study on Emergency Decision-making Based on Generalized Knowledge Coupling, Grant No. 70901047).

References

1. Fan, W.C., Liu, Y., Weng, W.G.: Triangular framework and "4+1" methodology for public security science and technology. Science & Technology Review (2009)
2. Wang, H.B., Sassa, K., Xu, W.Y.: Analysis of a spatial distribution of landslides triggered by the 2004 Chuetsu earthquakes of Niigata Prefecture, Japan. Natural Hazards 41, 43–60 (2007)
3. Chenga, J.D., Huang, Y.C., Wu, H.L., et al.: Hydrometeorological and landuse attributes of debris flows and debris floods during typhoon Toraji, July 29-30, 2001 in central Taiwan. Journal of Hydrology 306, 161–173 (2005)
4. Dai, F.C., Deng, J.H., Tham, L.G., et al.: A large landslide in Zigui County, Three Gorges area. Canadian Geotechnical Journal 41(6), 1233–1240 (2004)
5. Iverson, R.M.: Landslide triggering by rain infiltration. Water Resources Research 36, 1897–1910 (2000)
6. Cozzani, V., Salzano, E.: The quantitative assessment of domino effects caused by overpressure, Part I. Probit models. Journal of Hazardous Materials A107, 67–80 (2004)
7. Cozzani, V., Salzano, E.: The quantitative assessment of domino effects caused by overpressure, Part II. Case studies. Journal of Hazardous Materials A107, 81–94 (2004)
8. Fabbrocino, G., Iervolino, I., et al.: Quantitative risk analysis of oil storage facilities in seismic areas. Journal of Hazardous Materials A123, 61–69 (2005)
9. Antonioni, G., Spadoni, G., Cozzani, V.: A methodology for the quantitative risk assessment of major accidents triggered by seismic events. Journal of Hazardous Materials 147, 48–59 (2007)
10. Glickman, T.S., Khamooshi, H.: Using hazard networks to determine risk reduction strategies. Journal of the Operational Research Society 56, 1265–1272 (2005)
11. Zhong, S.B., Shu, X.M., Yuan, H.Y., et al.: Preliminary Study on Synthetical Forecast Based on Incident Chain in Emergency Platform of Public Safety. In: Proceedings of EMS 2008 (2008)
12. Zhang, N.L., Poole, D.: A simple approach to Bayesian network computations. In: Proceedings of the10th Canadian Conference on Artificial Intelligence, pp. 171–178. Morgan Kaufmann Publishers, Waltham (1994)
13. Cozzani, V., Salzano, E.: The quantitative assessment of domino effects caused by overpressure Part I: Probit models. Journal of Hazardous Materials (A107), 67–80 (2004)
14. Cozzani, V., Salzano, E.: The quantitative assessment of domino effects caused by overpressure Part II: Case studies. Journal of Hazardous Materials (A107), 81–94 (2004)
15. Baclawski, K.: Bayesian network development. In: Fujita, H., Gruhn (eds.) New Trends in Software Methodologies, Tools and Techniques. IOS Press (2004)
16. Dai, Q., Ma, J.W., Li, Q.Q., et al.: The Study on Remote Sensing Data Classification Using Bayesian Network. Journal of Electronics & Information Technology 27(11), 1782–1785 (2005)
17. Li, M., Chen, J.G., Chen, T., et al.: Probability for disaster chain in emergencies. Journal of Tsinghua University (Science & Technology) 50(8), 1173–1177 (2010)

Forest Fire Risk Mapping Based on Spatial Logistic Model of Northeastern China Forest Zone

Ou Deng[1,2], GuoFeng Su[1,*], QuanYi Huang[1], and YiQiu Li [2,3]

[1] Institute of Public Safety Research, Department of Engineering Physics, Tsinghua University, 100084 Beijing, China
[2] Ecological Security and Protection Key Laboratory of Sichuan Province Mianyang Normal University, 621006 Mianyang, China
[3] Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, 100101 Beijing, China
cnfly@foxmail.com

Abstract. Forest fire risk and danger mapping is about forest fire potential estimating. Spatial differentiation of forest fire risk and danger is obvious in the Northeastern China Forest Zone, and it is of crucial importance in setting up management objectives in this region. Based on RS and GIS technology and spatial logistic regression analysis, the relations between the burn scar and its influencing factors were evaluated using Wald step forward logistic regression. Using logistic regression model, forest fire risk probability of each pixel was calculated and five-class forest fire risk and danger zones were identified. The result shows that spatial differentiation of forest fire risk and danger is obvious in the region. Different forest fire management can be applied in different fire risk zone. To eradicate forest fire from the beginning, budgets, personnel, technology, equipment, early warning and monitoring systems need to be adjusted in fire management systems.

Keywords: forest fire, northeastern China, spatial logistic regression model, MODIS.

1 Introduction

Forest fires are a recurrent environmental and economic emergency worldwide. Since 1980s, forest fires caused by global warming appeared to be on the rise. The protection of life property and forest resources requires increasingly more effective forest fire management, and it is possible to map forest fire risk and danger zones and thereby minimize the frequency of fire, avert damage, etc.

Fire risk and danger refers to the probability of ignition, as determined by the presence and activity of causative agents (i.e., man, lightning, etc.) [1]. The term "danger" refers to sum of constant and variable factors affecting the ignition, spread, and resistance to control, and subsequent fire damage. Forest fire risk and danger

* Corresponding author.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 181–192, 2013.

zones are locations where a fire is likely to start, and from where it can easily spread to other areas.

A precise evaluation of forest fire problems and decisions on solution methods can only be satisfactorily made when a fire risk zone map is available. Statistical methods such as linear regression, classification regression trees, logistic regression analysis, neural networks or Bayesian probability have been used in fire risk mapping to generate local risk models[2], [3], 4], [5], [6], [7], [8], [9]. A geographic information system (GIS) can be used effectively to combine different forest-fire-causing factors for demarcating the forest fire risk zone map. Remote Sensing (RS) data plays a vital role in identifying and mapping forest fires and in recording the frequency at which different zones are affected. Based on GIS and RS, a spatial logistic regression model is developed in this study to predict the forest fire risk and danger probability. Northeast China Forest Zone is selected for this study, because it continually faces a forest fire problem, and the large range, low population density, limited traffic infrastructures and weak fire suppression capabilities, combined with the continental, dry character of the climate, created the conditions for large conflagrations.

2 Methods

2.1 Study Area

Northeast China includes Heilongjiang Province, Jilin Province, Liaoning Province and Inner Mongolia autonomous region. The Northeastern China Forest Zone is stretching over ten degrees latitude from South to North (41°30′-53°30′ N) and nearly five degrees longitude from East to West (115°30′-135°20′ E), including territories of Liaoning, Jilin, Heilongjiang and the eastern part of the Inner Mongolia Autonomous Region. The Northeastern China Forest Zone locates in the Greater and Lesser Hinggan Mountains and the Changbai Mountains, and it is the largest natural forest area in China. The main forests in the region are cold-temperate conifer mixed forests, temperate conifer and broadleaf mixed forests, warm-temperate deciduous broadleaf mixed forests, and the chief tree species include larch and Korean pine. The region is an important forestry of China. In the region Greater Hinggan Mountains forest district, Lesser Hinggan Mountains forest district, Changbai Mountains forest district, the main timber bases are located, with its forest coverage and timber reserves accounting for over one third of the national totals, and the area turns out half of the national total timber output.

The climate in the extreme northeastern region of China is continental cold-temperate, with long, cold, dry winters and short, warm, humid summers. The infrequency of winter storm passages produces relatively little precipitation. As spring approaches the storm frequency increases, temperature rapidly increases and humidity remains low, the combination of conditions leads to the drying of forest fuels, creating a prime situation for fire development. In the summer, the frequency of storm passages decreases, but the storm intensity increases, so the greater amount of precipitation associated with the stronger storms reduces the fire hazard. With the cooling of autumn the behavior of the spring storms is repeated, but the storms lack the intensity and strong winds common in spring, and smaller fires occasionally

occur. By November the dry period begins again. During this annual climatic cycle, forest fires are most frequent and severe in the spring. There are few fires during the summer, and small fire development is common during the autumn.

Due to the high fire hazard and risk in the grass and forest cover, these districts were also the first forest districts, where forest fire prevention was launched. In the past, the large range, low population density, limited traffic infrastructures and weak fire suppression capabilities always lead to large forest fires in the region. Since 1987, the prevention, fighting, and management of forest fires have been strongly enhancing, forest fires still occurred frequently.

2.2 Materials

The main characteristic of the forest fires emergency is its extreme dimension in terms of affected area and number of events. The only technology capable of responding to the requirements of cost effective monitoring and impact estimation is remote sensing. Many different satellite sensors, with complementary capabilities, have been employed for forest fire studies. The Moderate Resolution Imaging Spectroradiometer (MODIS) possesses the most useful channels for forest fire monitoring and mapping [10], [11],[12]. Nowadays, MODIS has more than 10 year's history and has sufficient temporal sampling, which is enough to exploit the long historical archive at a spatial resolution suitable for regional-scale studies. We process 2000-2010 MCD45A1, MOD/MYD14A1, MOD13Q1 and the ground-based fire reports to extract and validate burned area, and mapping the history forest fire distribution map in the Northeastern China Forest Zone. Average temperature distribution map, average precipitation distribution map, vegetation regions, landforms data, topographic data, settlement and road distribution map are also employed for the study to develop the fire risk and danger driving factor maps. All the fire risk and danger driving variables (maps) and the history forest fire distribution map are then integrated through spatial logistic analysis using a sample size of 34233 pixels. The forest fire risk and danger probability density distribution map is generated using the logistic regress model. By rating the probability density map, the forest fire risk and danger rating map of Northeastern China Forest Zone is integrated.

2.3 Dependent Variable: Burn Scar Distribution Map

The dependent variable of the fire risk and danger logistic regress model is generated using MCD45A1 (a monthly Level-3 gridded 500-meter fire burned area product). Nowadays, the fire potential burned pixels apply daily imagery acquired by MODIS sensors, on board of the TERRA and AQUA satellites from NASA (National Aeronautics and Space Administration). MCD45A1 was processed to develop the potential burned area distribution map. The data contains approximate Julian day of burning from eight days before the beginning of the month to eight days after the end of the month, or a code indicating unburned areas (0), approximate Julian day of burning (1-366), snow or high aerosol (900), internal water bodies (9998), seas and oceans water bodies (9999), or lack of data (10000).

MOD/MYD14A1, MOD13Q1 and the ground-based fire reports were used as validation data of burned area in the study. The MOD14 thermal anomaly algorithm uses a series of thresholds and contextual tests to assess whether a pixel contains a fire [13], [14]. The MOD/MYD14A1 daily Level 3 fire product is tile based; it contains eight successive daily fire masks for a specific MODIS tile. Each of these daily masks is a maximum value composite of ten classes for those swaths, with a value of 7 for a low confidence detection, a middle value of 8, and a value of 9 for a high confidence detection. MOD13Q1 is a 16-day Level 3 250-meter vegetation indices product, including a Normalized Difference Vegetation Index (NDVI) and an Enhanced Vegetation Index (EVI).

The validation algorithm takes advantage of the MOD/MYD14A1 hotspot detection, the changes in MOD13Q1 NDVI before and after burning, and the history in-situ fire records. First, the detected MOD/MYD14A1 hotspots are confirmed as real fires if there are significant decreases in MOD13Q1NDVI in the next month. Second, the confirmed hotspots are connected to MCD45A1 burned scar patches. Any burned patches that contain less than 5% of confirmed hotspot pixels are removed. Finally, the modified burned patches are connected to the history in-situ fire spot records, and the detection results agree well with the ground-based fire reports. There is a reasonable agreement with the correlation coefficient equal to 0.85. In this way, 11 years burn scar distribution map from 2000 to 2010 is created, shown as Fig.1.

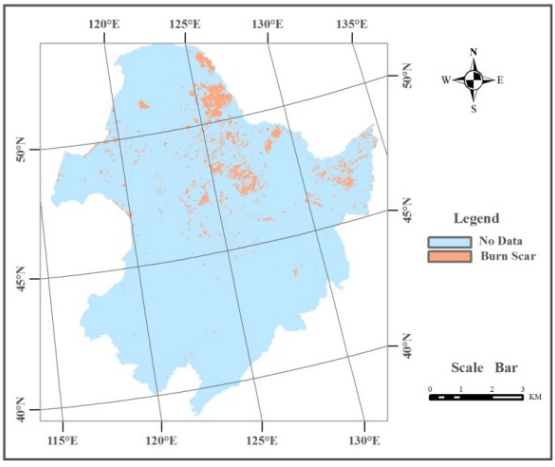


Fig. 1. Burn scar distribution map 2000-2010

2.4 Explanatory Variables: Forest Fire Risk and Danger Influencing Factors

Anticipation of factors such as climate, vegetation condition, topography and human activities are influencing the occurrence of a forest fire and understanding the dynamic behavior of a forest fire. These factors and forest fires are interrelated so that a change in any of the variables can affect the others.

Climate change directly and indirectly influences the wildfire regime through various mechanisms. Climate changemay alter the structure and composition of plant communities and thus affect the physical and chemical properties of fuels[15].

Vegetation type and condition is the most important elements in the evaluation of the forest fire risk and danger, since it characterizes the forest fuel[16]. Relative species composition is an indicator of site conditions and directly affects flammability of the fuel complexes[17]. Deciduous forest stands are considered to represent low fire danger areas, whereas coniferous stands are generally associated with high fire danger. Shrubs and grass are also known to have high fire danger potential due to the high flammability of the fuels and fast spread of fires in these vegetation types. Vegetation coverage is the percentage that a horizontal projective area of vegetation, including leaves, stems and branches, accounting for the total area of a study area. Vegetation coverage provides an indicator of the ease with which fire can spread. The higher the vegetation coverage is, the more intense the fires burns are.

Topography may influence fire occurrence and spread on landscape to regional scales. Particularly, elevation affects the temperature and water availability. In mountainous areas, elevation also influences the potential for ignition of fires by lightning. Terrain aspect directly affects fuel moisture status by altering the solar insolation, while the slope angle affects fire spread by increasing the efficiency of radiant energy transfer from flaming fronts to upslope fuels. A fire spreading up a steep slope resembles a fire spreading under the influence of a strong wind.

Fire risk was associated with ignition, and ignition in the Northeastern China Forest Zone was mostly triggered by human activities. Human-caused fires dominated there, and human fire causes were summed up as forestry operations, agricultural fires, smoking, and the use of fire in field etc.

2.5 Data Processing and Formats

The data of the forest fire risk and danger factors originate fromremote sensing (e.g., vegetation coverage), maps (e.g.,temperature, precipitation, vegetation, soil, landform), derived data (e.g.,slope and aspectderive from the digital elevation model (DEM), distance to river, settlements, and roadsderive from their distribution map), andother sources. To allow a straightforward analysis, Geographical Information Systems (GIS) are used to process all the spatial data. The data are converted into a grid based system with aresolution of 572×572m, same as MCD45A1 burned area data, and the data are projected into the same geo-referenceAlbers Conical Equal Area/Krasovsky.

Climate data is converted from vector to raster. The average temperature distribution map is divided into six classesfor forest fire risk and danger analysis: -6- -4℃, -4- -2℃, -2- 0℃,0-2℃,2-4℃,4-6℃, and the average precipitation map is divided intofour classes for fire risk and danger analysis: 300-400mm,400-500mm, 500-600mm,600-700mm.

Institute of Geography, Chinese Academy of Science produced a soil map of the People's Republic of China.The study area soil mapcontains ten soil types: Dark-brown earths, Dark meadow soils, Chernozems, Brown earths, Castanozems, Aeolian soils, Black soils, Brown earths, Bog soils, and others.

Institute of Botany, Chinese Academy of Science produced a vegetation map of the People's Republic of China; the study area vegetation map contains ten vegetation regions: coniferous forest, coniferous and broad-leaved mixed forest, broad-leaved forest, shrub, grassland, grass meadow, marsh vegetation, and alpine vegetation, without vegetation. Vegetation coverage is calculated according to the followingformula:

$$f_V = (NDVI - NDVI_{min}) / (NDVI_{max} - NDVI_{min}) \quad (1)$$

The Normalized Difference Vegetation Index (NDVI) was got from MOD13A3 images. Vegetation coverage wasdivided into five levels for forest fire risk and danger analysis:<0.2, 0.2-0.4, 0.4-0.6, 0.6-0.8, >0.8.

Topographic data was extracted from the Digital Elevation Model (90mDEM), which provided three different base maps: slope, elevation and aspect.Elevation was divided into five zones for forest fire risk and danger analysis: elevation ≤200 m, 201–500 m, 501–800 m, 801–1200mand >1200 m.Slope was classified into five classes for a similar gross pattern analysis: ≤5°, 5°-10°, 10°-15°, 15°-25°, and >25°.The terrain aspect was classified into nine classes: N (north), NE (northeast), E (east), SE (southeast), S(south), SW (southwest), W (west), NW (northwest) and flat, based on a general aspect classification method. Distance to rivers was divided into seven classesfor forest fire risk and danger analysis:<0.5km, 0.5-1km,1- 2km, 2-3 km, 3-4km, 4-5km, >5km.

Distance to settlements and distance to roads were considered to be theforest fire risk and danger factors to represent the influence of human activities. The distance was also divided into seven classesfor forest fire risk and danger analysis:<0.5km, 0.5-1km,1- 2km, 2-3 km, 3-4km, 4-5km, >5km.

2.6 Statistical Model

Logistic regression analysis is one of the most popular mathematical modeling approaches that can be used to describe the relationship of several variables to a binary dependent variable[18]. In logisticmodel, the dependent variable is a dichotomous dummy variable - the variable can take only two values, 1 and 0.These models are typically used to predict whether or not some event will occur. In this study, the forest fire occurrence or no occurrenceis evaluated using Wald step forward methodstepwise logistic regression model following:

$$\ln\left\{\frac{P}{1-P}\right\} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \quad (2)$$

Where P is the probability of a grid cell for the occurrence of a forest fire, x_1 , x_2 ,$\cdots$, x_n are the influence factors, β_0 , β_1 ,$\cdots$, β_n are the coefficient of explanatory variables of logistic equation. The stepwise procedure is used to help us select the relevant driving factors from a set of factors that is assumed to influence the forest fire risk and danger. Variables that have no significant contribution to the explanation of the forest fire risk and danger are excluded from the final regression equation. Forward entry adds factors to the model until no omitted variable would contribute significantlyto the model. This is the usual option, starting withthe constant-only model and adding variables one at a timein the order they are best by

some criterion until some cutofflevel is reached (i.e. until the step at which all variablesnot in the model have a significance higher than 0.05). The influence of the explanatory variables on the dependent variable can be measured by Exp(β). Exp(β) indicates the frequency ratio of the event occurrence or no occurrence, and indicates the dependent variable change while the explanatory variable adds each unit.

2.7 Validation

The goodness of the logistic regressionmodelfit can be evaluated with the receiver-operating characteristic (ROC) method[19]. ROC analysis is an ideal technique to quantify the tradeoffs of test sensitivity and specificity. In ROC analysis, the area under the ROC curve (AUC) is a common metric that can be used to compare different tests (indicator variables). The AUC is a measure of test accuracy. AUCis the portion of the area of the unit square, andits value will always be between 0 and 1.0. However,because random guessing produces a diagonal line between(0.0) and (1.0), which has an area of 0.5, no realisticclassifier should have an AUC less than 0.5.The larger the AUC is, the more accurate the test is. An AUC 0.5-0.69 corresponds to a poor diagnostic test, 0.7-0.79 corresponds to a reasonable diagnostic test, and 0.8-0.89 corresponds to an excellent diagnostic test, 0.9 or higher corresponds to an exceptional diagnostic test.

3 Results

The influence of spatial autocorrelation on the regression results can be minimized by only performing theregression on a sample of pixels that are selected at more or less equidistant from one another. Such aselection method is adopted in order to maximize the distance between the selected pixels to attenuate theproblem associated with spatial autocorrelation. By using the selection method, a sample size of 34233 pixels was extracted to develop a logistic regression model. 70%of the selected pixelswere used to generate themodel and the remaining 30% of the cells are used forvalidation. Predictions are finally obtained for 100% ofthe data, then the probabilities of fire occurrencefor each pixel are obtained and five-class forest fire risk and danger zones are identified.

3.1 Model Development

The statistical package SPSS v.18 is used to reveal and quantify the relations between the location of forest fire risk and the set of explanatory factors. The Wald step forward technique is applied, and ten candidate models are obtained. The following table1 shows the logistic regression results for the model built with the calibration sample. The vegetation coverage have an frequency ratio larger than 1, meaning that there was an increase in the frequency, i.e. when vegetation coverage increases a unit keeping the probability of fire increases by 14.4%. The “Sig.” values in table1 are all less than 0.05, then the variables are significant >95% level, and parameter estimation of the model successfully passed the Wald inspection.

Table 1. Logistic regression resultsfor the forest fire risk and danger model

Variables	B	S.E	Sig.
Average temperature X_1	-0.387	0.029	.000
Average precipitation X_2	-0.116	0.027	.000
Soil type X_3	-0.040	0.008	.000
Vegetation type X_4	-0.025	0.012	.044
Vegetation coverage X_5	0.135	0.018	.000
Landform X_6	-0.021	0.009	.026
Elevation X_7	-0.282	0.026	.000
Slope X_8	-0.135	0.053	.011
Distance to rivers X_9	-0.065	0.016	.000
Distance to settlements X_{10}	0.032	0.013	.016
Constant	-0.868	0.245	.000

3.2 Omnibus Tests of Model Coefficients and ROC Analysis

Omnibus tests are a kind of statistical test. They test whether the explained variance in a set of data is significantly greater than the unexplained variance. In the study, at a given significance levelα=0.05, degree of freedomdf=10, Chi-square critical value χ^2=18.307 , the model Chi-square is 884.597 (Table 2), greater than Chi-square critical value, and the "Sig." values are all ≤0.05. The results show that at the given significance level, model coefficients pass the test.

Table 2. Omnibus tests of model coefficients

		Chi-square	df	Sig.
Step10	Step	4.045	1	.044
	Block	884.597	10	.000
	Model	884.597	10	.000

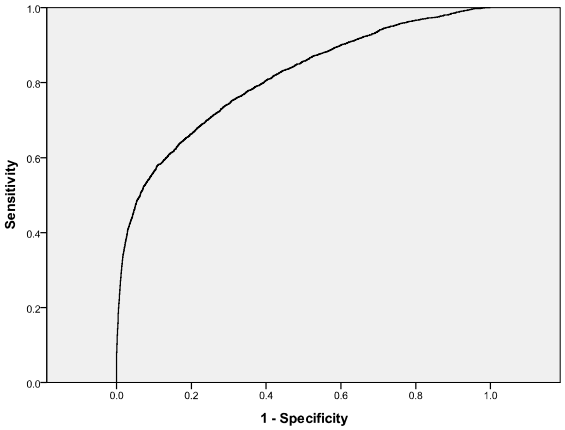


Fig. 2. ROC Curve analysis

According the ROC analysis of the study, the areaunder the curve (AUC) is0.809 (Fig.2). That is to say the regression equation of dependent variable has anexceptionalexplanatory capacity.

3.3 Forest Fire Risk and Danger Mapping

According to table1, the logistic regression model can be expressed as following:

$$Ln\left\{\frac{p}{1-p}\right\}=-0.868-0.387x_1-0.116x_2-0.040x_3-0.025x_4+0.135x_5 \\ -0.021x_6-0.281x_7-0.135x_8-0.065x_9+0.032x_{10} \tag{3}$$

Let

$$Z=-0.868-0.387x_1-0.116x_2-0.040x_3-0.025x_4+0.135x_5 \\ -0.021x_6-0.281x_7-0.135x_8-0.065x_9+0.032x_{10} \tag{4}$$

Then

$$P=\frac{e^Z}{1+e^Z} \tag{5}$$

According to equation (5), the forest fire risk probability of each pixel can be calculated. The probability values and their spatial distribution inthe region is a predictor of forest fire risk and danger. Based on theforest fire risk and danger probability density distribution, five-class (no or very low, low, moderate, high, extreme) forest fire risk and danger zones are identified in the Northeast China Forest Zone(Fig.3). The classes are based on natural groupings inherent in the data. ARCGIS identifies break points by picking the class breaks that best group similar values and maximize the differences between classes. The features are divided into classes whose boundaries are set where there are relatively big jumps in the data values.

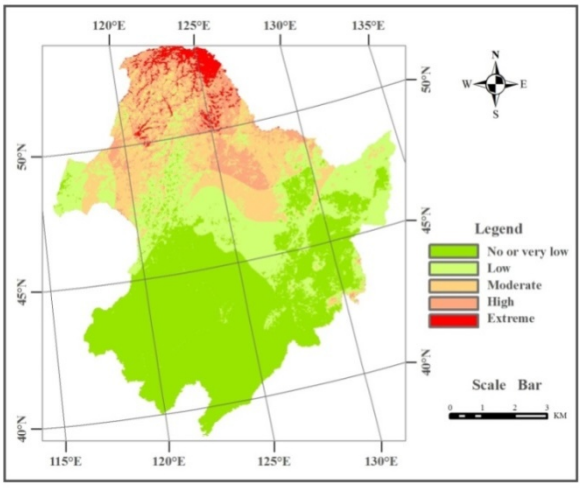


Fig. 3. Forest fire risk and danger zones distribution map

3.4 Spatial Differentiation Analysis

Spatial differentiation of forest fire risk and danger was obvious in the Northeast China Forest Zone: The Greater Hinggan Mountains cold temperate coniferous forest area was centralized high fire risk and danger zone and extreme fire risk and danger zone. The Lesser Hinggan Mountains temperate coniferous and broad-leaved mixed forest area basically belonged to high fire risk and danger zone and moderate fire risk and danger zone. The Changbai Mountains warm-temperate deciduous broadleaf mixed forests small area belonged to moderate fire risk and danger zone, and most areas belonged to low or no fire danger zone.

Burn scar distribution map (Fig.1) and forest fire risk and danger zones distribution map (Fig.3) overlay analysis showed that the burnt ratio in the five-class forest fire risk and danger zones is 0.58%, 5.18%, 7.74%, 12.60%, and 14.21% in turn (Table 3).

Table 3. Burnt ratio ofthe forest fire risk and danger zones

Fire risk and danger zones	Burnt ratio (%)
no or very low	0.58
low	5.18
moderate	7.74
high	12.60
extreme	14.21

4 Conclusion and Discussion

The forest fire risk and danger measure is an important part of forest fires management and a very powerful tool for supporting the prevention and combat activities. RS and GIS technology can be used for forest fire risk analysis and forest fire danger division to provide detailed, fast, high reliability, and direct information. Remote Sensing data plays a vital role in identifying and mapping forest fires and in recording the frequency at which different zones are affected. MODIS active fire products provide a resourceful data to develop the potential burned area distribution map and have sufficient temporal sampling, which are enough to exploit the long historical archive at a spatial resolution suitable for regional-scale studies. Geographic Information System can be effectively to combine different forest-fire-causing factors for demarcating the forest fire risk zone map. Spatial logistic regression analysis is an effective method to calculate the forest fire risk and danger probability, and the computational method gives a relevant support to the decision maker for solving the specific requirements related to forest fires management. In the study, based on RS and GIS technology, a large volume of data of 2000-2010 MCD45A1, MOD/MYD14A1, MOD13Q1and the ground-based fire reports to extract and validate burnt area were processed, and the forest fire influence factors data originate from remote sensing, maps, and other sources were analyzed. The relations between the burn scar and its influencing factors were evaluated using Wald step forward logistic regression.

The result shows that the regression equation of dependent variable has an exceptional explanatory capacity. Based on the logistic regression model, the forest fire risk probability of each pixel was calculated and five-class forest fire risk and danger

zones were identified. The Greater Hinggan Mountains cold temperate coniferous forest area is centralized high and extreme fire risk zone. The Lesser Hinggan Mountains temperate coniferous and broad-leaved mixed forest area basically belongs to moderate fire risk zone. The Changbai Mountains warm-temperate deciduous broadleaf mixed forests most areas belongs to low fire danger belongs to low fire danger and no fire zone. The higher the forest fire risk and danger rating are, the higher the burnt ratio in the study region is.

Forest fire risk and danger mapping is about forest fire potential estimating, but forest fire occurrence and spread has its randomness and uncertainty. Fire prevention should stand to the guideline of “take prevention first and extinguish actively". Spatial differentiation of forest fire risk and danger is obvious in the Northeast China Forest Zone, and it is of crucial importance in setting up management objectives in this region. Different forest fire management should apply in different fire risk zone. Fire management systems need to make adjustments in budgets, personnel, technology, equipment, early warning and monitoring systems. Strengthen publicity and education in order to improve the fire safety awareness, and eradicate forest fire from the beginning.

Acknowledgments. This study is supported by the National Natural Science Foundation of China (Grant No. 91024017), the National Key Technology R&D Program (Grant No. 2011BAK09B01), State Key Program of National Natural Science of China(No. 91224004) and Major projects in Guizhou province(No. [2012]6007). The MODIS data used in this analysis were received from https://wist.echo.nasa.gov. The ground-based fire reports were gotten from Heilongjiang provincial forest agencies.

References

1. Hardy, C.C.: Wild land fire hazard and risk: Problems, definitions, and context. Forest Ecology Management 211(1-2), 73–82 (2005), doi:10.1016/j.foreco.2005.01.029
2. Chao-Chin, L.: A preliminary test of a human caused fire danger prediction model. Taiwan Journal of Forest Science 17(4), 525–529 (2002)
3. Vega García, C., Woodard, P.M., Titus, S.J., Adamowicz, W.L., Lee, B.S.: A logit model for predicting the daily occurrence of human caused forest fires. International Journal of Wildland Fire 5(2), 101–111 (1995)
4. Koutsias, N., Kalabokidis, K.D., et al.: Fire occurrence patterns at landscape level: beyond positional accuracy ofignition points with kernel density estimation methods. Nature Resource Model 17(4), 359–375 (2004)
5. Robin, J.G., Carrega, P., Fox, D.: Modeling fire ignition in the Alpes-Maritimes Department, France. A comparison. In: Proceedings of V International Conference on Forest Fire Research, Figueira da Foz, November 27-30, pp. 27–30 (2006)
6. Amatulli, G., Camia, A.: Exploring the relationships of fire occurrence variables by means of CART and MARS models. In: Proceedings of IV International Wildfire Conference, Seville, May 13-17 (2007)
7. Amatulli, G., Pérez-Cabello, F., de la Riva, J.: Mapping lightning/human-caused wildfires occurrence under ignition point location uncertainty. Ecological Modelling 200(3-4), 321–333 (2007)

8. Yang, J., He, H.S., Shifley, S.R., Gustafson, E.J.: Spatial patterns of modern period human-caused fire occurrence in the Missouri Ozark Highlands. Forest Science 53(1), 1–15 (2007)
9. Syphard, A.D., Radeloff, V.C., Keeley, J.E., Hawbaker, T.J., Clayton, M.K., Stewart, S.I., Hammer, R.B.: Human influence on California Fire Regimes. Ecolology Application 17(5), 1388–1402 (2007)
10. Justice, C.O., Giglio, L., Korontzi, S., Owens, J., Morisette, J.T., Roy, D., et al.: The MODIS fire products. Remote Sensing of Environment 83(1-2), 244–262 (2002)
11. Kaufman, Y.J., Justice, C.O., Flynn, L.P., Kendall, J.D., Prins, E.M., Giglio, L., et al.: Potential global fire monitoring from EOS-MODIS. Journal of Geophysical Research-Atmospheres 103(D24), 32215–32238 (1998), doi:10.1029/98JD01644
12. Kaufman, Y.J., Ichoku, C., Giglio, L., Korontzi, S., Chu, D.A., Hao, W.M., et al.: Fire and smoke observed from the Earth Observing System MODIS instrument-products, validation, and operational use. International Journal of Remote Sensing 24(8), 1765–1781 (2003)
13. Giglio, L., Csiszar, I., Morisette, J.T., Schroeder, W., Morton, D., Justice, C.O.: Active Fire Detection with the Advanced Space borne Thermal Emission and Reflection Radiometer (2006) (manuscript in preparation)
14. Giglio, L., Descloitres, J., Justice, C.O., Kaufman, Y.: An enhanced contextual fire detection algorithm for MODIS. Remote Sensing of Environment 87(2-3), 273–282 (2003)
15. Pu, R., Li, Z., Gong, P., et al.: Development and analysis of a 12-year daily 1-km forest fire dataset across North America from NOAA/AVHRR data. Remote Sensing of Environment 108(2), 198–208 (2007)
16. Chuvieco, E., Salas, F., Vega, C.: Remote Sensing and GIS for long-term fire risk mapping. In: Chuvieco, E. (ed.) A Review of Remote Sensing Methods for the Study of Large Wild Land Fires, pp. 91–107 (1997)
17. Sağlam, B., Bilgili, E., Durmaz, B.D., et al.: Spatio-temporal analysis of forest fire risk and danger using LANDSAT Imagery. Sensors 8(6), 3970–3987 (2008), doi:10.3390/s8063970
18. Hosmer, D.H., Lemeshow, S.: Applied logistic regression. Wiley series in probability and mathematical statistics, p. 307. Wiley, NewYork (1989)
19. Pontius, R.G., Schneider, L.C.: Land-use change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. Agriculture, Ecosystems and Environment 85(1-3), 239–248 (2001)

Eco-environment Assessment in Gannan Former Central Soviet Area—A Case Study in Ruijin City

Zhubin Zheng[1] and Jingli Ren[2]

[1] School of Geography and Planning, Gannan Normal University, Ganzhou., China
{zhengzhubin,jingliren321}@gmail.com
[2] School of Journalism and Communication, Gannan Normal University, Ganzhou., China

Abstract. Eco-environment change, as one of the key impacts of the urbanization and economy development, has taking place at an unprecedented rate around China in the past few decades. This paper explores the method of comprehensive assessment of extracted eco-environmental factors and sets up an assessment model applying the eco-environment assessment method. Understanding the change of eco-environment on county level city is crucial for improving the ecology and sustainability for Jiangxi province, in the undeveloped areas in central China. With the aid of an integrated GIS/RS-based approach, this study investigated how many ecological factors would affect eco-environmental change in the Ruijin, a county level city in the region of Jiangxi province, based on the analysis of land surface temperature (LST), normalized difference vegetation index (NDVI), vegetation fraction (Fv), tasseled cap brightness (TCB) considered as soil brightness, tasseled cap wetness (TCW) considered as soil wetness, DEM and Slope. Three Landsat TM images acquired on November 2, 2000, November 3, 2006 and January 14, 2010 were used to estimate LST, NDVI, Fv, TCB and TCW. These data in the new form were then combined with the extracted eco-environment key factors by way of re-projection to form comprehensive image . Our results have showed that, although there are significant variations in LST at a given fraction of vegetation on a per-pixel basis, NDVI, Fv, TCB and TCW are all good predictors of eco-environment on the regional scale. In north, southeast of the study area, the quality of eco-environment assessment is generally better than the other region, such as central and west directions. Meanwhile the place out of the central urban area, special in the northwest, due to the high altitude, less human intervention, eco-environment in most of these regions is good. Thus, it shows that quality of the eco-environment assessment in Ruijin city is conformable with the actual situation and feasible. Finally, the quality of eco-environment assessment in Ruijin city was extracted through an RS-based model in the terms of maps and tables.

Keywords: Eco-Environment Assessment, Ruijin city, Assessment factors, RS, GIS.

1 Introduction

Regional eco-environment dynamic monitoring and trend forecasting is an important studies in eco-environmental directions. According to this method, continuous

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 193–203, 2013.

compilation map of different periods of eco-environment assessment, will get information on dynamics eco-environment monitoring, which might be predicted trends in eco-environment system. In recent years, Remote Sensing and GIS have been commonly used to identify, characterize and measure eco-environmental factors, for example, soil, geology, water, vegetation, transportation, and infrastructures. Eco-environment research has been conducted in many countries such as the United States, Canada, Australia and Russia since the 1980s, and China started to catch up in recent years. Scholars have advocated establishing a national resources environment database.

Ruijin city, as the central part of the Former Central Soviet Area, has not been researched yet. Furthermore, this year is the occasion of the first anniversary of introduction of Several Opinions of the State Council on Supporting Gannan and Some Other Former Central Soviet Areas' Revitalization and Development. Ruijin city, also famous by the “river sources” and the “ecology sources” of Ganjiang river during the development of the eco-system, has a very special ecological status, functions and effects, which has significance for the Poyang Lake Ecological Economic Zone and Jiangxi survival and development. The ecological benefits from the protection of the eco-system in Ruijin can influence the sustainable development of midstream and downstream areas of the Ganjiang River, even the Poyang Lake directly as well as the local economic and social development. So far, what’s the status of eco-environment in the Ruijin city, how to protect it? These become questions for realistic significance by improving the sustainable development in Jiangxi province.

2 Study Area and Data

2.1 Study Area

The study area (115°42′ E-116°22′ E , 25°30' N-26°20′ N) is Ruijin city, which is the capital of Former Central Soviet Area. Ruijin is a county-level city of Ganzhou in Jiangxi Province, southeast of China, which bordering Fujian Province and south-eastern Jiangxi. With a surface of approximately 2448 km^2 (Fig.1). It is most famous as one of the earliest centers of Gannan Former Central Soviet . In 1931, under Mao Zedong's leadership, the Chinese Soviet Republic was established here. It is from here that the famed "Long March" began in 1934. Ruijin city is in the transition zone of Central China climate zone and South China climate zone, it located in the subtropical monsoon humid climate with a mean annual temperature of 18.9 °C (from minimum 7.6 °C to maximum 28.5 °C), and a mean annual precipitation of 1710 mm(data calculated from the observation of 1951 to 2008) [1].In this area, Soils are predominantly red soil, yellow soil and paddy soil with native vegetation is characterized by the subtropical evergreen broadleaved forest and the evergreen broadleaved-deciduous broadleaved mixed forest. There has abundant rainfall , adequate light , four distinct seasons with the mean frost-free period of 286 days and mean annual number of 163.7 rainy days .

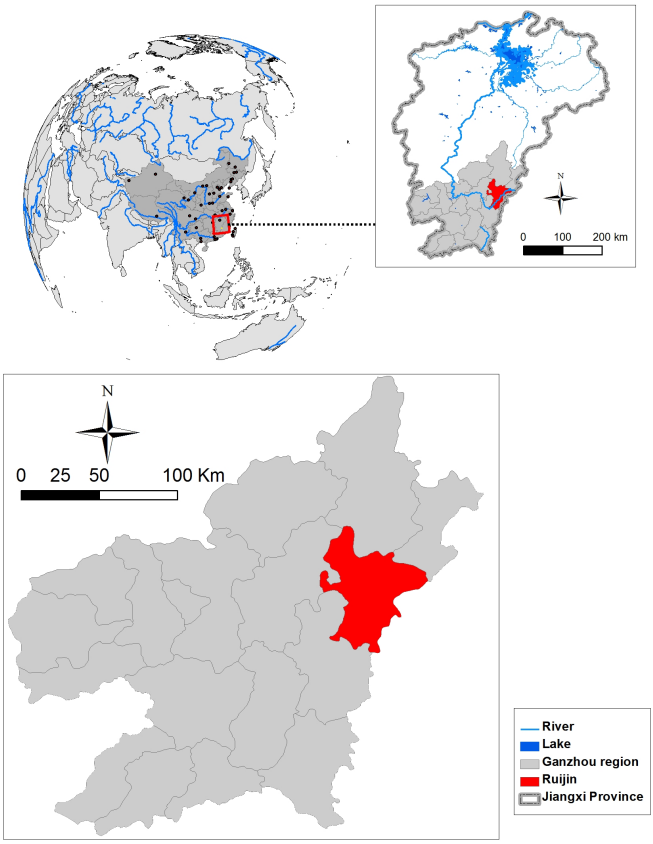


Fig. 1. Location map of Ruijin city

2.2 Data

The Thematic Mapper (TM) is an advanced, multispectral scanning, Earth resources sensor designed to achieve higher image resolution, sharper spectral separation, improved geometric fidelity and greater radiometric accuracy and resolution than the MSS sensor. TM data are sensed in seven spectral bands simultaneously. Band 6 senses Thermal (heat) Infrared Radiation(TIR). A TM scene has an Instantaneous Field Of View (IFOV) of 30m x 30m in bands 1-5 and 7 while band 6 has an IFOV of 120m x 120m on the ground [2]. This data was used as supporting land cover information, evaluate the forest cover rate and extracting other eco-environment factors at higher resolution for the automated classification.

In this research, three Landsat-5 Thematic Mapper (TM) imageries (Row/Path: 042/121), which were collected on November-02-2000, November-03-2006 and January-14-2010, respectively, were acquired through U.S. Geological Survey's Global Visualization Viewer (http://glovis.usgs.gov/) with very low cloud/haze cover (less than 10%) , the imagery's details can be seen in Table 1.

Table 1. Landsat TM imagery used in this study

Path/row	Acquisition date (MM/DD/ YY)	Spectral bands	Spatial resolution of the bands (m)
12142	11022000	VNIR-SWIR-TIR	30m,30m,120m separately
12142	11032006	VNIR-SWIR-TIR	30m,30m,120m separately
12142	01142010	VNIR-SWIR-TIR	30m,30m,120m separately

These images were further rectified to the Universal Transverse Mercator project system (UTM Zone, N50, WGS84), and resampled using the cubic convolution algorithm with a pixel size of 30 by 30 m for all bands. The root mean squared errors were smaller than 0.5 pixel (7.5 m) for each of the two images.

3 Methods

3.1 Radiometric Calibration

Usually, radiometric calibration and atmospheric correction are the most important steps in image preprocessing before extracting vegetation fraction and estimating land surface temperature. In this study, radiometric calibration and atmospheric correction was conducted with the image-based dark-object subtraction method due to its simplicity [3].

Radiometric calibration is the processing of converting Digital Number (DN) value of the image to the apparent reflectance. In order to doing it, it should be followed three steps:

First, converted the DN values of the image into the radiance values. Second, converted the radiance values into apparent reflection. Therefore, Apparent reflection for Landsat TM imagery was calculated as:

$$\rho = \pi * (\text{gain} * \text{DN} + \text{offset}) * d^2 / (\text{ESUN} * \cos(\theta)) \quad (1)$$

Where ρ is the planetary TOA re reflectance, d is the earth–sun distance, ESUN is the mean exo-atmospheric solar irradiance in W /($m^2 \cdot \mu m$), and θ is the solar zenith angle. Offset is the slope of the radiance/DN conversion.

After the implementation of radiometric calibration, all the images were rectified and geo-referenced to the UTM map projection(North 50 Zone) prior to interpretation. Subsequently, the images were resampled to 30 m using the nearest Neighbor algorithm to keep the unchanged original brightness values of pixels, and the RMSE were both found within 1 pixel. The image processing and data manipulation were conducted using algorithms supplied with the ENVI 5.0 image processing software. Furthermore, ARCGIS 10.1 was used for spatial analysis.

3.2 Retrieval of Land Surface Temperature

So far there have been several well developed algorithms for retrieval of LST from Landsat TM/ETM+ data, including mono-window algorithm [4], single-channel

algorithm [5]. However, there are some difficulties constraining their application. For instance, it is difficult to acquire near real time atmospheric profile data when the satellites pass over the study area. Alternatively, in this study the image-based method, which were successfully used in some typical case studies, is adopted to retrieve LST due to its simplicity and validity. Prior to retrieval of LST, a quadratic model was used to convert the digital number (DN) of Landsat TM thermal TIR band into radiant temperatures [6]:

$$T_B = 209.831 + 0.834DN - 0.00133DN^2 \quad (2)$$

For Landsat TM TIR band, another model was used as follows:

$$L_\lambda = \text{gain} \times DN + \text{offset} \quad (3)$$

Where L_λ is the radiance of thermal band pixels in $W/(m^2 \cdot sr \cdot \mu m)$, gain is the slope of the radiance/DN conversion function, and offset is the slope of the radiance/DN conversion.

Adopting the conversion formula, the spectral radiance was then converted to at-satellite brightness temperature under the assumption of uniform emissivity [7].

$$T_B = K_2 / \ln(1 + (K_1 / L_\lambda)) \quad (4)$$

Where T_B is the effective at-satellite temperature in K, both K1 and K2 are pre-launch calibration constants (As TM: K1= 607.76 $W/(m^2 \cdot sr \cdot \mu m)$, K2=1260.56 K; As ETM^+: K1= 666.09 $W/(m^2 \cdot sr \cdot \mu m)$, K2=1282.71 K), $L_\lambda$is the spectral radiance at the sensor's aperture in W/(m2 . sr . μm)

According to Eq.(**4**), the temperature values obtained above were referenced to a black body, which is quite different from the properties of real objects. Therefore, correction of spectral emissivity (e) is a must. Furthermore, the emissivity corrected land surface temperature was computed as follows:

$$T_s = T_B / (1 + (\lambda \times T_B / \alpha) \ln\varepsilon) \quad (5)$$

Where T_s is the surface radiant temperature in Kelvin (K), T_B is the black body temperature in Kelvin(K), the wavelength of emitted radiance, herein, $\lambda = 11.5\mu m$ (Markham and Barker, 1985), $\alpha = hc/K$ (1.438×10^{-2}mK), h = Planck constant($6.626\times10^{-34}Js^{-1}$), and c = velocity of light ($2.998 \times 10^8 ms^{-1}$), K = Boltzman constant ($1.38 \times 10^{-23} JK^{-1}$), ε = surface emissivity.

3.3 NDVI, Impervious Surface Fraction and Vegetation Fraction Calculation

In this study, based on the Normalized Difference Vegetation Index (NDVI), fractional vegetation cove was used as the key indicator describing intensity of human activities and land surface properties of the study area [8].

NDVI for Landsat TM imagery was calculated as follows:

$$NDVI = (\rho_{NIR} - \rho_{RED}) / (\rho_{NIR} + \rho_{RED}) \quad (6)$$

Where ρ_{NIR} and ρ_{RED} are the reflectance values in the near-infrared and red bands of Landsat TM imagery.

The scaled value (N*) was computed as follows:

$$N^* = (NDVI - NDVI_0)/(NDVI_S - NDVI_0) \quad (7)$$

Where NDVIs and $NDVI_0$ are the NDVIs in complete vegetation cover region and no vegetation cover region respectively.

Further, fractional vegetation cover (Fr) was computed as follows:

$$F_r = fN^* \times N^* \quad (8)$$

Where fN* refers to a polynomial fit for Fr as a function of N*, and based on a consistent relationship between F_r and N*, Fr is computed as the square of N* [9].

3.4 TCB (Tasseled Cap Brightness) and TCW(Tasseled Cap Wetness)

The Tasseled Cap Transformation in remote sensing is the conversion of the readings in a set of channels into composite values; i.e., the weighted sums of separate channel readings. One of these weighted sums measures roughly the brightness of each pixel in the scene, and another of these composite values might represent the degree of the wetness of the soil [10].

The Brightness and Wetness weights for the Thematic Mapper(TM) bands were calculated as follows [11]:

$$\text{Brightness} = 0.3037 * (TM1) + 0.2739 * (TM2) + 0.4743 * (TM3) + 0.5585 * (TM4) + 0.5082 * (TM5) + 0.1863 * (TM7) . \quad (9)$$

$$\text{Wetness} = 0.1509 * (TM1) + 0.1973 * (TM2) + 0.3279 * (TM3) + 0.3406 * (TM4) - 0.7112 * (TM5) - 0.4572 * (TM7) . \quad (10)$$

3.5 Model of the Eco-environment Assessment

The model and index system is the key to the eco-environment assessment. Because of the important of the eco-environment factors and its architecture, it needs a deeply understanding and necessary research for the area. As well as its complexity and uncertainty, combined with the actual situation of the study area and the existing data, we choose the six factors: land surface temperature (LST), vegetation fraction (Fv), tasseled cap brightness (TCB) considered as soil brightness, tasseled cap wetness (TCW) considered as soil wetness, DEM and Slope as the key factors of the eco-environment assessment.

The Comprehensive Assessment Model of Ecological-environment assessment for Ruijin city was calculated as follows:

$$E = \sum_{i=1}^{n} P_i \bullet A_i \tag{11}$$

Where: E is Comprehensive Assessment Index of the ecological-environment
P_i is value of the i-th principal component.
A_i is weight value (contribution rate) of the i-th main component.

We classified the result of eco-environment into 5 classes: Unacceptable, Acceptable, Satisfied, Good, Excellent, it depended on the quality value we got combined with the results of field survey. Tab. 2 shows the detail of quality classification.

Table 2. Index types classified of Eco-environment assessment in Ruijin city

Classification	Comprehensive Evaluation Index	Characteristic or description?
Excellent	≥60	Undisturbed , ecosystem structure and function is excellent
Good	45～60	Almost undisturbed , ecosystem structure and function is good
Satisfied	35～45	Little bit disturbed , ecosystem structure and function is satisfied
Acceptable	25～35	Little bit disturbed , ecosystem structure and function is acceptable
Unacceptable	≤25	Disturbed , ecosystem structure and function is unacceptable

4 Results and Discussion

Table 3&4 shows the quality of Eco-environment assessment of Ruijin city in 2000, 2006, 2010. From Figs. 2, Table 3&4, it shows quite different patterns of Eco-environment assessment over the study period. In Table 3 and Figure 3, it can be clear seen the results of Eco-environment assessment in Ruijin city and its changes from 2000 to 2010.

From the temporal view, in 2010 eco-environment in Ruijin city is good type, this can be seen from ratio more than 50% of Excellent plus Good type. But it also shows some signs of deterioration : Excellent, Good, Satisfied, acceptable, Unacceptable 5 types of evaluation performance in 2000 accounted of the entire Ruijin city area, 33.35%, 45.91%, 15.86%, 3.55%, 1.18% respectively; In 2006, the Excellent and

Table 3. Area and percentage for Eco-environment assessment in Ruijin city 2000～2010

Quality type	2000		2006		2010	
	Area (km²)	Percent	Area (km²)	Percent	Area (km²)	Percent
Excellent	810.75	33.35	677.84	27.88	488.71	20.10
Good	1116.08	45.91	985.77	40.55	789.09	32.46
Satisfied	385.65	15.86	470.85	19.37	638.93	26.28
Acceptable	86.23	3.55	226.87	9.33	391.99	16.13
Unacceptable	28.59	1.18	65.57	2.70	118.58	4.88

Good proportion dropped to 27.88%, 40.55% but the Satisfied acceptable Unacceptable proportion rose to 19.37%, 9.33%, 2.70% ; In 2010 Excellent Good continue to decline to 20.10%, 32.46%, Satisfied, acceptable, Unacceptable proportion continues to rise to 26.28% , 16.13% and 4.88%.

Table 4. Change rate for Eco-environment assessment in Ruijin city 2000～2010

Quality type	2000-2006		2006-20010		2000-2010	
	Change rate	Annual rate	Change rate	Annual rate	Change rate	Annual rate
Excellent	-16.37	-2.73	-27.93	-6.98	-39.72	-3.97
Good	-11.66	-1.94	-19.96	-4.99	-29.30	-2.93
Satisfied	22.10	3.68	35.69	8.92	65.68	6.57
Acceptable	163.09	27.18	72.78	18.19	354.57	35.46
Unacceptable	129.35	21.56	80.86	20.21	314.79	31.48

Fig.4 and Tab.4 show the changeable data for the Eco-environment assessment result from 2000-2006 and 2006-2010. From 2000-2006, it shows that the acceptable type is the maximum changeable types, the average annual change rate increased 10.47 %, and the unacceptable type is the second maximum changeable types, the average annual change rate increased 10.47%. Nevertheless, it changed from 2006-2010. the unacceptable type became the maximum changeable types since the average annual change rate increased 20.21%, and the acceptable type was the second maximum changeable type, the average annual change rate increased 18.19%.

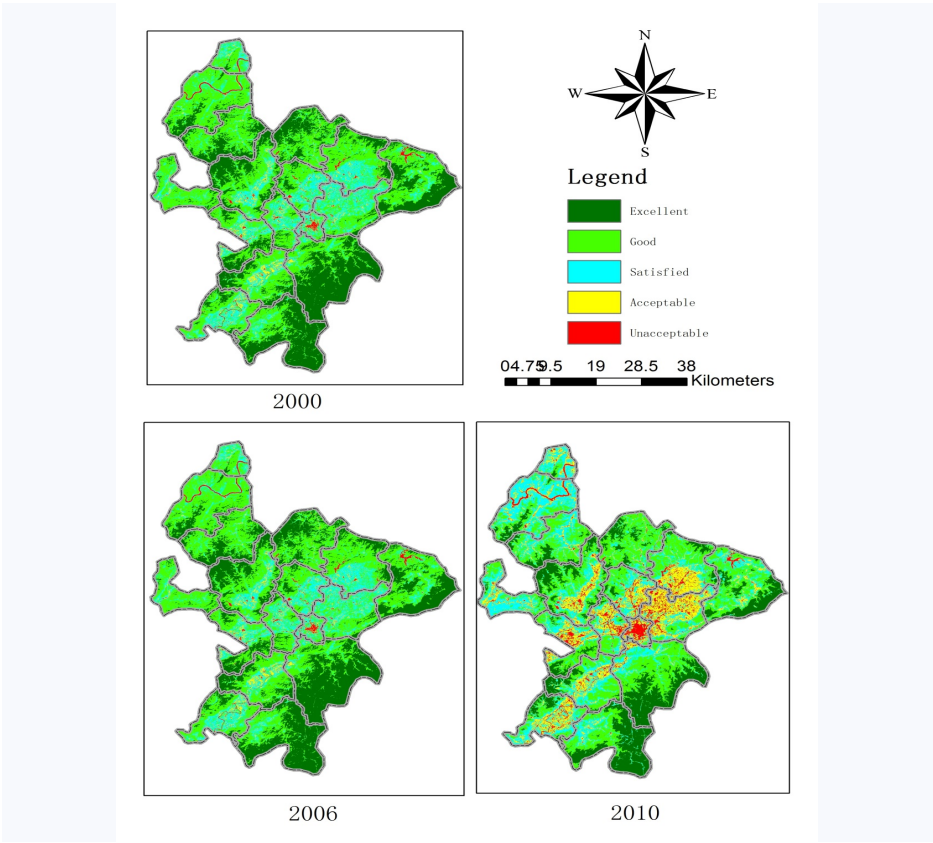


Fig. 2. Change Map of Eco-environment assessment in Ruijin city, 2000-2010

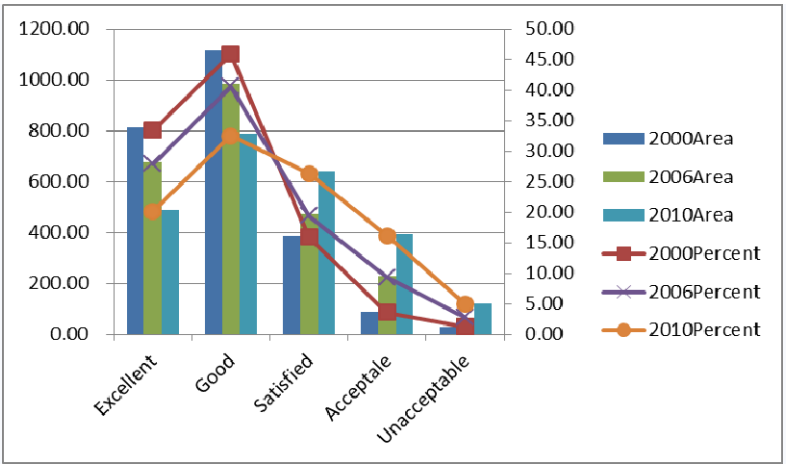


Fig. 3. Map of the area and percentage for Eco-environment assessment in Ruijin city

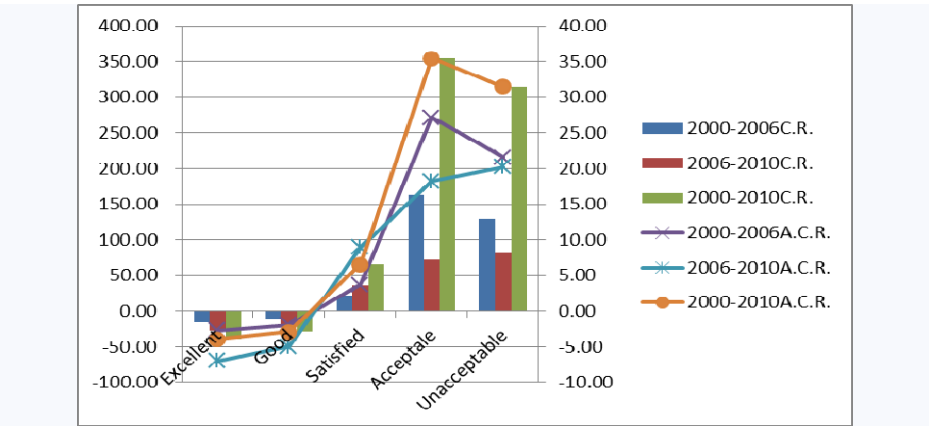


Fig. 4. Map of the change rate for Eco-environment assessment in Ruijin city

The good type was the minimum changeable types, the average annual change rate decreased 1.94% from 2000-2006, and 4.99% from 2006-2010. The same as good type, excellent type was the second minimum changeable type, since the average annual change rate decreased 2.73% from 2000-2006, and 6.98% from 2006-2010. For the midst was the Satisfied type, the average annual change rate increased 3.68% from 2000-2006, and 8.92% from 2006-2010.

From the spatial view, Fig.2, Tab. 3&4 show that the unacceptable and acceptable types are most in urban area of source area of Ruijin city from 2000-2010. With the development of urbanization in the past decade, the unacceptable and acceptable type's total area is increasing very fast in Ruijin central urban area, special from 2006-2010. The Excellent and Good types are mainly distributed in the mountainous region, such as east, northeast and southeast of Ruijin city area. The satisfied type mostly distributed in the border area of urban and rural places.

5 Conclusion

Evaluation results from the above data can know, quality of eco-environment assessment in Ruijin city from 2000-2010 basic distribution trend is: In north, southeast of the study area, the quality of eco-environment assessment is generally better than the other region, such as central and west directions. The unacceptable and acceptable types of the quality result mainly in residential areas, along the rivers, construction land, railway, highway, poor natural conditions and other human activities,. This is mainly due to the central and west of Ruijin city, Xianghu town, Yeping xiang and Yunshishan xiang, are under the pressure of population and economic development, such as urban development and cultivated land, garden land, forest land, meadow, deforestation and excessive exploitation as well as the natural fire, caused serious eco-environment destruction. Meanwhile Ridong xiang, Ruilin xiang and other places out of central urban area, due to the high altitude, less human intervention, eco-environment in most of these regions is good. Therefore, it shows that quality of

the eco-environment assessment in Ruijin city is conformable with the actual situation and feasible.

Acknowledgement. Foundation item: Under the auspices of Humanities and Social Sciences Project of Jiangxi Province (No. JC1206); China Scholarship Council; Young college teachers' developmental program of Jiangxi Province.

References

1. Baidu Encyclopedia, http://baike.baidu.com/view/44669.htm
2. National Aeronautics and Space Administration, http://landsat.gsfc.nasa.gov/about/tm.html
3. Chander, G., Markham, B.L., Helder, D.L.: Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+ and EO-1 ALI sensors. Remote Sensing of Environment 113(5), 893–903 (2009)
4. Qin, Z., Karnieli, A., Berliner, P.: A mono-window algorithm for retrieving land surface temperature from Landsat TM data and its application to the Israel–Egypt border region. International Journal of Remote Sensing 22(18), 3719–3746 (2001)
5. Jimenez-Munoz, J.C., Sobrino, J.A.: A generalized single channel method for retrieving land surface temperature from remote sensing data. Journal of Geophysical Research 108(D22), 4688 (2003), http://dx.doi.org/10.1029/2003JD003480
6. Malaret, E., Bartolucci, L.A., Lozano, D.F., Anuta, P.E., McGillem, C.D.: Landsat-4and Landsat-5 Thematic Mapper data quality analysis. Photogrammetric Engineering and Remote Sensing 51, 1407–1416 (1985)
7. Landsat Project Science Office. Landsat 7 Science Data User's Handbook. Goddard Space Flight Center, NASA, Washington, DC (2002), http://ltpwww.gsfc.nasa.gov/IAS/handbook/handbook_toc.html
8. Kerr, J.T., Ostrovsky, M.: From space to species: ecological applications forremote sensing. Trends in Ecology and Evolution 18, 299–305 (2003)
9. Gillies, R.R., Kustas, W.P., Humes, K.S.: A verification of the 'triangle' method for obtaining surface soil water content and energy fluxes from remote measurements of the Normalized Difference Vegetation Index(NDVI) and surface radiant temperature. International Journal of Remote Sensing 18, 3145–3166 (1997)
10. Kauth, R.J., Thomas, G.S.: The tasseled cap – a graphic description of the spectraltemporal development of agricultural crops as seen in Landdat. In: Proceedings on the Symposium on Machine Processing of Remotely Sensed Data, June 29-July 1, pp. 41–51. LARS, Purdue University, West Lafayette, Indiana (1976)
11. Crist, E.P., Cicone, R.C.: A physically-based transformation of Thematic Mapper data – the TM Tasseled Cap. IEEE Trans. on Geosciences and Remote Sensing GE-22, 256–263 (1984)

An Algorithm of the Constrained Construction for Terrain Morse Complexes

Hongbin Wang, Xuesheng Zhao, Chunkang Zhang, and Ting Xiong

College of Geoscience and Surveying Engineering, China University of Mining & Technology (Beijing), Beijing, China
{whbcumtb,chkang.chd}@163.com, zxs@cumtb.edu.cn

Abstract. The correct connection of the topological relationships between critical points (or lines) is the basis of the earth's surface description, terrain topological simplification, or geomorphic generalization of relief. However, the intersection of valley and ridge line at regular points often occurs in the existing algorithms of constructing terrain Morse complexes. In this paper, a novel universal algorithm of the constrained construction for terrain Morse complexes is presented. In our approach, the separatrix of descending (or ascending) Morse complex is regarded as the constrained boundary of extracting the separatrix of the dual Morse complex, and the terrain feature line starting from end (or start) saddle coincides exactly with the "macro-saddle line". As a result, the intersections are prevented absolutely and the macro-saddles can be identified to achieve complete decomposition of the whole terrain surface. In the end, an experiment is done to validate the correctness and feasibility of this algorithm.

Keywords: intersection, macro-saddle, terrain morphology, Morse complex.

1 Introduction

Terrain morphological information consists of feature points (pits, peaks and passes), feature lines (such as ridges and ravines), or segmentation of the terrain in regions of influence of pit and peak or in regions of uniform gradient field [8]. Extracting and representing the terrain morphology is widely applied in several contexts including terrain analysis and understanding, knowledge-based reasoning and hydrological simulation, to gain and maintain the knowledge about terrain surface.

Over the years, based on Morse (or Morse-Smale) complex, there has been a lot of research focusing on the representation of the terrain morphology and many algorithms have been proposed. Generally speaking, these algorithms can be classified into *boundary-based* or *region-based* techniques [5]. The former [8, 12, 2, 9, 1, 18] extracts the feature lines before the construction of the complex, while the latter [10, 6, 14, 3] constructs the complex by region growing without the extraction of feature lines. In fact, these techniques are rooted in Morse theory and try to simulate the decomposition of a terrain induced by C^2 Morse functions in the discrete grid, often known as *Regular Square Grid* (RSG) or *Triangulated Irregular Network* (TIN).

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 204–213, 2013.

However, some issues will occur in the actual implementation. First of all, valley and ridge often intersects at regular points, which is inconsistent with Morse theory and induces the error of acquiring the main terrain features in the applications such as topological simplification, since it distorts the correct connection of the topological relationships between critical points. Some of those algorithms, like [10] ignore this issue, while the others, such as [15, 12, 2], either provide a relatively simple but not comprehensive discussion or increase the computing cost, or the scope is limited, although have considered to prevent it. Secondly, the macro-saddle cannot be extracted and identified effectively, though the concept has been proposed in some literatures (e.g. [8]). It is just the macro-saddle that extends the basic operators for topological simplification [7, 8], and complicates the multi-resolution construction for the real terrain morphology. In this paper, a novel universal algorithm for terrain morphology is proposed in order to overcome the above shortcomings of the existing techniques.

The remainder of this paper is organized as follows. In Section 2, we introduce some basic notions on Morse Theory. In Section 3, we review the related works on preventing the intersections after addressing the reasons and types of which appears in the real terrain. In Section 4, we present our algorithm in details. In Section 5, we show some experimental comparisons about the terrain morphology, In Section 6, we draw some concluding remarks.

2 Morse Theory

Morse theory [11, 17] is a powerful tool to capture the morphology of manifold on which a scalar function f is defined. It decomposes the shape into cells in which the gradient flow is uniform, and encodes the adjacencies of these cells in a complex which represents the topology as well as the geometric properties of the gradient of f.

Let f be a C^2-differentiable real-valued function defined over a domain D in the 2D space. A point p of D is a *critical point* of f if and only if the gradient of f vanishes on p. Function f is said to be a *Morse function* when all its critical points are non-degenerate. This implies that the critical points of f are isolated. The number of negative eigenvalues of $Hess_p f$ is called the *index* of a critical point p. There are three types of non-degenerate critical point p: a minimum (pit), a saddle (pass), or a maximum (peak) if and only if p has index 0, 1 or 2, respectively. Points which are not critical are said to be *regular*.

An *integral line* of f, going through a point p, is a maximal path which is everywhere tangent to the gradient of f. Integral lines that converge to (originate from) a critical point p of index i form an i-cell ((2-i)-cell) called the *descending (ascending) cell* of p. The collection of all descending cells form a Euclidean cell complex, called the *descending Morse complex* (Figure 1a), and the collection of all ascending cells form also a Euclidean cell complex, called the *ascending Morse complex* (Figure 1b), which is dual with respect to the descending Morse one. A Morse function f, whose descending and ascending Morse complexes intersect only transversally, is called a *Morse-Smale function*. This means that the intersection of the edges of the descending and ascending cells is a saddle point. In this case, we can

define a complex, called the *Morse-Smale complex*, whose cells are the connected components of the intersections of the descending and ascending cells (Figure 1c). Integral lines connecting saddle points to other critical points are called *separatrices*. All separatrices forms a network, called the *critical net*. It is the 1-skeleton of the Morse-Smale complex.

In terrain analysis, the separatrix starting from saddle to peak is called a *radge line*, while the one starting from saddle to pit is called a *ravine line*. Essentially, the *surface network* [13, 16] consisted of ridges and ravines is the critical net. According to Morse theory, the case that valley and ridge line intersect at regular points is impossible. In this paper, both Morse (or Morse-Smale) complex and critical net (or surface network) for representing the terrain morphology are called the *terrain morphological model*.

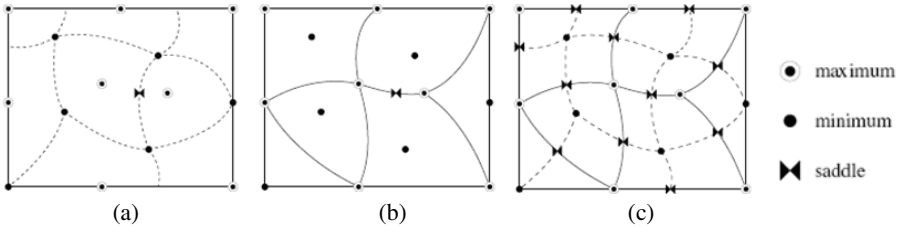


Fig. 1. Morse and Morse-Smale complex: (a) Descending Morse complex, (b) Ascending Morse complex, (c) Morse-Smale complex

3 Related Work

For the convenient and clear discussion, we reclassify those algorithms mentioned above for terrain morphology into the ones based on the original grid vertex [9, 1, 18, 10] and others based on the actual gradient of terrain surface [12, 2], according to the routing location of the separatrix in the construction of terrain morphological model.

The routing path of the separatrix extracted by the first group algorithms is strictly located on the original terrain gird vertexes, which grows the complex cell directly by including original triangles in a fast computation speed. However, just because of this routing method, the extracted ridge and ravine are only the approximate but not really tangent to the actual gradient vector field and often intersect at the regular points. Probably, there are four types of intersections between a ravine and ridges in the real terrain morphological model constructed by this approach, as shown in Figure 2. Here, the red line represents the ridge while the blue one the ravine.

To prevent the intersection, Schneider [15] proposed a strategy that if it occurs, the routing path of one of the two feature lines can be alternated into another. Nevertheless, there are some shortcomings in this technique. Firstly, intersection must be identified in advance through the comparison between all the ravines and ridges in the critical net, and it will increase the calculation cost when there are plenty of feature lines. Moreover, it is only the discussion about the intersection at a single regular point, without considering the complexity and variety of the real terrain

morphology (as shown in Figure 2). Furthermore, it will increase a lot of redundant computation and reduce the calculation efficiency if using this method to rectify the intersections caused by those region-based algorithms, because the separatrix requires to be extracted to identify the intersections after both descending and ascending Morse complex are constructed, and one Morse complex need to be reconstructed again using the rectified feature lines.

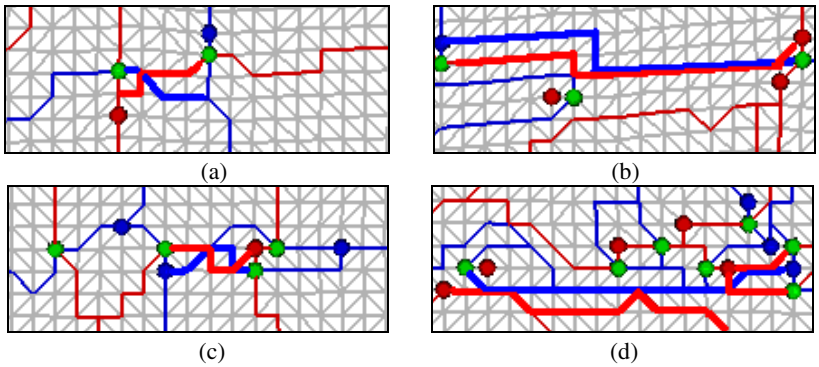


Fig. 2. Four types of intersections caused by traditional algorithms: (a) Intersection at a single point, (b) Intersections forming a line, (c) Multiple intersections, (d) Multi-intersections between a ravine and some ridges

By the second group algorithms, the actual gradient of terrain is calculated and the feature lines are extracted according to the gradient vector, simulating the smooth case as accurately as possible. As a result, intersections might be prevented to some extent, but there still remain some problems. First of all, the computation of the real gradient vector of every triangle is more complex and costly than that of the approximate one along the original terrain gird vertex. In addition, since some triangles are spitted and terrain grid is destroyed, in order to guarantee the topological consistency, both the original vertexes and new path points need to be retriangulated, regarding the extracted feature lines as the triangulation constraint. Consequently, it also affects the construction efficiency. And even more important, the actual gradient calculation can not be done when encountering the horizontal areas that often appear in terrain.

4 An Algorithm for Constrained Construction

In this section, an algorithm for the constrained construction of terrain Morse complexes is addressed in detail to construct the correct connection of the topological relationships between critical points and to achieve a complete decomposition of the whole terrain surface.

4.1 The Main Principle and Steps

The principle of our algorithm is as follows: from the dual structure of the ascending and descending Morse complex (as shown in Figure 1), it can be observed that the separatrix of an ascending cell is consisted of ravines and saddles, and that the ridges starting from the boundary saddles to the central peak are routing within the cell. Similarly, the separatrix of a descending cell is consisted of ridges and saddles, the ravines starting from the boundary saddles to the central pit are routing within the cell.

Our algorithm mainly consists of there steps. Firstly, ascending (descending) Morse complex is constructed by the existing algorithm which can be either the region-based technique or the boundary-based one. Therefore, our algorithm is universal for the prevention of the intersections at regular points. Moreover, our approach is flexible for either the descending Morse complex or the ascending one can be built in the first place. Secondly, ridges (ravines), that is, the boundary of descending (ascending) Morse complex is extracted within every constructed ascending (descending) cell to prevent the intersection. That is to say, the separatrix of descending (or ascending) Morse complex is regarded as the constrained boundary of extracting the separatrix of the dual Morse complex. Here, we call this condition the *global constraint* of this algorithm for it is applied to the whole terrain surface. Finally, descending (ascending) Morse complex is constructed by taking advantage of the extracted feature lines and the terrain morphology is represented without intersections.

For the horizontal area in terrain, there is not yet a better method to deal with all the cases [19]. In our algorithm, lexicographic ordering [18] is adopted for its simplicity and high efficiency as well as the university.

4.2 Identifying the Macro-saddles

In the real terrain surface, there may be some macro-saddles, that is, one feature line has two saddles. By this algorithm, complete decomposition of the whole terrain surface can not be achieved without the consideration of the macro-saddle. So, all the macro-saddles are identified when extracting the feature lines in the second stage of our approach. For brevity, we firstly introduce two definitions.

Definition 1. Let point p be a vertex of the terrain grid, the adjacent points with p which are adjacent each other is called a *higher neighbor set of p* if there are higher than p, otherwise, called a *lower neighbor set of p* if they are lower than p.

The number of higher and lower neighbor set of a point on terrain surface is as different as the type of the point varies. For example, as shown in Figure 3, a saddle S has two higher and lower neighbor sets at least, respectively. Moreover, if the feature line routs on some vertex of the original terrain grid, ridge will pass through a vertex of the higher neighbor set while ravine through a vertex of the higher neighbor set.

Definition 2. If a feature line passes through a macro-saddle, the path between two saddles is called the *macro-saddle line*. The lower saddle is called the *start saddle* while the higher the *end saddle*. As a part path, a macro-saddle line plays the dual role of both ravine line and ridge line.

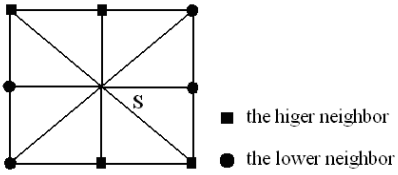


Fig. 3. The higher and lower neighbor sets of the saddle *S*

Here, the implementation of identifying the macro-saddle and extracting a ravine is detailed after the descending Morse complex is constructed before the ascending one (the case is similar that the ascending Morse complex is constructed before the descending one).

First of all, assign a macro-saddle-line flag with the value FALSE to every vertex. After the descending Morse complex is constructed, the cell saddles are extracted and this flag becomes TRUE for the boundary vertex of each descending cell.

Because of the strangulation [4], the saddles belongs to a cell are grouped into the boundary and the internal. Therefore, a cell saddle whether located on the boundary or not need be determined. If it is internal, select the lowest vertex of one lower neighbor sets of this saddle as the second point of the corresponding ravine path, then route like this in turn until reaching at a pit. In this way, a complete ravine is extracted. On the other side, if the cell saddle is a boundary saddle, all the vertexes of every lower neighbor sets of this saddle are checked according to the macro-saddle-line flag. If there is any vertex with the flag TRUE, it indicates that a macro-saddle is located on the boundary of the cell and this saddle is the end saddle. Here, starting from this saddle, extract the macro-saddle line, that is, select the vertexes with the TRUE flag as the path points until to the start saddle, and then store this macro-saddle line as a ravine if it had not been extracted. In this way, the ravine starting from end saddle coincides exactly with the macro-saddle line. So, we call this case the *local constraint* of our algorithm for the macro-saddle is only located in some region of the terrain surface. In the lower neighbor sets, if there is no vertex with the TRUE flag and all the vertexes belong to this cell, recursively select the lowest vertex as the path point of the ravine until reaching at a pit.

If all the ravines and macro-saddle lines of each cell have been extracted, the ascending Morse complex can be constructed with these feature lines as its separatrix. The flow chart of the above procedure is shown in Figure 4.

To satisfy the definition of the feature line, any of the macro-saddle lines still can be complemented for the endpoint is an extreme point. Here, we provide a method from the view of topological simplification. Specifically, for a macro-saddle line, if it to be a ridge, find out the two ridges starting from the end saddle and select the one with higher peak as the remaining path. Similarly, if it to be a ravine, find out the tow ravines starting from the start saddle and select the one with lower pit as the remaining path. In this way, the complete feature line is given a larger persistence [9] and can not be deleted easily in the multi-scale representation of terrain morphology.

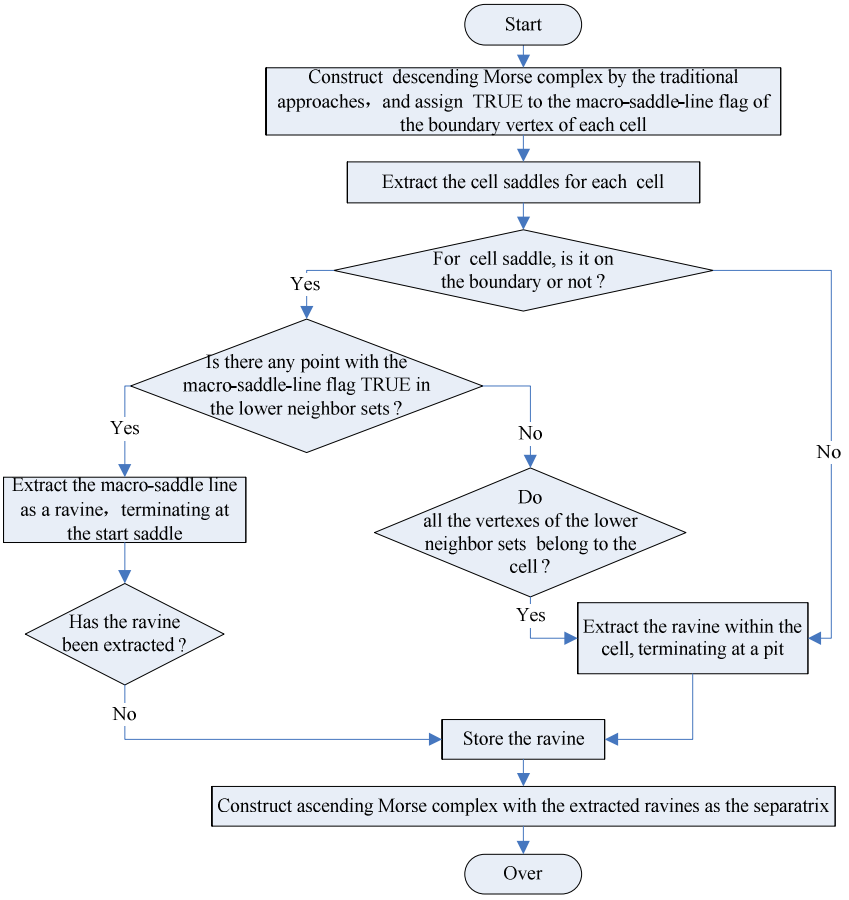


Fig. 4. The flow chart of the constrained construction for terrain Morse complexes

This algorithm is still high-efficiency, because the path points of the feature lines are located on the original vertex of the terrain grid and neither any Morse complex nor constrained Delaunay triangulation is reconstructed in the implementation.

5 Experiments and Analasis

To validate our algorithm, an experiment system is developed by using C++ and OSG (Open Source Graph), selecting Qinghai-Tibet Plateau of China as the experimental region, with the 90m-resolution DEM data from CGIAR-CSI GeoPortal (`http://strm.csi.cgiar.org`).

The critical net constructed by traditional algorithm [10] and our method are shown in Figure 5(a) and 6(a), of which the enlarged result of the local selected region by square frame are shown in Figure 5(b) and 6(b), respectively. Intersections appear in Figure 5(b) as marked by thick lines but not in Figure 6(b), so it indicates that intersection of ravine and ridge is prevented absolutely by our algorithm.

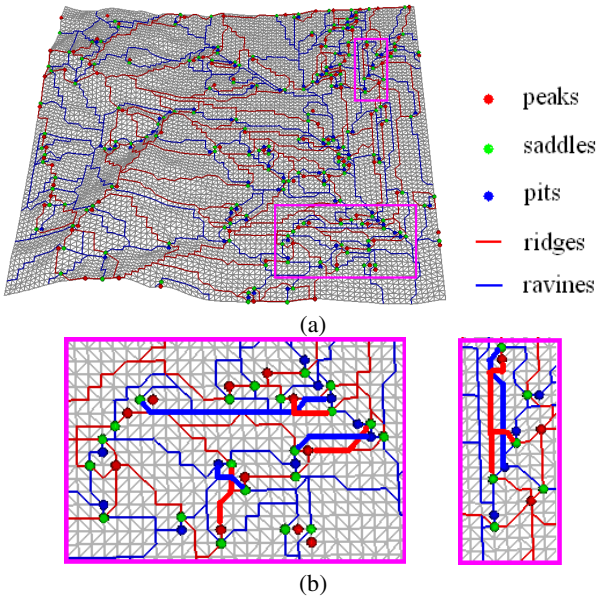


Fig. 5. Experimental region and intersections caused by traditional algorithm: (a) The terrain morphology by the STD algorithm, (b) The enlarged chart of local intersections

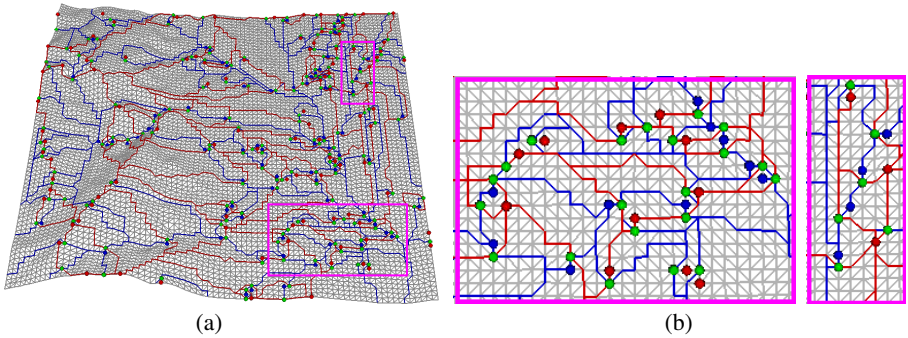
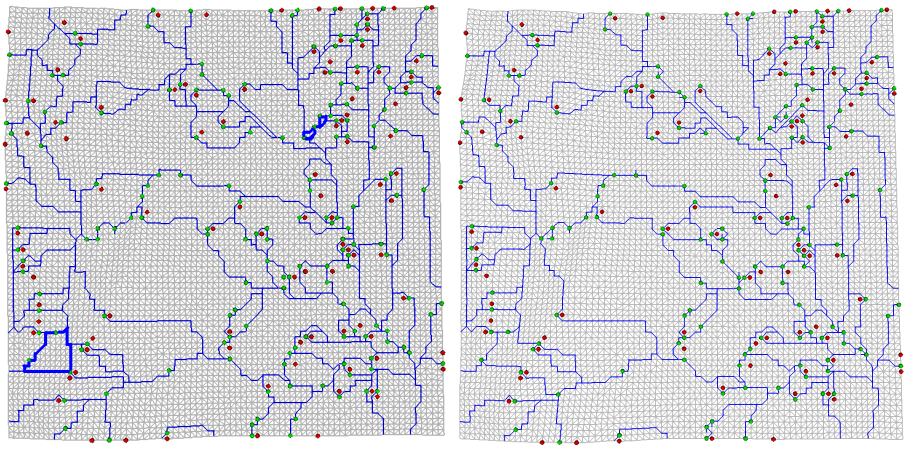


Fig. 6. Results of our algorithm: (a) The terrain morphology by our method, (b) The prevention of intersection

The ascending Morse complex constructed by the extracted ravines is shown in Figure 7. In Figure 7(a), without identifying the macro-saddles, there is no peak in each region with the thick blue boundary, where triangles do not belong to any ascending Morse cell and the decomposition of the whole terrain surface is not complete. On the contrary, as shown in Figure 7(b), after identifying the macro-saddles in our approach, not only every cell does not overlap each other, but also there is only one peak in each cell. Therefore, it reveals that the complete decomposition of the whole terrain surface is achieved by our algorithm.



(a) (b)

Fig. 7. Identifying the macro-saddles in the ascending Morse complex: (a) Before identification, (b) After identification

6 Concluding Remarks

Intersection at regular points between ravine and ridge is not only inconsistent with Morse theory, but also induce the error of acquiring the main terrain features in those applications such as the topological simplification or the geomorphic generalization of relief. In this paper, a novel universal algorithm of the constrained construction for terrain Morse complexes is proposed to construct the correct connection of the topological relationships between critical points and an experiment is designed to validate the correctness and feasibility of this algorithm, the main contributions of this paper are as followings:

1) This algorithm prevents the intersections absolutely by regarding the separatrix of descending (or ascending) Morse complex as the constrained boundary of extracting the separatrix of the dual Morse complex.

2) The macro-saddles can be identified to achieve complete decomposition of the whole terrain surface.

3) In our approach, neither any Morse complex nor constrained Delaunay triangulation is reconstructed, which optimizes the calculation steps in a higher efficiency and guarantees the consistency and integrity of the multi-resolution construction of the terrain morphology or the DEMs.

References

1. Bajaj, C.L., Shikore, D.R.: Topology Preserving Data Simplification with Error Bounds. Computers and Graphics 22(1), 3–12 (1998)
2. Bremer, P.T., Edelsbrunner, H., Hamann, B., Pascucci, V.: A Multi-resolution Data Structure for Two-dimensional Morse Functions. In: Turk, G., van Wijk, J., Moorhead, R. (eds.) Proceedings of the IEEE Visualization 2003, pp. 139–146. IEEE Computer Society (2003)

3. Bieniek, A., Moga, A.: A Connected Component Approach to the Watershed Segmentation. In: Mathematical Morphology and its Application to Image and Signal Processing, pp. 215–222. Kluwer Acad. Publ., Dordrecht (1998)
4. Bremer, P.: Topology-based Multi-resolution Hierarchies, PhD. University of California (2004)
5. Čomić, L., De Floriani, L., Papaleo, L.: Morse-Smale Decompositions for Modeling Terrain Knowledge. In: Cohn, A.G., Mark, D.M. (eds.) COSIT 2005. LNCS, vol. 3693, pp. 426–444. Springer, Heidelberg (2005)
6. Danovaro, E., De Floriani, L., Mesmoudi, M.M.: Topological Analysis and Characterization of Discrete Scalar Fields. In: Asano, T., Klette, R., Ronse, C. (eds.) Geometry, Morphology, and Computational Imaging. LNCS, vol. 2616, pp. 386–402. Springer, Heidelberg (2003)
7. Danovaro, E., De Floriani, L., Papaleo, L., Vitali, M.: Multi-resolution Morse-Smale Complexes for representing terrain morphology. In: Proceedings of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems, Seattle, Washington, Article No: 29 (2007)
8. De Floriani, L., Magillo, P., Vitali, M.: Modeling and Generalization of Discrete Morse Terrain Decompositions. In: 20th International Conference on Pattern Recognition, Istanbul, Turkey, pp. 999–1002 (2010)
9. Edelsbrunner, H., Harer, J., Zomorodian, A.: Hierarchical Morse Complexes for Piecewise Linear 2-manifolds. In: Proceedings 17th ACM Symposium on Computational Geometry, pp. 70–79. ACM Press (2001)
10. Magillo, P., Danovaro, E., De Floriani, L., Papaleo, L., Vitali, M.: A Discrete Approach to Compute Terrain Morphology. Computer Vision and Computer Graphics. Theory and Applications 21, 13–26 (2009)
11. Milnor, J.: Morse Theory. Princeton Univ. Press (1963)
12. Pascucci, V.: Topology Diagrams of Scalar Fields in Scientific Visualization. In: Rana, S. (ed.) Topological Data Structures for Surfaces, pp. 121–129. John Wiley and Sons Ltd. (2004)
13. Pfaltz, J.L.: Surface Networks. Geographical Analysis 8, 77–93 (1976)
14. Stoer, S.L., Strasser, W.: Extracting Regions of Interest Applying A Local Watershed Transformation. In: Proc. IEEE Visualization 2000, pp. 21–28. IEEE Computer Society (2000)
15. Schneider, B.: Extraction of Hierarchical Surface Networks from Bilinear Surface Patches. Geographical Analysis 37, 244–263 (2005)
16. Schneider, B., Wood, J.: Construction of Metric Surface Networks from Raster-based DEMs. In: Rana, S. (ed.) Topological Data Structures for Surfaces. John Wiley and Sons Ltd., Chichester (2004)
17. Smale, S.: Morse Inequalities for a Dynamical System. Bulletin of American Mathematical Society 66, 43–49 (1960)
18. Takahashi, S., Ikeda, T., Kunii, T.L., Ueda, M.: Algorithms for Extracting Correct Critical Points and Constructing Topological Graphs from Discrete Geographic Elevation Data. Computer Graphics Forum 14(3), 181–192 (1995)
19. Vitali, M., Floriani, L.D., Magillo, P.: Analysis and Comparison of Algorithms for Morse Decompositions on Triangulated Terrains. Technical report DISI-TR-12-03. DISI, University of Genova (2012)

MODIS Satellite Data Coupled with a Vegetation Process Model for Mapping Maize Yield in the Northeast China

Jiahua Zhang[1,2] and Fengmei Yao[3,*]

[1] College of Geosciences, Yangtze University, Wuhan, 430100, China
[2] Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094 , China
jhzhangcma@gmail.com
[3] College of Geosciences, Graduate University of Chinese Academy of Sciences, Beijing, 100049, China
yaofm@ucas.ac.cn

Abstract. In this study, the regionalestimation ofmaize yield was reported with integrating a process-based mechanism model and MODIS remote sensing data. AC4 plant photosynthetic pathway mode was developed and to substitute for the C3 plant photosynthetic pathway in the remote-sensing–photosynthesis –yield estimation for crops (RS-P-YEC) model, and the Harvest Index (HI) derived from the ratio of grain to stalk yield was adopted in the developed model. We performed maize yield simulation by using the developed model in the Northeast China (NEC) region from 2007-2009. The selected county-leveldata at from the NEC region was validatedwith the MODIS-simulated results. We found that that the correlation coefficientbetween the simulated yield and the statistics yield is high (R^2=0.637, n=69),and the spatial pattern of MODIS-simulated yield was agree with the statisticaldistribution in the NEC. It indicatedthe improved model has ability to estimateC4 crops in large area.

Keywords: BEPS model, Maize, C4 crop yield, MODIS data.

1 Introduction

Accurate estimation of the crop yield is of great importance for food production, grain policy making, andfood security warning [1,2]. The statistical models, light use efficiency models, and processed-based models have been used to predict crop yield in recent years [3-5].One kind of statistical models was established based on field experiments, which are time-consuming, expensive, and prone to large errors in estimating at the regional scale[6]. Others statistics models are established based on remote sensing technology to regionally predictcrop yield. The inverse information derived from remote sensing, for example, normalized difference vegetation index (NDVI) and LAIare key parameters of crop photosynthesis product accumulation, which areof importance for crop yield to a certain extent [7-9].The light use efficiency modelsuse the proportional relation between net primary productivity derived from

* Corresponding author.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 214–222, 2013.

light energy utilization ε and crop biomass or yield to spatially estimate crop yield[5,10], and the proportional relation could be built in different stages[11]. However, it keeps a problem that the specific ε does not distinguish the crop typesin the same region and change with the crop growth stages [12]. The proportional relation changes with regions and are vulnerable to the impact of other environment factors[13]and lack of mechanistic explanation to the crop ecological andphysiological changes.

Over the last decades, the process-based models have been developing to estimate crop yield. Especially, with development of the remote sensingtechnology, the regional estimation of crop yieldintegratedsatelliteinformation and process-based crop models has been performed [3,14,15]. The simplest is to use the remote sensing data, such as LAI and land cover, using as classification of crops to form a GIS database,as input variables to drive the model is namely the driving method. To improve simulation precision, more complex assimilation method coupling remote sensing data and crop modelsis used to estimate theregional yield. With the ensuring of the special and temporal resolution, and explanation of crop growth mechanism, the later method is more and more widely used.

In this study, the process-basedremote sensing BEPS model is initially used to simulate the net primary productivity of boreal forest ecosystem in Canada [3, 14].It has been used to estimate the NPP of vegetation, and is suitable for applying in China.Wang et al. (2009, 2011) developed a RS-P-YEC model based on the BEPS model, and to simulate the white wheat yield (C3 crop) [16,17].It is generally accepted that the carbon isotopic composition of plant material is correlated with C3 and C4 pathways of carbon fixation in photosynthesis, the maize is C4 crop according to its photosynthetic pathway. However, in recent year, there was seldom study focusing on C4 crop yield regional simulationintegratingsatellite information and crop models. In this paper, based on the remote-sensing-photosynthesis-yield estimation for crops (RS-P-YEC) and the Boreal Ecosystem Productivity Simulator BEPS), we develop a process model for yield estimation for C_4 crops, and the developed model was used to estimate maize yield in the Northeast China.

2 Study Area and Method

The Northeast China (NEC) (118.83°~135.09°E, 38.72°~54.56°N) covers Heilongjiang, Jilin and Liaoning Provinces. The total arable land of 21,526,200 hectares, accounts for 16.6% of the nation. With continental monsoon climate,the annual rainfall in the NEC is 400~800mm, which is mainly concentrated in July to September. The annual average temperature is -3~11ºC, the accumulated temperature of ≥10ºC is 1600~3600ºC and the frost-free period is 155~257 days.The rainfall and the heat are in the same period, which is suitable for maize growth. The soil is mainly black and chernozem, which is fertile and has a strong ability of water-holding.

2.1 Method

In this study, the BEPS model wasmodified suit for C4 crop yieldestimation;and a hypothesiswith horizontally homogeneous and vertically laminar structure was made, and the structure of two-big-leaf model was still retained, based on the BEPS model. Integrating the photosynthesis of each leaf layer to get the canopy photosynthesis and subtracting the autotrophic respiration to get the NPP of C_3crop, the crop yield was estimated according the ratio of NPP and the yield with the correlation coefficient up to 0.9[16,17].In this study, the developed model calculated the photosynthesis of C_4 crop, based on the RS-P-YEC model, and introduced the harvest index derived from the ratio of the grain yield to the stalk yield for maize.

2.1.1 Photosyntheticrate of Leaf-Level

1) Instantaneous photosyntheticrate (A,$\mu molCm^{-2}s^{-1}$)

The model accumulates the assimilation rateof C_3 plants using the Farquhar model[18]. It could describe the photosynthetic pathway of C_4 plants after modifying [19].

$$A = \min\{w_{v,}w_j\} - R_d \tag{1}$$

$$R_d = 0.015V_{cmax} \tag{2}$$

$$w_v = V_{cmax} \tag{3a}$$

$$\mathrm{w}_j = \mathrm{J} \tag{3b}$$

Where,A is the net CO_2 assimilation rate, w_v is the rate limited by Rubisco, w_j is the rate limited by photoelectron transfer rate, R_d is dark respiration,$\mu molCm^{-2}s^{-1}$;V_{cmax} is the maximum rate of carboxylation, and J is the electron transfer rate.The calculation of V_{cmax} and Jcan be found in Chen *et al.* (1999) [14].

2) Daily integration of photosynthetic rate

Because C_4 photosynthesis is nearly saturated at current CO_2 concentration, the photosynthetic rate is almost free from theinfluence of CO_2concentration [20], and the daily total photosynthesis can be made with respect to time.

$$A_v = \frac{\alpha}{2}(V_{cmax} - R_d) \tag{4a}$$

$$A_j = \frac{\alpha}{2}(\mathrm{J} - R_d) \tag{4b}$$

Where, A_v and A_j correspond toand w_v and w_j, respectively, after a small reduction for dark respiration. The parameter αcan be multiplied to the integral and the calculation can be expressed as [14]:

$$\alpha = \frac{1}{0.5\pi/2}\int_0^{\pi/2} cos\theta \,\mathrm{d}\theta = \frac{4}{\pi} \approx 1.27 \tag{5}$$

Where, θ is the solar zenith angle.

2.1.2 Solar Radiation

The solar radiation received by leaves is a key factor to determine their photosynthetic rate. Calculate the radiation by making leaves divided into N layers. The top of the canopy is the sunlit layer. The solar radiation received by the top sunlit leaves of the canopy includes the direct and scattered radiation from the sky.The radiation received by the lower layers is the sum of multi-scatted and reflect solar radiation from the canopy and soil surface, and the scattered radiation meet the radiation transfer equation [21].

$$S_{sun}(0) = S_{shade}(0) + S_0\mu_0 \tag{6}$$

$$-\mu\frac{dS_{in}}{dL} = -S_{in} + \frac{\omega}{2}\int_{-1}^{1} S_{in}(\mu')d\mu' + \frac{\omega S_0}{2}\exp\left(\frac{LG_0}{\mu_0}\right) \tag{7}$$

Where, $S_{sun}(0)$is the photosynthesis available radiation received by the top leaves of the top of the canopy, $S_{shade}(0)$is the scatted radiation, $S_0$is the direct radiation on theunderlying surface and $\mu_0$is the cosine of the zenithangle. $S_{in}$is the scattered radiation received by the internal of the canopy, $L$is the distance between internal leaves and the top of the canopy, μis the cosine of the zenith in scatter direction, $\omega$is the single scatteralbedo of leaves and $G_0$is the projection on the direction of the reflect radiation.

The scattered radiation in the horizontal ofL could be expressed as,

$$S_{shade}(L) = 2\int_0^1 S_{in}(L,\mu)\mu d\mu \tag{8}$$

Where, $S_{shade}(L)$ is the scattered radiation received by the shade leaves in L.

2.1.3 Photosynthetic Rate of the Canopy

The photosynthetic rate of the sunlit and shaded leaf, and the maizecanopyis divided into N (N>1) layers to make spatial scale expansion. Photosynthetic rate of the canopy is expressed below[22],

$$A_{canopy} = \frac{1}{N}A_{sun}LAI_{sun} + \frac{N-1}{N}A_{shade}LAI_{shade} \tag{9}$$

Where, the subscripts ‘sun’ and ‘shade’ denote the sunlit and shaded components of photosynthesis and LAI.

$$LAI_{sun} = 2cos\theta(1 - \exp(-0.5\Omega LAI/cos\theta)) \tag{10a}$$

$$LAI_{shade} = LAI - LAI_{sun} \tag{10b}$$

Where, Ω is foliage clumping index, and Ω for crop is 0.9[3].

2.1.4 Conversion of NPP to Crop Yields

Gross primary productivity (GPP) could be calculated from photosynthetic rate A, and net primary productivity is equal to GPP minus autotrophic respiration.

$$\mathrm{NPP} = \mathrm{GPP} - R_a \tag{11}$$

$$\mathrm{GPP} = A_{canopy} \times L_{day} \times F_{GPP} \tag{12}$$

Where, $L_{day}$is the length of day, $F_{GPP}$is a conversion factor of photosynthesis to GPP; $R_a$is the autotrophic respiration, including maintenance respiration and growth respiration.

There is a great relationship between NPP and aboveground biomass, so the crop yield could be obtained by the NPP and harvest index(HI).

$$\text{Yield} = \text{NPP} \times \alpha \times \text{HI} \tag{13}$$

Where α is the conversion ratio between carbon content and dry matter (∼45%)[23].

2.1.5 Key Parameters

V_{cmax}, the maximum rate of carboxylation, is one of key parameters in the model, representing the leaf of maize maximum photosynthesis capacity.The value of V_{cmax} is mainly affected by the environmental factors such as temperature and fertilizer. J,the electron transfer rate, has a strong correlation with the V_{cmax}.

2.1.6 Model Running

Due to the difference of geographical latitudes, the sowing dates of maize in the three administrative are April 25th, May 1th and May 7th, respectively. The latest harvest date in Heilongjiang Province is October 8th. In order to make more use of remote sensing data and simplify the simulation process, April 23th (22th in the leap year) was uniformly set as the sowing date, and October 8th (7th in the leap year) was set as the mature date, with the reason that the farmland is almost bare soil before sowing and after harvesting and the NPP can be ignored.

2.2 Data

Running model, the meteorological data, remote sensing data, land use and soil data were used.Daily meteorological data were collected at 72 meteorological stations in Northeast China [24], including maximum, minimum, relative humidity and rainfall which were fromtheNational Meteorological Information Center in China. The MODIS (Moderate-resolution Imaging Spectroradiometer)derivedLAI data was MOD15A2 product, with the temporal resolution of 8-day and the spatial resolution of 1km [25]. The land use data used in this study isthe IGBP global land cover classification data produced by the United StatesGeological Survey, with a temporal resolution of 1-year and the spatial resolution of 500m. It was resampled to the spatial resolution of 1km using the method of nearest neighbor method (see Figure 1).

The ratio of grain to stalk yield and harvest index(HI) is defined as the ratio of crop grain yield to aboveground biomass. The HI is also one of important variables, which affect the crop yield.The data was derived from the ratio of grain to stalkyield[26],which was derived from the National Meteorological Information Center in China. The HI at stations was obtained and then interpolated to the space with the method of IDW, resampled to 1km.

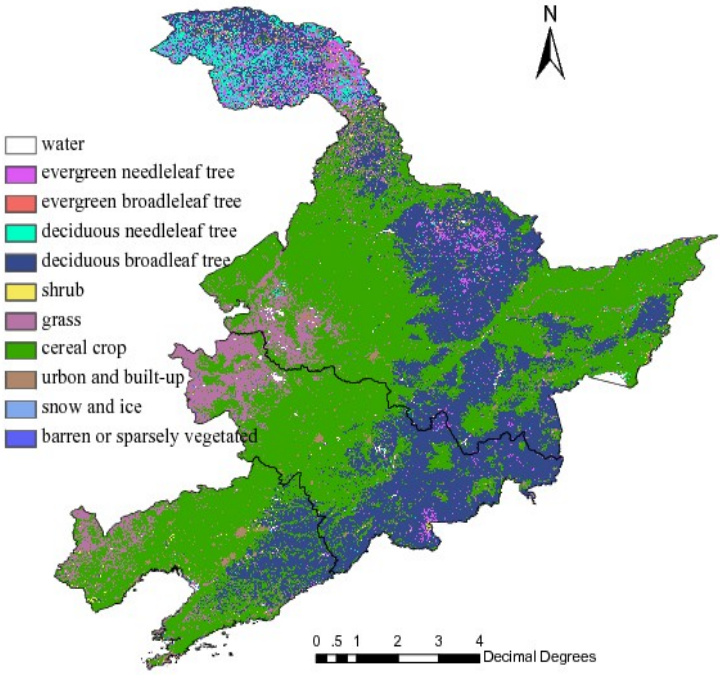


Fig. 1. Classification of land cover in Northeast China

$$HI = y/(y+1) \tag{14}$$

Where, y is the ratio of grain to stalk yield.

Soil available water capacity (AWC) varies with soil texture. The soil texture of the main growing areas of maizein Northeast China is of little difference and the AWC data is difficult to obtain, so it could be made a constant 0.17[27].

3 Result and Discussion

3.1 Validation of Simulated Maize Yield

We validated the simulated maize yieldat a county-level, thestatisticalmaize yield data arecollected in the provincialStatistical Yearbook.The regression analysis between simulated yield and statistics yield in county-level was performed, and the Pearson correlation coefficient R^2 is equal to 0.637(Figure 2) ($p<0.01$). The statistics yield of 69 counties/cities were used to calculate the Absolute deviation(ABS), root mean square error(RMSE) and relative error(RE) .The results showed that, ABS is 829 $kgha^{-1}$, RMSE is 1060 $kgha^{-1}$, and RE is 11.9%. Figure 2 shows that most of the data fell in the confidence range of 75%.The statistics yield of Shangzhi in 2007 is an unusually high value and maybe is a bias from the statistical method.

3.2 The Spatial Distribution Pattern of Maize Yield in the NEC

The yield of maize in the Northeast China Plain in 2008was estimated using the developed process-based model (Figure 2). The simulated results showed that maize yield was high in north and low in west parts.Sanjiang plain in Heilongjiang province, Songnen plain, the middle of Jilin province and the central plain of Liaoning province are the main producing area of maize with fertilesoil, and the maizeyields were relativelyhigher with values between7000-9000 kgha^{-1},some of the areas were even up to 10000 kgha^{-1}.The southeast of Heilongjiang province, the east of Jilin province and the southeast of Liaoning province aremountainous areas withless cultivated area, and maizeyields varied between 5000-6000 kgha^{-1}. Because of less rainfall, the yield of west of Jilin province and northwest of Liaoning Province was the lowest and less than 3000kgha^{-1}.On the whole, the maizeyields in the Northeast China Plain mainly ranged in 5000-9000kgha^{-1} according to the results.

3.3 Discussion

A key parameter of model, LAI, could well reflect the crop growth status. The precision of the model depends heavily on the accuracy of LAI. LAI used to drive the model was derived from MODIS products. The resolution of 1 km might affect the accuracy of the prediction of the NPP and the yield. Some studies have suggested that MODIS LAI is lower than measured values, and this could lead to low simulated value.These errors may be decreased through using higher precision LAI.

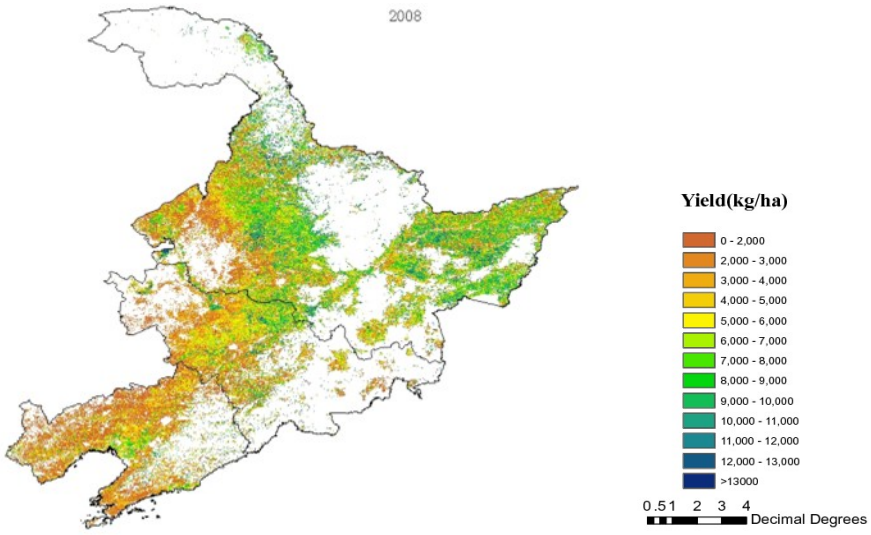


Fig. 2. The simulation results spatial distribution of maize yield in 2008in Northeast China

The HI was critical for accurate prediction of crop yield by the model. Taking into account the small change between years, The HI was derived from the ratio of grain to stalk yield, assuming that changes from crop varieties and cultivars did not occur. In addition, the influence of environmental condition and field management did not take into consideration. The rough assumption must affect the accuracy of prediction of the crop yield. The accurate estimation of The HI is an effective way of improving the prediction of yield. There is a certain uncertainty in statistics yield. A more accurate yield data to the validation of results can also reduce the error.

4 Conclusion

We developed the process-based model for estimating the C4 crop yield based on the RS-P-YEC model in this paper, and the translation of NPP was realized to yield using the relationship of NPP to biomass by introducing the harvest index. The results indicated that there was a good relationship between yield model-simulated and statistical yield at the county level(R^2=0.637, n=69), and the developed process-based model was suitable for the maize yield estimation in a large scale.

Acknowledgements. This study is supported the Social Commonweal Meteorological Research Project (Grant No. GYHY201106027), the Global Change Global Research Key Project of the National Science Plan (Grant No. 2010CB951302), the Fund of Agricultural Science &Technology Achievements (No.2011GB249100), and One Hundred Person Project of the Chinese Academy of Sciences.

References

1. Moulin, S., Bondeau, A., Delecolle, R.: Combining agricultural crop models and satellite observations: From field to regional scales. Int. J. Remote Sens. 19, 1021–1036 (1998)
2. Basso, B., Ritchie, J.T., Pierce, F.J., Braga, R.P., Jones, J.W.: Spatial validation of crop models for precision agriculture. Agr. Syst. 68, 97–112 (2001)
3. Liu, J., Chen, J.M., Cihlar, J.: A process-based boreal ecosystem productivity simulator using remote sensing inputs. Remote Sensing of Environ. 62(2), 158–175 (1997)
4. Fraisse, C.W., Sudduth, K.A., Kitchen, N.R.: Calibration of the CERES-Maize model for simulating site-specific crop development and yield on claypan soils. Appl. Engin. Agricul. 17(4), 547–556 (2001)
5. Lobell, D.B., Asner, G.P., Ortiz-Monasterio, J.I.: Remote sensing of regional crop production in the Yaqui Valley, Mexico: estimates and uncertainties. Agricul. Ecosys. Environ. 94(2), 205–220 (2003)
6. Reynolds, C.A., Yitayew, M., Slack, D.C.: Estimating crop yields and production by integrating the FAO Crop Specific Water Balance model with real-time satellite data and ground-based ancillary data. Int. J. Remote Sens. 21(18), 3487–3508 (2000)
7. Shanahan, J.F., Schepers, J.S., Francis, D.D.: Use of remote-sensing imagery to estimate corn grain yield. Agron. Journ. 93(3), 583–589 (2001)
8. Baez-Gonzalez, A.D., Kiniry, J.R., Maas, S.J.: Large-area maize yield forecasting using leaf area index based yield model. Agron. Journ. 97(2), 418–425 (2005)

9. Prasad, A.K., Chai, L., Singh, R.P.: Crop yield estimation model for Iowa using remote sensing and surface parameters. Inter. J. App. Ear. Obser. Geoin. 8(1), 26–33 (2006)
10. Bastiaanssen, W.G.M., Ali, S.: A new crop yield forecasting model based on satellite measurements applied across the Indus Basin, Pakistan. Agricul. Ecosys. Environ. 94(3), 321–340 (2003)
11. Liu, J., Pattey, E., Miller, J.R.: Estimating crop stresses, aboveground dry biomass and yield of corn using multi-temporal optical data combined with a radiation use efficiency model. Remot. Sens. Envir. 114(6), 1167–1177 (2010)
12. Turner, D.P., Gower, S.T., Cohen, W.B.: Effects of spatial variability in light use efficiency on satellite-based NPP monitoring. Remot. Sens. Envir. 80(3), 397–405 (2002)
13. Gower, S.T., Kucharik, C.J., Norman, J.M.: Direct and Indirect Estimation of Leaf Area Index, fAPAR, and Net Primary Production of Terrestrial Ecosystems. Remot. Sens. Envir. 70(1), 29–51 (1999)
14. Chen, J., Liu, J., Cihlar, J.: Daily canopy photosynthesis model through temporal and spatial scaling for remote sensing applications. Ecol. Model. 124(2), 99–119 (1999)
15. Ma, Y.P., Wang, S.L., Zhang, L.: Monitoring winter wheat growth in North China by combining a crop model and remote sensing data. Inter. J. App. Ear. Obser. Geoin. 10(4), 426–437 (2008)
16. Wang, P.J., Xie, D.H., Zhang, J.H.: Application of BEPS model in estimating winter wheat yield in North China Plain. Tran. CSAE 25(10), 148–153 (2009)
17. Wang, P.J., Sun, R., Zhang, J.H.: Yield estimation of winter wheat in the North China Plain using the remote-sensing–photosynthesis–yield estimation for crops (RS–P–YEC) model. Int. J. Remote Sens. 32(21), 6335–6348 (2011)
18. Farquhar, G.D., Caemmerer, S., Berry, J.A.: A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species. Planta 149(1), 78–90 (1980)
19. Zhang, F.M., Ju, W.M., Chen, J.M.: Characteristics of terrestrial ecosystem primary productivity in East Asia based on remote sensing and process-based model. Chin. J. Appl. Ecol. 23(2), 307–318 (2012)
20. Leakey, A.D.B., Uribelarrea, M., Ainsworth, E.A.: Photosynthesis, productivity, and yield of maize are not affected by open-air elevation of CO2 concentration in the absence of drought. Plant Phy. 140(2), 779–790 (2006)
21. Huang, H.F.: The study on Theory and Simulation of the Interactions of Soil, Vegetation and Atmosphere. Metereological Press, Beijing (1997)
22. Norman, J.M.: Simulation of microclimates. In: Hatfield, J.L., Thomason, I.J. (eds.) Biometeorology in Integrated Pest Management, pp. 65–99. Academic Press, New York (1982)
23. Schlesinger, W.H.: Biogeochemistry: an analysis of global change. Academic press (1997)
24. Chinease Climate Data Information, `http://cdc.cma.gov.cn`
25. MODIS Data Information, `http://reverb.echo.nasa.gov/reverb/`
26. Zhang, F.C., Zhu, Z.H.: Harvest index for various crops in China. Sci. Agricul. Sin. 23(2), 83–87 (1990)
27. Feng, X., Liu, G., Chen, J.M.: Net primary productivity of China's terrestrial ecosystems from a process model driven by remote sensing. J. Environ. Manag. 85(3), 563–573 (2007)

Runoff Response of Zamu River Basin to IPCC Climate Change Scenarios in Northwest China

Sufen Wang and Xin Liu

Center for Agricultural Water Research in China, China Agricultural University, Beijing, 100083, China
wangsuf@cau.edu.cn, wwwsf71@163.com

Abstract. Predicting runoff response to climate change is useful in making the decision of water resources management in arid region. This study investigated the impact of climate change on the runoff of Zamu River, one of the inland rivers in the arid region of northwest China using Soil and Water Assessment Tool (SWAT) model. Climate-change was predicted by the UK Hadley Centre's Climate Model (HadCM3) under IPCC A2 and B2 scenarios, and downscaled by statistical downscaling model (SDSM) for two periods: 1961–1990 (control) and 2010–2099 (scenario) to drive the SWAT model. SDSM predicted an increase trend of maximum and minimum temperature and precipitation in the study area during the period of 2010–2099. Simulated runoff under IPCC SRES A2 and B2 scenarios changed by -10.6% - +1.17% and - 4% - +13%, respectively. The runoff tended to decline more significantly under SRES A2 (high GGa emissions) than under SRES B2 (low GGa emissions) in the future. The linear trend values were -0.048 and -0.018, respectively.

Keywords: Climate change; Runoff; Downscaling; SWAT; SDSM.

1 Introduction

There are scientific evidences about hydrologic system affected by the global climate change. The global temperature is increasing and the 100-year trend (1906-2005) of 0.74 [0.56 to 0.92] °C is larger than the corresponding trend of 0.6 [0.4 to 0.8] °C (1901-2000) given in the TAR (IPCC, 2007) [1]. The global climate change has impacts on regional precipitation, precipitation distribution and runoff [2],[3],[4],[5],[6]. Global warming results in evaporation increase. Many studies have proved that runoff is very sensitive to climate change [7],[8],[9],[10]. Runoff conditions are strongly controlled by climate [11]. Climate change could therefore have positive or negative impacts on runoff [12].Hydrological model sensitivity to climate change can be defined as the response of a particular hydrological model to a known quantum of climate change [13],[23],[24],[25].

One way to assess possible impact of climate change on hydrological cycle is to apply different climate change scenarios to hydrological models to estimate hydrologic cycle factors [14],[15]. General climate models (GCMs) and regional climate models (RCMs) are frequently used to model future climate scenarios. The building of hydrological model and the generation of future climate change scenarios are essential

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 223–231, 2013.

to water cycle assessment. The output of climatic factors by GCMs was used to drive hydrological models to study hydrological response to global climate change [16], [17],[18]. Many studies have investigated the impact of climate change on annual mean water flow under IPCC A2 and B2 GHG scenarios [19],[22]. The simulating scale of climate models has great difference to that of hydrological models. However, the general circulation models and regional climate models are among the most advanced tools in estimating future climate change scenarios. Therefore the output from GCMs and RCMs has to be downscaled to obtain the information relevant to hydrologic studies. Downscaling approach was widely used for constructing climate scenarios to drive the hydrologic models. The approaches to downscale the outputs of GCMs are as the follows: Dynamic downscaling method, statistical downscaling method and interpolation method. Dynamic downscaling method, as a Regional Climate Model, is embedded into GCM, but the method had complicated design and application condition, so it is not widely selected for downscaling. Compared to the dynamic downscaling method, the statistical downscaling method is most widely used to downscale the climate scenarios because of less demanding application condition. Statistical downscaling method is to derive empirical relationships that transform large scale features of the GCM (Predictors) to small scale variables (Predictants) based on the basic data. Precipitation and temperature can be predicted. There are three implicit assumptions involved in statistical downscaling method [21]. Interpolation methods include bilinear interpolation and non-equidistant Lagrange three-point interpolation method, through which the output of GCMs can be interpolated to appropriate site.

Arid regions frequently suffer from years of acute shortages of water resources. Climate is a key factor affecting the runoff formation of inland river basin of the arid area in the northwest China. The objective of this study was to evaluate the runoff changes under different climate change scenarios in the Zamu river basin of northwest China. The Statistical downscaling approach (Wilby et al., 2001)[21] and Soil and Water Assessment Tool (SWAT) (Arnold et al., 1998)[20] distributed hydrological model were chosen for this study.

2 Materials and Methods

2.1 Study Watershed

The Zamu River originates in the Qilian Mountains and has a catchment area of 851 km^2. It is the only unregulated river in the Shiyang river basin in the arid region of northwest China and has a glacier area of 3.74 km^2 in the mountain area in the upper reach. The location of the study area was shown in Fig. 1. The only gauging station in the catchment is Zamusi hydrologic station. The elevation of the catchment varies from 2000 m to 4802 m above sea level and its catchment shape is plumose. Mean flow of the river (1955-2005) measured by the Zamusi hydrologic station is 7.75 m^3/s. The upper Zamu river catchment has good vegetation cover, with alp meadow, alp grassland, shrub and arbor. Forestland is patchy, with mixed distribution of grassland and forestland. Major land use type can be divided into grassland, forestland, farmland, rural resident land and uncultivated land [26].

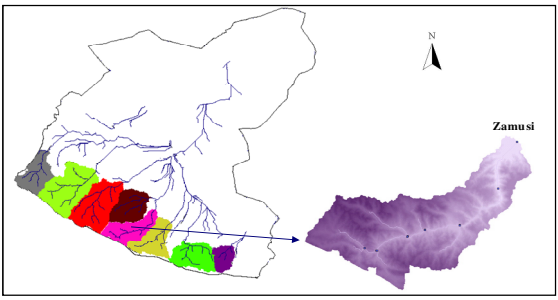


Fig. 1. The map of location o f the study area

2.2 SWAT Hydrological Model

SWAT (Soil and Water Assessment Tool) (Arnold et al., 1998)[20] is physically based hydrological model. Hydrologic Response Units (HRUs) is the basic calculating units, which is consisting of unique combinations of land cover and soils in each sub-basin. SWAT allows a number of different physical processes to be simulated in a basin. It can be used to simulate the hydrological response to changed environment in different time steps (http://www.brc.tamus.edu/swat/swatmanual, swat2000theory). The hydrologic cycle as simulated by SWAT is based on the water balance equation:

$$SW_t = SW_0 + \sum_{i=1}^{t} (R_{day} - Q_{surf} - E_\alpha - w_{seep} - Q_{gw}) \quad (1)$$

where SW_t is the final soil water content (mm H_2O), SW_0 is the initial soil water content on day i (mm H_2O), t is the time (days), R_{day} is the amount of precipitation on day i (mm H_2O), Q_{surf} is the amount of surface runoff on day *i* (mm H_2O), E_a is the amount of evapotranspiration on day i (mm H_2O), w_{seep} is the amount of water entering the vadose zone from the soil profile on day i (mm H_2O), and Q_{gw} is the amount of return flow on day i (mm H_2O).

2.3 Statistical Downscaling Model

Statistical downscaling model (SDSM) was used to calculate statistical relationships between large-scale (the predictors) and local climate variables (Predictants) based on multiple linear regression technique. These relationships are developed using the observed weather data, assuming that these relationships remain valid in the future. They can be used to obtain downscaled local information for some future time period by driving the relationships with predictors simulated by GCMs [21]. There are following key steps: 1) verifying of observing materials; 2confirming predictors; 3) model calibration and verification and 4) driving future climate change scenarios. The established model for predicting daily maximum and minimum temperatures is an unconditional model, but the model for predicting precipitation is a conditional model.

2.4 Climate Change Scenario

Scenarios of climate change used in this study were IPCC SRES A2 and B2, which were projected by the UK Hadley Centre's HadCM3 model under corresponding emissions scenarios. The predictor variables can be obtained online (http://www.grida.no/climate/ipcc/emission). A2 describes a very heterogeneous world with high population growth, slow economic development and technological change. B2 describes a world with intermediate population and economic growth, emphasizing local solutions to economic, social, and environmental sustainability (IPCC, 2007). Climate scenarios A2 and B2 are close to the development of study area. The output of HadCM3 was downscaled to the daily series data of corresponding weather station.

2.5 SDSM Calibration and Validation

Weather factors under different emission scenarios will be obtained based on the NCEP (National Centre for Environmental Prediction) data from 1961 to 1990. Daily rainfall and maximum and minimum temperatures were analyzed in this study.

Correlation coefficient (R), relative error (RE), Nash–Sutcliffe efficiency (NSE) are the criterion to evaluate the model performance . The Nash–Sutcliffe efficiency was calculated as follows:

$$NSE = 1 - \sum_{i=1}^{i=n} (V_{obs_m} - V_{dow_m})^2 / \sum_{i=1}^{i=n} (V_{obs_m} - \overline{V_{obs}})^2 \quad (2)$$

where $Vobs_m$ is observed value, $Vdow_m$ is downscaled value, $\overline{V_{obs}}$ is mean observed value, and n is the number of measurement. NSE value can range from -∞ to 1. 1 corresponds to perfect match of downscaled value to the observed data.

The representation meteorological stations selected are Tianzhu and Wuwei in Zamu river basin. The positions and averages of the temperature and precipitation were shown in Table 1. Tianzhu is the mountain observing station whereas Wuwei station is the plain observing station. Selected predictors for established downscaling model are shown in Table 2. The positive correlation coefficients of the variables are the selected predictors for establishing the downscaling model. Data from 1961-1975 was used for calibration and data from 1976-1990 was used for validation. Table 2 shows that predictor variables of P500 (500 hPa geopotential height) and tem (the average temperature of the ground 2 meters) are important in predicting the climate variables.

Table 1. Statistics of two representative meteorological stations

Station	Longitude (°N)	Latitude (°E)	Elevation (m)	Mean precipitation (mm)	Mean temperature (°C)	data
Tianzhu	102°52′	37°12′	3045	411.1	0.05	1951-2005
Wuwei	102°40′	37°55′	1531	167.2	8	1951-2005

Table 2. Predictor variables selected for downscaling

Predictor	Predictant					
	T_{max}(℃)		T_{min}(℃)		Prec(mm)	
	Tianzhu	Wuwei	Tianzhu	Wuwei	Tianzhu	Wuwei
p-u(Surface zonal velocity)	0.13	0.26	0.09	0.02	-0.01	-0.06
p-v(Surface meridional velocity)	-0.18	-0.18	-0.26	-0.18	-0.09	-0.13
p-z (Surface vorticity)	0.18	-0.28	-0.07	-0.26	-0.12	-0.03
p500(500 hPa geopotential height)	0.27	0.20	0.43	0.23	0.03	0.05
r500(Relative humidity at 500 hPa)	-0.30	-0.19	-0.07	-0.01	0.08	0.07
Shum(Surface specific humidity)	-0.08	-0.03	-0.07	-0.01	0.06	0.11
tem (Mean temperature at 2m)	0.59	0.69	0.56	0.53	0.04	0.03
rhum (Near surface relative humidity)	0.01	0.02	0.06	0.07	0.03	0.13

Table 3. Statistics of SDSM validation (1976-1990)

Item	T_{max}(℃)		T_{min}(℃)		maximum wet-spell length(days)		Mean wet-day precipitaion(mm)	
	Tianzhu	Wuwei	Tianzhu	Wuwei	Tianzhu	Wuwei	Tianzhu	Wuwei
Observed	5.59	15.17	-4.72	1.28	15.40	8.00	3.21	2.65
Downscaled	5.40	14.69	-4.70	1.31	15.20	7.65	2.89	3.38
R^2	0.98	0.99	0.99	0.99	0.96	0.76	0.96	0.96
RE(%)	-3.40	-3.16	-0.42	2.34	-1.30	-4.38	-9.97	27.55
NSE	0.99	0.997	0.99	0.996	0.96	0.965	0.99	0.837

Table 3 shows the results of observed and downscaled daily maximum temperature, minimum temperature and precipitation for the validation period in Tianzhu station and Wuwei station. Both correlation coefficient and Nash-Suttclife efficiency coefficient (NSE) are more than 0.75, especially the result of downscaled temperature was better than that of daily precipitation. So daily maximum and minimum temperatures under climate change scenarios A2 and B2 from 2010 to 2099 were generated by SDSM based on the data from 1961 to 1990. The results show that maximum temperature (T_{max}) and minimum temperature (T_{min}) have an increasing tendency in the future under different IPCC scenarios in the Zamu river basin (Fig.2). The precipitation have an increasing tendency begin 2010s under SRESA2 and SRESB2 (Fig.3).

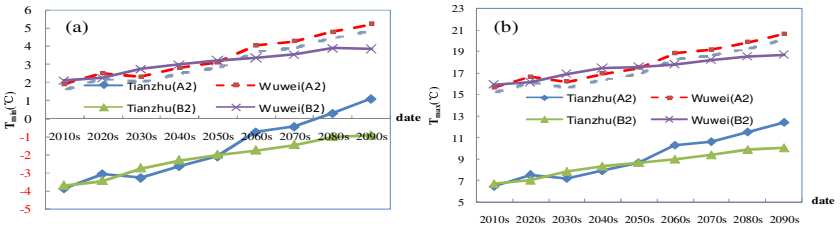


Fig. 2. Average Tmin (a),Tmax (b),under SRES A2 and B2

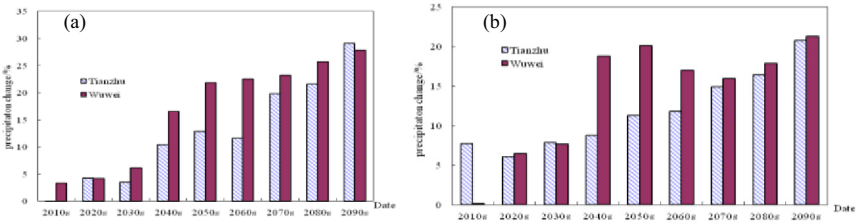


Fig. 3. Precipitation under SRES A2(a) and B2(b) relative 1980s

2.6 SWAT Model Validation

SWAT model can be used to simulate the runoff under the observed climate data in Zamu river basin [26] .Downscaled climate data based the NCEP from 1985 to 1990 was used to driven the SWAT hydrological model for validation in this study. Fig. 4 shows monthly simulated runoff with downscaled climate data well matched the observed value. Correlation coefficient (R) of validation of SWAT model with downscaled climate data is 0.79. SWAT model can be used to simulate runoff driven by downscaled data. Peak observed and simulated runoff under downscaled climate data is greater than those under observed weather data (Fig.4). The difference is caused by different spatial scale between observed data and downscaled data.

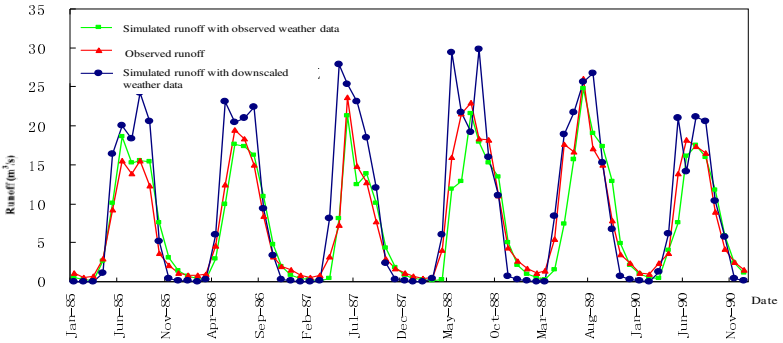


Fig. 4. Validation of SWAT model using downscaled data (1985-1990)

3 Results and Discussion

Tables 4 and 5 show the comparison of the baseline runoff and projected values for different scenarios corresponding to the downscaled precipitation and temperature under SERS A2 and B2, respectively. The projected runoff shows the decreasing tendency under SERS A2 climate scenario in future except that in the early 21st century. The runoff decreased with the increasing of precipitation in 2070s, 2080s and 2090s. Runoff was influenced by precipitation and temperature. The temperature has negative effects on runoff. The runoff was reduced by less than 10% in 2080s, 5% in

2050s and 10.6% in 2060s under SRES A2 respectively. From Table 5, the runoff was reduced by less than 5% in 2020s, 2040s and 2080s under SERS B2 scenario. The runoff generally was changed by less than 10% in the future decades along with the change of temperature and precipitation in Heihe river basin [22], which is similar to the results of this study.

Table 4. Change of Simulated Runoff, Precipitation, T_{max} and T_{min} under SRES A2 relative to baseline(1980~1989)

Years		2010s	2020s	2030s	2040s	2050s	2060s	2070s	2080s	2090s
Runoff change(%)		1.2	0.8	-5.4	-8.9	-2.2	-10.6	-2.4	-8.9	-3.4
Precipitation Change(%)	Tianzhu	0.2	4.3	3.5	10.4	12.9	11.6	19.8	21.6	29.1
	Wuwei	3.3	4.1	6.1	16.5	21.8	22.5	23.2	25.6	27.8
T_{max} Change(℃)	Tianzhu	1.2	2.2	1.9	2.6	3.4	5	5.3	6.2	7.1
	Wuwei	1.1	2.1	1.7	2.4	2.9	4.3	4.6	5.3	6.1
T_{min} Change(℃)	Tianzhu	0.9	1.7	1.5	2.1	2.7	4	4.3	5	5.8
	Wuwei	0.6	1.2	1	1.5	1.8	2.7	2.9	3.5	3.9

Table 5. Change of Simulated Runoff, Precipitation, T_{max} and T_{min} under SRES B2 relative to baseline(1980~1989)

Years		2010s	2020s	2030s	2040s	2050s	2060s	2070s	2080s	2090s
Runoff change(%)		12.9	-1.8	1.2	-3.9	8.3	1.3	2.4	-2	6.8
Precipitation Change(%)	Tianzhu	7.7	6.1	7.8	8.8	11.2	11.8	14.8	16.4	20.7
	Wuwei	0.1	6.5	7.6	18.8	20.1	17	15.9	17.8	21.2
T_{max} Change(℃)	Tianzhu	1.4	1.7	2.5	3	3.4	3.7	4.1	4.6	4.7
	Wuwei	1.4	1.6	2.3	2.9	3	3.2	3.6	4	4.1
T_{min} Change(℃)	Tianzhu	1.1	1.3	2	2.5	2.8	3	3.3	3.7	3.8
	Wuwei	0.8	0.9	1.4	1.7	1.9	2	2.2	2.5	2.5

The runoff in the mountain reaches varied from -10.6% to +1.17% under SRES A2 and -4% to +13% under SRES B2 in the future. The runoff had a decline tendency in the future (Fig. 5), The linear trend values under SRES A2 and B2 are -0.048 and -0.018 separately.The runoff declined more obviously under SRES A2 than under SRES B2.

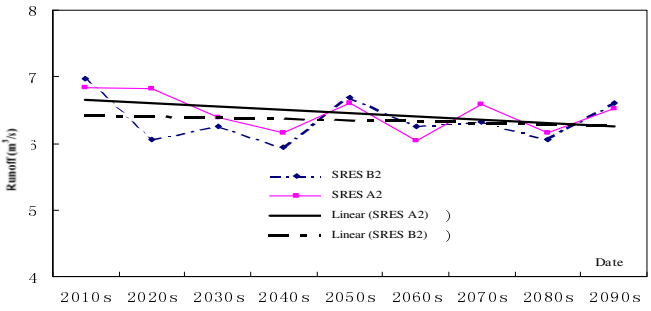


Fig. 5. Simulated runoff under SRES A2 and B2 in future

4 Conclusions

The response of the runoff in the mountain reaches to the climate change under IPCC SRES were simulated by combining the climate model and distributed hydrological model. According to the predictions made by the HADCM3 model, which is downscaled by SDSM, future climates predict increased warming under two different climate change scenarios.

With climate change, runoff change in Zamu river basin appeared. The simulated runoff of the mountain reaches under different climate change scenarios shows that runoff under high GGa emissions has a more obvious decline tendency than that under low GGa emissions in the future. Predictions regarding runoff response to climate change in this paper can give some advices for water resources management and ecological environment decisions in arid regions.

Acknowledgements. This research was supported by Chinese National Nature Science Fund (51109211).

References

1. Intergovernmental Panel on Climate Change (IPCC), Climate Change 2007: the Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA (2007)
2. Sefton, C.E.M., Boorman, D.B.: A regional investigation of climate change impacts on UK steamflows. Journal of Hydrology 195, 26–44 (1997)
3. Guo, S.L., Wang, J.X., Xiong, L.H., Ying, A.W., Li, D.F.: A macro-scale and semi-distributed monthly water balance model to predict climate change impacts in China. Journal of Hydrology 268, 1–15 (2002)
4. Chen, M., Pollard, D., Barron, E.J.: Hydrologic processes in China and their association with summer precipitation anomalies. Journal of Hydrology 301, 14–28 (2005)
5. Hans, T.: The influence of climate change on stream flow in Danish rivers. Journal of Hydrology 333, 226–238 (2007)
6. Gardner, L.R.: Assessing the effect of climate change on mean annual runoff. Journal of Hydrology 379, 351–359 (2009)
7. Gleick: Methods for evaluating the regional hydrologic impacts of global climatic changes. Journal of Hydrology 88(1), 97–116 (1986)
8. Jones, R.N., Chie, F.H.S., Boughton, W.C., Zhang, L.: Estimating the sensitivity of mean annual runoff to climate change using selected hydrological models. Advances in Water Resources 29, 1419–1429 (2006)
9. Vaze, J., Post, D.A., Chiew, F.H.S., Perraud, J.M., Viney, N.R., Teng, J.: Climate non-stationarity – Validity of calibrated rainfall–runoff models for use in climate change studies. Journal of Hydrology 394, 447–457 (2010)
10. Zarghami, M., Abdi, A., Babaeian, I., Hassan, Y., Kana, R.: Impacts of climate change on runoffs in East Azerbaijan, Iran. Global and Planetary Change 78, 137–146 (2011)
11. Kling, H., Fuchs, M., Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. Journal of Hydrology (424-425), 264–277 (2012)

12. Nunes, J.P., Seixas, J., Keizer, J.J.: Modeling the response of within-storm runoff and erosion dynamics to climate change in two Mediterranean watersheds: A multi-model, multi-scale approach to scenario design and analysis. Catena 102, 27–39 (2013)
13. Viner, D., Hulme, M., Raper, S.C.B.: Climate change scenarios for the assessments of the climate change on regional ecosystems. J. Therm. Biol. 20(1/2), 175–190 (1995)
14. Booij, M.J.: Impact of climate change on river flooding assessed with different spatial model resolutions. Journal of Hydrology 303, 176–198 (2005)
15. Nigel, W.A.: Climate change and global water resources. Global Environmental Change (9), 31–49 (1999)
16. Nigel, W.A.: Relative effects of multi-decadal climatic variability and changes in the mean and variability of climate due to global warming: future streamflows in Britain. Journal of Hydrology 270, 195–213 (2003)
17. Sperna, W.F.C., Van, L.P.H., Weerts, A.H., Bierken, M.F.P.: Extracting information from an ensemble of GCMs to reliably assess future global runoff chang. Journal of Hydrology 412-413, 66–75 (2012)
18. Wilby, R.L., Dawson, C.W., Barrow, E.M.: SDSM – a decision support tool for the assessment of regional climate change impacts. Environmental and Modelling Software 17, 145–157 (2001)
19. Shivam, T., Sriniva, V.V., Nanjundiah, R.S.: Downscaling of precipitation for climate change scenarios: A support vector machine approach. Journal of Hydrology 330, 621–640 (2006)
20. Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R.: Large area hydrologic modeling and assessment part I: model development. American Water Resources Association 34(1), 73–89 (1998)
21. Arnell, N.W.: Factors controlling the effects of climate change on river flow regimes in a humid temperate environment. Journal of Hydrology 132, 321–342 (1992)
22. Lan, Y.C., Ding, Y.J., Liu, J.Q., Kang, E.S., Wei, Z.: Change of Water Resources in Mountainous Area of Heihe River under Global Warming Scene. Journal of Desert Research 25, 863–868 (2005)
23. Andersson, L.W.J., Martin, C.T.: Impact of climate change and development scenarios on flow patterns in the Okavango River. Journal of Hydrology 331, 43–57 (2006)
24. Chiew, F.S., Teng, J., Vaze, J., Post, D.A.: Influence of global climate model selection on runoff impact assessment. Journal of Hydrology 379, 172–180 (2009)
25. Julie, W., Martin, C.T., Denis, A.H., Anton, E., Dominic, K., Russel, L., Hubert, H.G.: Impact of climate change and development scenarios on flow patterns in the Okavango River. Journal of Hydrology 331, 43–57 (2006)
26. Wang, S.F., Kang, S.Z., Zhang, L., Li, F.S.: Modeling the hydrological response to different land-use and climate change scenarios in the Zamu River Basin of Northwest China. Hydrological Processes 22, 2502–2510 (2008)

Utilizing Cloud-Computation to Analyze the Causative Factors of Rainfall-Induced Landslide

Junyi Huang[1] and Qiming Zhou[1,2,*]

[1] Department of Geography and Centre for Geo-Computation Studies,
Hong Kong Baptist University, Hong Kong, China
jrhuang@life.hkbu.edu.hk
[2] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
qiming@hkbu.edu.hk

Abstract. Among the various natural hazards, landslides are among the most widespread and damaging one. It can be triggered by various external stimuli and pose significant threat to human safety and natural environment. Geospatial computingis currentlyfacing a daunting challenge in data management and processing with ever-increasing complexity and heterogeneity .Different approaches have been developed to produce landslide susceptibility maps. This paper reports a pilot study of analyzing the causative factors of landslide and proposes to utilize the geospatial cloud-computing method in mapping the rainfall-induced landslide susceptibility. A series of geospatial data of triggering stimulus are acquired. The cloud computation platform is utilized to analyze some selected environmental parameters in Lantau Island, Hong Kong. The emergence of cloud computing brings potential solutions to solve the geospatial intensity problems and will enable the public to better prepare for such deadly events and to help mitigate any potential damages.

Keywords: cloud-computation, landslide, Hong Kong.

1 Introduction

1.1 Landslide Susceptibility Mapping

Landslide, defined as the mass movement of rock, debris or earth down a slope results in a geomorphic makeover of the Earth's surface [1], can be triggered by various external stimuli. In recent years, landslide hazard assessment has played an important role in developing land utilization regulations aimed at minimizing the loss of lives and damage to property [2]. In spite of some of the highest urban population densities in the world, 40% of the land area of the Hong Kong Special Administrative Region

* This study was supported by International Science and Technology Collaboration Project of China (2010DFA92720-24), National Basic Research Program of China (Project No. 2012CB719901), Hong Kong Baptist University Faculty Research Grant (FRG1/12-13/070) and The State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS) Open Research Fund.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 232–239, 2013.

(HKSAR) is designated as country parks. Natural terrain covers about 60% of the total land area of Hong Kong.The seeming incongruity is explained by the fact that the mountainous terrain is subject to landslide hazard, with urban development confined to 10% of the land area: a narrow coastal strip backed by steep convex slopes [3].

Landslide susceptibility: A quantitative or qualitative assessment of the classification, volume (or area), and spatial distribution of landslides which exist or potentially may occur in an area [4]. To obtain the most reasonable interpretations, it is therefore important to establish a systematic usage of spatial data and methodologies that quantify and efficiently integrate the spatial relationships. The use of remote sensing data (satellite and air-borne imagery) is well suited to landslide-related studies such as detection, identification and monitoring [5]. Remote-sensing and spatial analysis tools are widely used in landslide studies including landslide detection, landslide assessment, natural hazard, landslide mapping, and landslide inventories (e.g., [6], [7], [8]). Through statistically based prediction models, future landslide-prone areas could be identified through susceptibility and hazard mapping and therefore proper planning for the landscape could be designed by using those maps [9], and thus reduce landslide damages through proper preparation and/or mitigation. The quantitative methodologies are relatively objectives and can be grouped into three categories: statistical analysis, geotechnical engineering approaches and artificial intelligence methods [10].

1.2 Geospatial Cloud-Computation

Geospatial computing, like many other disciplines of science and engineering, is currently facing a daunting challenge in data management that compels researchers and engineers to address the need for handling voluminous geospatial data with ever-increasing complexity and heterogeneity [11]. Toward the resolution of these problems, data-intensive computing has focused on the development of scalable software and hardware solutions for the effective and efficient storage, manipulation, analysis, and distribution of massive amounts of complex data [12].The technology of cloud computing is ushering great change and the potential in the world of Geographic Information System (GIS).

As the geospatial sciences face grand information technology (IT) challenges in the twenty-first century: data intensity, computing intensity, concurrent access intensity and spatiotemporal intensity [13], when comparing Cloud GIS applications to the traditional desktop environments, it becomes quite apparently how useful it could be. Over a few decades efforts are being made to upgrade the conventional GIS applications in order to provide broad spectrum services to the users across the globe. ‘Cloud computing’, a term which has become popular in recent years, has been described as “the next natural step in the evolution of on-demand information technology services and products” [14]. Cloud Computing can be applied to solve and overcome the challenges in GIS applications [15]. It enables people to run modeling applications on a virtual server or thousands of servers. With the development of cloud computing technology, the public can use computing power just like the way they use water, electricity, gas and telephone [16].

1.3 Research Objectives

The high-performance benefits of cloud computing must be discussed on a case-by-case basis and computational performance depends on the design and implementation of individual geospatial applications in specific cloud environments. Therefore, it's imperative to conduct real-world data-intensive geospatial computing studies using cloud computing to determine whether these technologies are viable solutions for data-intensive geospatial computing. The purpose of this research is to investigate and evaluate the feasibility and advantages/disadvantages of adopting cloud-computation technology in solve real-world geographical problem.

1.4 Study Area

Lantau Island is situated in the southwest of Hong Kong, China. In spite of some of the highest urban population densities in the world, 40% of the land area of the Hong Kong Special Administrative Region (HKSAR) is designated as country parks. The seeming incongruity is explained by the fact that the mountainous terrain is subject to landslide hazard, with urban development confined to 10% of the land area: a narrow coastal strip backed by steep convex slopes. Lantau Island is just characterized by its high occurrence of landslide induce by monsoon rainfall.

2 Methodology

2.1 Data and Analysis

2.1.1 Elevation

Elevation, from which the slope and aspect can be retrieved, essentially reflects the topography. The Digital Elevation Model (DEM) represents spatial variation in altitude. The slope and aspect data are derived from SRTM DEM with 90m resolution. Elevation values in the data of study area ranged from -15m to 888m (Fig 1).

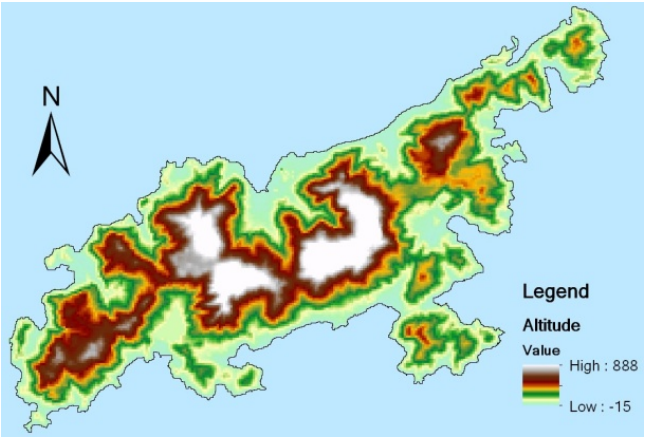


Fig. 1. Digital Elevation Model (DEM)

2.1.2 Aspect and Slope

Aspect is also an important factor in studying and producing land-slide susceptibility maps, since it will affect rainfall and exposure to sunlight [17]. In the geochaining platform, the topography aspect in Lan Tau island can be classified into nice categories: flat (−1°),north (0°–22.5°, 337.5°–360°), northeast (22.5°–67.5°), east (67.5°–112.5°), southeast (112.5°–157.5°), south (157.5°–202.5°), southwest(202.5°–247.5°), west (247.5°–292.5°) and northwest (292.5°–337.5°) (Fig 2).

The slope is an important parameter for stability consideration. The frequency of landslide is likely to be the highest on moderate steep terrain.

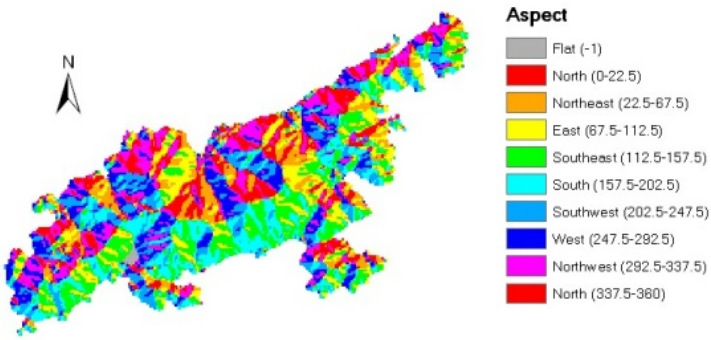


Fig. 2. Topography aspect of the study area

2.1.3 NDVI

Land cover extensively changes soil hydrology by raising rainfall interception, infiltration and evapo-transpiration [10]. In our study, a single data of Landsat TM image in 2007 with 30m resolution is acquired to calculate NVDI. The higher the NDVI is, the larger the area covered by vegetation. The experiment result indicates that area with low vegetation coverage is prone to landslide occurrence.

2.1.4 Landslide Inventory

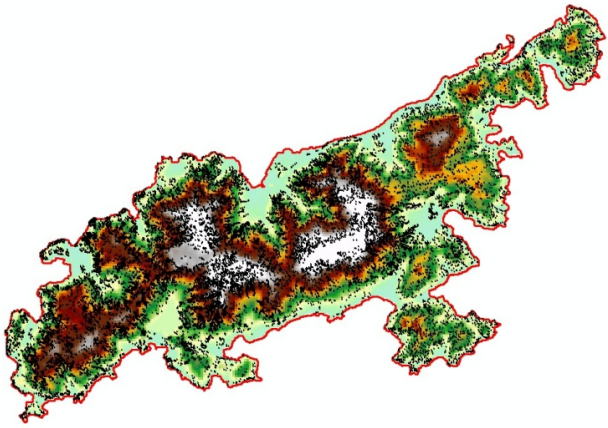


Fig. 3. Enhanced Natural Terrain Landslide Inventory for Lantau Island (1924-2009)

Most land resources of the Hong Kong Special Administrative Region (HKSAR) can never be utilized because of landslide hazard. The Hong Kong government has established the Geotechnical Engineering Office (GEO), which has created the Natural Terrain Landslide Inventory (NTLI). With such high quality data available, there is little need for other more, or equally efficient methods for landslide monitoring in Hong Kong [3] (Fig 3).

2.2 Cloud-Computation Platform

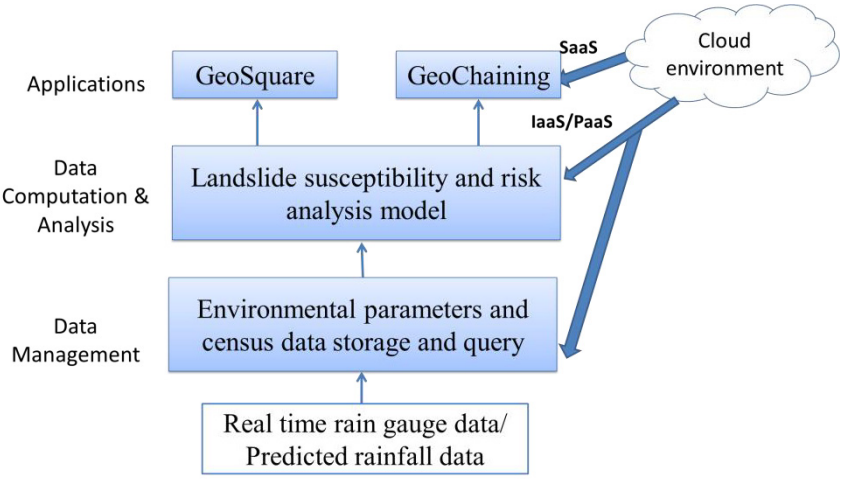


Fig. 4. Cloud-enabled system establishment

In order to reduce the difficulty of constructing Geospatial Web Service Chain (GWSC), so users without workflow and web services related knowledge can also easily create models, GeoChaining adopts a simple data-flow based GWSC meta-modelas model language. Users can perform model design only using several simple "drag-and-drop" operations. These operations include model and model element creation,properties editing, validation, model diagram enhancement. This platform isdeveloped by State Key Laboratory for Information Engineering in Surveying, Mappingand Remote Sensing, Wuhan University (Fig 4).

2.3 Landslide Susceptibility Model

Landslide susceptibility: A quantitative or qualitative assessment of the classification, volume (or area), and spatial distribution of landslides which exist or potentially may occur in an area (Fig 5).

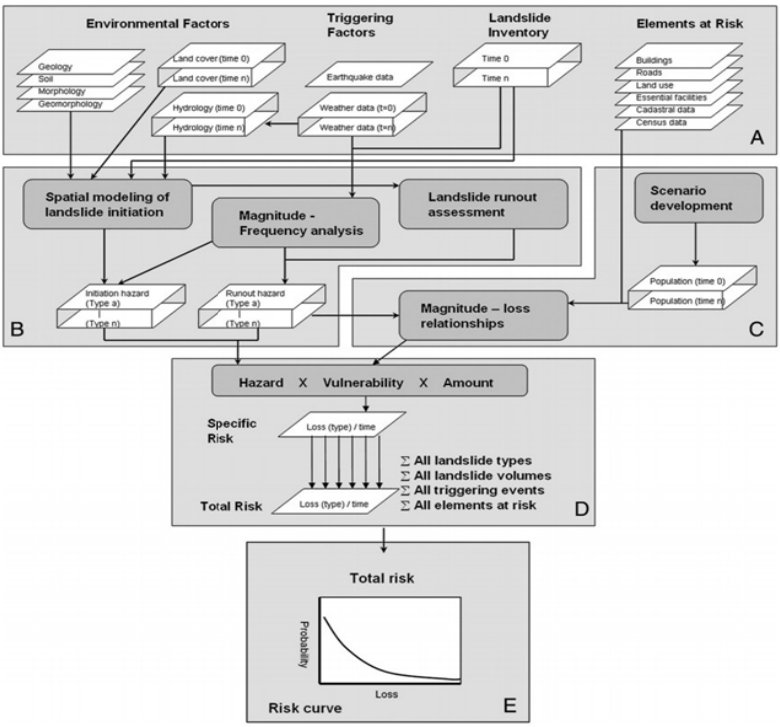


Fig. 5. Schematic representation of the landsliderisk assessment procedure. A: Basic data setsrequired, both of static, as well as dynamic (indicatedwith"time…")nature, B: Susceptibility and hazardmodelingcomponent,C:Vulnerabilityassessmentcomponent, D: Risk assessment component, E: Totalrisk calculation in the form of a risk curve [18].

3 Experiment Result

Landslides are caused by mutual interaction of various factors [19]. The above analysis presents a preliminary investigation of the causal relationship of some causal factors of landslide. According to the processing and analyses in the cloud-platform, It's well recognized that landslide occurrence in the study area have significant relationship with a variety of natural parameters.

4 Discussion

The cloud-computation platforms described above are still on a trial run, therefore may suffer from unstable network connection or hardware condition. Another major difficulty could be: When compared to the traditional desktop platform (ArcGIS), the cloud platform might not be offering sufficient geospatial analysis tool for our problem solving. It's also been recognized that there is an urgent need to investigate

how geoscience can leverage cloud computing to improve the performance to further improve the compatibility of scientific problems, and hide the complexity of computing infrastructure so that scientists can focus on geospatial problems solving [20].

This research is expected to provide a thorough evaluation of adopting the cloud-computing technology in solving big data geospatial computation problem. The final product will primarily be a geo-spatial hazard and risk information system, comprising graphic and numeric geo-spatial data, aerial and satellite images and ground thematic data, as an effort to evaluate the feasibility and advantage of cloud-computing method.

By solving a real-world problem, we will compare the cloud-computing method and traditional mapping method in a variety of aspects: scalability, computation efficiency, expressiveness and flexibility of cloud-enabled solutions.

5 Conclusion

This paper reports our preliminary analyses of the relationship among landslide and some environmental parameters on cloud-computation platform and traditional desktop platform (i.e. ArcGIS). The evaluation of landslide hazard is a complex task as the occurrence of landslide is dependent on many factors, in which large volume of spatial data processing and rapid computation process involved. It has been recognized that the emergence of cloud computing brings potential solutions to solve the geospatial intensity problems with elastic and on-demand access to massively pooled, instantiable and affordable computing resources [13]. On the other hand, a cloud-enabled landslide susceptibility warning mechanism would be even more helpful to enable the public to better prepare for such deadly events and to help mitigate any potential damages.

References

1. Cruden, D.M.: A simple definition of landslide. Bull. Int. Assoc. Eng. Geol. 43(1), 27–29 (1991)
2. Wang, H., Liu, G., Xu, W., Wang, G.H.: GIS-based landslide hazard assessment: an overview. Prog. Phys. Geog. 29(4), 548–567 (2005)
3. Nichol, J., Wong, M.S.: Detection and interpretation of landslides using satellite images. Land Degrad. Dev. 16, 243–255 (2005)
4. Ercanoglu, M.: An Overview on the Landslide Susceptibility Assessment Techniques. In: The 1st WSEAS International Conference on Environmental and Geological Science and Engineering (EG 2008), pp. 131–134 (2008)
5. Lee, S., Lee, M.–J.: Detecting landslide location using KOMPSAT 1 and its application to landslide-susceptibility mapping at the Gangneung area, Korea. Adv. Space Res. 38, 1095–1113 (2006)
6. Gorsevski, P.V., Gessler, P.E., Jankowski, P.: Integrating a fuzzy k-means classification and a Bayesian approach for spatial prediction of landslide hazard. J. Geog. Syst. 5(3), 223–251 (2003)

7. Guzzetti, F., Carrrarra, A., Cardinali, M., Reichenbach, P.: Lanslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. Geomorphology 31, 181–216 (1999)
8. Pradhan, B., Singh, R.P., Bucheroithner, M.F.: Estimation of stress and its use in evaluation of landslide prone regions using remote sensing data. Adv. in Space Res. 37, 698–709 (2006)
9. Pradhan, B., Mansor, S., Pirasteh, S., Buchroithner, M.F.: Landslide hazard and risk analyses at a landslide prone catchment area using statistical based geospatial model. Int. J. Remote Sens. 32(14), 4075–4087 (2011)
10. He, S., Pan, P., Dai, L., Wang, H., Liu, J.: Application of kernel-based Fisher discriminant analysis to map landslide susceptibility in the Qinggan River delta, Three Gorges, China. Geomorphology 171-172, 30–41 (2012)
11. Li, Q., Zhang, T., Yu, Y.: Using cloud computing to process intensive floating car data for urban traffic surveillance. Int. J. Geogr. Inf. Sci. 25(8), 1303–1322 (2011)
12. Gorton, I., Greenfield, P., Szalay, A., Williams, R.: Data-intensive computing in the 21st century. Computer 41(4), 30–32 (2008)
13. Yang, C., Goodchild, M., Huang, Q., et al.: Spatial cloud computing: how can the geospatial sciences use and help shape cloud computing? Int. J. Digit. Earth 4(4), 305–329 (2011)
14. Hearn, J., Zhang, Y.: Geographic Information Systems and Cloud Computing. Poster Presenting at Undergraduate Research Conference, Stephen F. Austin University. Nacogdoches, Taxes (2012)
15. Bhat, M.A., Shar, R.M., Ahmad, B.: Cloud Computing: A solution to Geographical Information Systems (GIS). Int. J. Comp. Sci. Eng. 3(2), 594–600 (2011)
16. Yang, X., Deng, Y.: Exploration of cloud computing technologies for geographic information services. In: The Proceeding of the 18th International Conference on Geoinformatics, pp. 18–20 (2010)
17. Süzen, M.L., Doyuran, V.: Data driven bivariate landslide susceptibility assess-ment using geographical information systems: a method and application to Asarsuyu catchment, Turkey. Eng. Geol. 71(3-4), 303–321 (2004)
18. van Westen, C.J., Castellanos, E., Kuriakose, S.L.: Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. Eng. Geol. 102, 112–131 (2008)
19. Saha, A.K., Gupta, R.P., Arora, M.K.: GIS-based landslide hazard zonation in the Bhagirathi (Ganga) valley, Himalayas. Int. J. Remote Sens. 70, 824–829 (2002)
20. Huang, Q., Yang, C., Benedict, K., et al.: Utilize cloud computing to support dust storm forecasting. Int. J. of Digit. Earth 6(4), 338–355 (2013)

Spatial Database Design and Realization of the HuaiHe Detention Basin in HeNan Province

ChengCai Zhang[1],WeiRan Luo[1], XiMei Sun[1], and XiHong Lv[2]

[1] College of Water Conservancy & Environmental Engineering, Zhengzhou University, Zhengzhou 450001, China
[2] The First Engineering Bureau of HenNan Water Conservancy, Zhengzhou 450000, China

Abstract. Using geodatabase spatial data model, combined with the actual situation of the detention basin of the Huaihe River Basin in Henan Province, through the steps of detention basin multi-source data preprocessing, spatial database designation, attribute database designation, etc., a detention basin spatial database of the Huaihe River Basin were established in Henan Province. The detention basin information management system of the Huaihe River Basin has been developed in Henan Province using C # language and ArcEngine component library. This system can realize the query, search and the statistics of detention basin, data input and output, and spatial data visualization, providing support for the scientific use of detention basin.

Keywords: Geodatabase, Spatial Data, Attribute Data, Detention basin.

1 Introduction

Flood storage and detention area is an important part of river basin's flood control engineering system. Due to the fact that our nation has a large population with relatively little land, flood storage and detention areas shall, in a long term, serve two functions: flood control and social economic development [1]. Huaihe River Basin has five flood storage and detention areas, which have been applied frequently, for example, Nihewa flood storage and detention area had diverted flood for 8 times in 1957, with a flood storage and detention duration of 22 days. During the flood seasons in 2007, Huaihe River basin had suffered the second biggest river-basin flood ever seen since 1954. In order to ensure the flood-control safety of important embankments and major places, multiple flood storage and detention areas had been applied in Henan and Anhui. With the rapid development of information technology, the traditional management way of flood storage and detention areas is no longer adapted to the demand of scientific management and application, requiring a more reasonable and advanced method to manage flood storage and detention areas. Geodatabase is a new-generation spatial data model introduced by ArcGIS, and is an uniform, smart spatial database built on DBMS. Managing spatial data with Geodatabase model can effectively enhance the management efficiency, and convenient geo-information combination and the integration of GIS [2]. Construction of the spatial database of flood storage and detention areas is of great significance to improving the scientific management level in flood storage and detention areas.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 240–248, 2013.

2 Design of the Spatial Database of the Flood Storage and Detention Areas in Huaihe River Basin in Henan Province

Main data used in the research: flood control chart of Huaihe River basin, thematic map data, digital orthophoto map data, remote sensing data of flood storage and detention areas. Several datasets have been built, based on data type, in the ArcCatalog of ARCGIS, with each containing the relevant feature class of certain data type, constituting the structure of the spatial database [3].

2.1 Design of the Spatial Database of the Flood Storage and Detention Areas

Design of the spatial database of the flood storage and detention areas is divided into two parts, which are graphic database and attribute database. The graphic database is mainly used to save the flood storage and detention areas' geo-data, thematic map data, digital orthophoto map data and DEM. The attribute database is mainly used to save the flood storage and detention areas' basic information, application data, user data and access data etc. As shown in Figure 1.

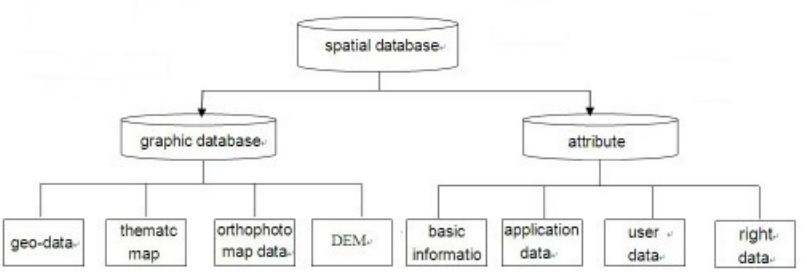


Fig. 1. Overall design of the Spatial Database

2.2 Design of the Graphic Database

Graphic entity data are abstracted as point, line and polygon, and organized by the data structures of network, path, area, raster, surface and so on, with different types of data expressed as different spatial data structures. Main considerations in designing the database are: which kind of layer information is needed, which kind of graphic data is fit to express the layer information, how to organize the layers as well as spatial scale and precision control. Graphic layer is a basic factor of spatial data organization, and the design and organization of graphic database apply the principle of layer. Graphic database of the flood storage and detention areas is made up of certain number of layer data, and the layers are saved in the form of Shapefile data in the system.

Table 1. Design of the graphic database

LAYER	TYPE	DESCRIPTION
the county administrative center	Point	Describe a flood storage and detention area in the county administrative center
towns	Point	Describe a flood storage and detention area of the township location
villages	Point	Describe a flood storage and detention area in each village
rivers	Line	Describe the shape of the river
a flood storage and detention area boundary	Line	Describe a flood storage and detention area boundary
the county boundary	Line	Describe a flood storage and detention area within the county boundary information
boundary of Henan Province	Line	Description of Henan province boundary information
a flood storage and detention area range	Polygon	Describe a flood storage and detention area range information
basin perimeter	Polygon	Information describing the flood storage and detention area in the basin
the scope of villages	Polygon	Describe a flood storage and detention area villages and towns range information

The Geodatabase model is described as follows:

Huaihe River Basin geo-data ……FeatureDataset

Layer feature factors ……FeatureClass

① River factors ……FeatureClass

② Flood storage and detention area factor ……FeatureClass

③ Iver basin factors ……FeatureClass

Map datum

Control data ……FeatureDataset

Control ……FeatureClass

Administrative data in Henan Province ……FeatureDataset

① Regional map of Henan province ……FeatureClass

② Boundary map of the county ……FeatureClass

The image database

① Imag data index charts ……FeatureClass

② A flood storage and detention area map ……RasterDataset

③ Flood storage and detention areas and remote sensing ……RasterDataset

DEM data

The flood storage and detention areas DEM ……Terrain

Map data index

① Administrative division index ……FeatureClass

② The important rivers index ……FeatureClass

③ Important gazetteer ……FeatureClass

④ The flood storage and detention areas index ……FeatureClass

2.3 Design of the Attribute Database

Based on the conditions of the flood storage and detention areas in Huaihe River basin in Henan province, design of the attribute database divides the data of the database into following categories:Basic information of the flood storage and detention areas: construction time, local river, local cities and towns, design water level for flood storage and detention, design volume of flood storage and detention, flooded area, chance of application, communication equipments and so on.Application information of the flood storage and detention areas: total times of flood storage and detention, flood storage and detention volume, flooded arable land, property loss etc.Villages and towns in the flood storage and detention areas: villages and towns, population, protected population, unprotected population, average elevation, statistic data etc.Infrastructure of the flood storage and detention areas: valve type, valve quantity, water level for flood diversion, opening and closing method.Table 2 shows the Flood Storage and Detention Areas attribute

Table 2. The Attribute Database

The serial number	Englishname	Function
1	BaseInform	Describe a flood storage and detention area
2	DetailHty	Use information describing a flood storage and detention area
3	Villages	Village information of flood storage and detention area
4	Gate	Describe a flood storage and detention area information infrastructure
5	Towns	Describe town information of flood storage and detention area

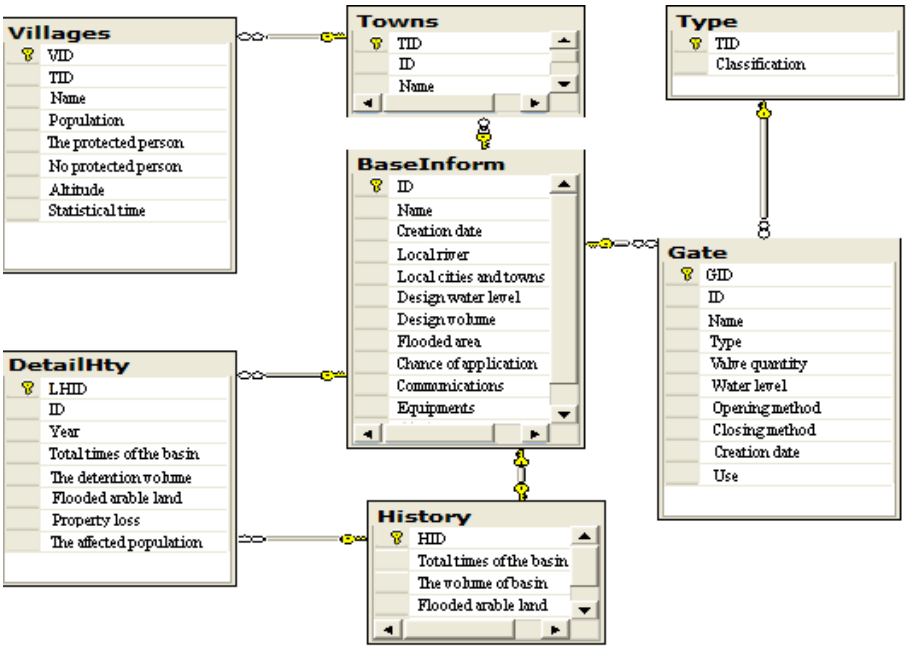


Fig. 2. Design of Attribute Database

3 General Design of the Information Management System of the Flood Storage and Detention Areas in Huaihe River Basin

3.1 Development Environment of the System

The system is developed by Client–server model, and the configuration of the system's development environment include the development environment of server end and client end.

3.1.1 Environment of Server End

Server end mainly realizes the maintenance and management of spatial graphic data and attribute database, and provides necessary spatial data support for the system. Environment of server end: Windows 2000 Advance Server is employed as the server's operation system. Database system: the Geodatabase of ESRI is applied as the system's database. Spatial data's engine: the ArcSDE developed by ESRI is applied as the system's spatial data management layer.

3.1.2 Environment of User End

The operation system of the user end of the system in this paper supports windows 2000 Professional, Windows2000 Server and Windows XP.

3.1.3 Environment of Development Platform

The system has applied C# and component library ArcEngine for system development, with the development platform Visual Studio 2008. ArcGIS Engine is a complete class library used by developers to build the component GIS for user defined applications, and its functions include five aspects: First, basic services provided by the GIS core: ArcObjects; Second, perform data access according to raster and vector's form, with a geo-database of abundant data; Third, it is used to create and display maps with marks and marking function, as well as create the map display functions such as thematic map and so on; Fourth, interface controls and comprehensive management system that can be applied rapidly for program development; Fifth, it can be arranged with the standard functions and other functions during operation.

3.2 General Structure of the System

The system consists four main function blocks: basic data operation block, data search block, data statistics block, system management block. As shown in Figure 3

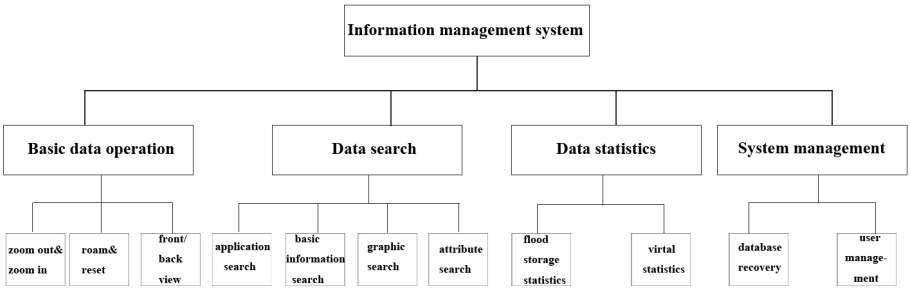


Fig. 3. System Function Module

4 Design of the System's Functions

4.1 Basic Operation Block

Basic operation block mainly includes the functions of zoom in, zoom out, roam, reset, front/back view, navigation and so on. There are three ways to zoom in: click-map zoom in, pull-box zoom in and wheel-roll zoom in; There are three ways to zoom out: click-map zoom out, pull-box zoom out and wheel-roll zoom out; The main way of roaming is: LB-click-map roam. Navigation is mainly used to fix the current main window's position in the whole map.

4.2 Data Search Block

The main functions of the data search block include: flood storage and detention areas' basic information search, application search, graphic search, attribute search etc. Flood storage and detention areas' basic information search refers to choosing flood storage and detention area and search the area's construction time, local rivers, local cities and towns, design water level for flood storage and detention, design volume of flood storage and detention, flooded area, chance of application, communication equipments and so on. Application search refers to choosing flood storage and detention area and search the area's total times of flood storage and detention, flood storage and detention volume, flooded arable land, property loss etc. Flood storage and detention areas' graphic search refers to choosing graphic layer through dialog box, and then use mouse to select the corresponding factors in the map, and view the selected factor's information, such as name, length and area etc., in the dialog box. Attribute search refers to choosing flood storage and detention area and the required information through dialog box, to acquire the application information of the corresponding flood storage and detention area, including local villages, valve type and so on. Figure 4 is Interface of Attribute Query in Lao Wangpo

Fig. 4. Interface of Attribute Query

4.3 Data Statistics Block

Data statistics block mainly collects the statistical data on the basic information and application information of the flood storage and detention areas, and then display the results in the form of statistical charts. Statistics of the basic information of the detention areas is mainly about certain flood storage and detention area's total times of flood storage and detention in history, flood storage and detention volume, flooded area of arable land and property loss, and display the results in the form of three-dimensional bar graphs. Statistics of the application information of the flood storage and detention areas is mainly about historical highest volume of flood storage and detention, the top water level for flood storage and detention, corresponding volume of flood storage and detention, duration of flood storage and detention, flooded area of arable land, flooded population, property loss etc, and display the results in the form of line graphs. Fig.5 is Histogram of flood detention's lasting time in Lao Wangpo Detention Basin

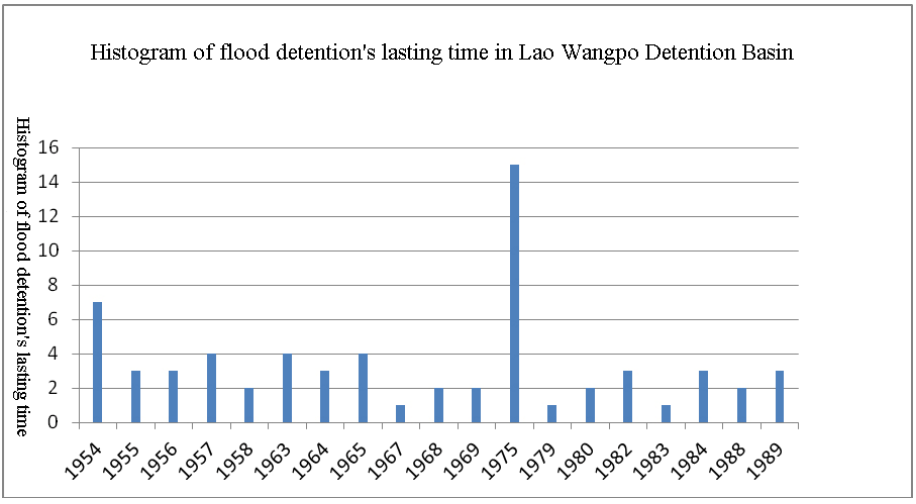


Fig. 5. Histogram of flood detention's lasting time in Lao Wangpo Detention Basin

4.4 System Management Block

Main functions of system management block: database backup, database recovery and user management. Database backup refers to making a backup of the data of the existing database in a magnetic disc, so as to prevent data loss in case of accidental database damage. Database recovery function can recover the data when the database has been damaged or lost. User management is mainly about user' setting and the right of access.

5 Conclusions

The spatial database model GeoDatabase has realized a seamless integration of spatial data and non-spatial data, and is able to make a full use of the functions of database management system (DBMS). This paper, by applying GeoDatabase model, has established the spatial database of the flood storage and detention areas in Huaihe River basin, and, by employing Structured Query Language(SQL), realized rapid two-way query between spatial data and non-spatial data.

The organization and management of the massive data are the difficult points and key points during system development. Spatial data's dynamic search has been realized by applying object-oriented technology, thus an information system has been built truly on client–server model, which, on the one hand, has resolved the management problems of massive data, and, on the other hand, fixed spatial data's problems like multi-user editing, data completeness, concurrency control and sate recovery etc. The information management system of has realized data functions like input, saving, maintenance, searching, analysis, output and so on for multiple kinds of data, such as geo-data, thematic map data, attribute data, digital orthophoto map data etc. The establishment of the system can effectively enhance the management level in the flood storage and detention areas in Huaihe River basin, and provide the basis for flood storage and detention areas' scientific, reasonable application and compensation.

Acknowledgement. Fund project: Basic research project of Henan Province, (No. 132300410031) Supported by Program for Innotive Research Team（in Science and Technology）in University of Henan Province（No. 13IRTSTHN030）.

References

1. Yang, B., Yang, P., Yang, X.: Design and Implementation of the Management System for the Detention Basin Based on WebGIS. Haihe Water Resources (6), 53–56 (2011)
2. Gong, J.Y.: Concepts and Development of Spatial Database Management System. Science of Surveying and Mapping 26 (2001)
3. Luo, Z., Liu, X.: Spatial Database Design and Realization Based on Geodatabase. Geo-Information Science 6(4), 105–109 (2004)
4. Xie, H., Chen, H., Wu, J.: Design and Development of Qingfengshan Basin SyntheticInquiry System Based on GIS. Journal of Yangtze River Scientific Research Institute 22 (2005)
5. Zhou, Z.: Design of Geospatial Database for infrastructure based on GeoDatabase. Journal of Geomatics 34(4), 48–49 (2009)

Exploring Location-Aware Process Management

Xinwei Zhu[1], Guobin Zhu[2], and Peichao Guan[3]

[1] International School of Software, Wuhan University, Wuhan 430079, China
{xinwei.zhu,gbzhu,guanpc}@whu.edu.cn
[2] International School of Software, Wuhan University, Wuhan 430079, China
[3] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China

Abstract. Context-awareness has emerged as an important principle in the design of flexible business processes. One important yet under-researched context variable is location; which we expect to have significant potential in making business processes truly agile. This paper proposes to extend context-aware Business Process Management (BPM) towards location-awareness in the management of business processes. We propose a model that explains where and how location variables matter to a business process. We presents a set of evaluation criteria to determine whether the business process location-aware or not. We expect our research-in-progress, together with further research, to bring an improvement of BPM development in theory and in practice.

Keywords: location-awareness, location-aware BPM, location-dependent control-flow patterns.

1 Introduction

Over the past years, Business Process Management (BPM) has become one of the abiding approaches to manage enterprises or organizations. In practice, BPM has led to significant improvements in business processes in terms of performance, operating cost, and compliance [1]. Organisations have committed large investments to identify, standardise, and document business processes with the intention to improve business performance or to enable subsequent automation through process-aware IS [2] and workflow management systems [3]. Yet, most of organizations continue to invest in the standardization and automation of core processes [4] leaving their enterprise systems insensitive to external process variables[5].

Context-awareness as a new paradigm of BPM deals with those dynamic business process environments, where processes can be rapidly changed and adapted to a new external context (such as weather, regulatory changes or location dependency). Contextualizing processes builds upon a more explicit consideration of the environmental setting of a process [4]. Location plays an irreplaceable role in better understanding of the context. The term of locality becomes a demonstrative pronoun to the context (e.g. what types of tools or equipment for emergency management staffs need to be pre-

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 249–256, 2013.

pared to rescue any victims in a specific location? How do I choose an un-ruined road to get to a safe place when I am driving a car?).

We suggest location-aware BPM as a research branch complementary to, and extending, context-aware BPM. In this paper, we attempt to distil and specify specific patterns in location-aware business process modelling as the first step of this research. The knowledge from complexity theory, dynamic theory, organizational theory, management science, information technology, especially existed research achievement from context-aware BPM and Business Process Modelling give a big confidence to the research of location-aware BPM [6].

The rest of this paper is structured as follows. Section 2 introduces the research background of context-aware BPM and existed research of location awareness. Section 3 builds a concept framework of location-aware BPM.. Section 4 provides evaluation criteria and an illustrative example for location awareness in business processes. Section 5 summarizes the contribution of the paper and discusses these research results and future work directions.

2 Research Background

2.1 Context-Aware Business Process Management

The business process management follows a lifecycle of design phase, implementation phase, enactment phase and diagnosis phase in support of operational business processes[7, 8]. Context-aware BPM complies with the lifecycle focusing on the extrinsic capabilities to assist the operation of the processes (e.g. decision-making). The foundations for the operationalization of this process-oriented approaches are business models[9]. They typically includes graphical depictions of at least the activities, events/states, and control flow logic that constitute a business process [10].There has been an increased recognition in process modelling to capture process-relevant information beyond the pure control flow and intend to model of the dynamic behaviour of organizations[9, 11].

Process modeling is a widely used approach for describing how businesses conduct their operations, as part of an effort to understand or analyze current operations (‘as is’ modeling), or as part of an effort to design improved blueprints for future operations (‘to be’ modeling). In either case, process models are specified using *process modeling grammars* [10]. These grammars provide sets of graphical constructs, together with rules for how to combine these constructs, to express graphically relevant aspects of business processes, such as the tasks that have to be performed, the actors that are involved in the execution of these tasks, relevant data, and, notably, the control flow logic that describes the logical and temporal order in which tasks are to be performed [12].

Looking at how these traditional models of business workflows can be extended to provide more contextual information for specialized decision-making. Two fields can be identified: One is concerned of the financial dimension (e.g. cost[13], risk[14], value[11]), an internal context perspective. [11]puts emphasis on the original extreme purpose of business activities–gaining economy value at maximum. [14]represents current research of managing risk in immediate context[4] aiming to improve the

ability to minimize risks at a design-time and to mitigate risks at a run-time. The other field of research prefers to study from the aspect of the context (e.g. context[5], green environment[15, 16], location), an environmental context perspective. [15, 16] attempt to optimize business process operations in a single process level from the environmental aspect aiming at Green-aware BPM or sustainability-aware BPM.

2.2 Location-Awareness

Location is a vital component in the environment context layer in the onion model for context awareness[4]. Different from some components (e.g. social-cultural system, political-legal system), which have a long life span, in this layer, location contextual variable can help process respond quickly and flexibly. We define a location as a geographical place or its topographical information (e.g. earth's surface features, vegetation, human-made structures, history, culture etc.). The location we mentioned in this paper refers to the geographical place (geo-location). The notion of location-awareness puts forwards the idea that location and location-based services can be sensed, aware and adapted in the processes.

Location-aware business process management defines a methodology to use location awareness in the management of business processes. It derives ability for a business process to sense the current process status in a specific location and to be aware of the whole process situation. Based on which, the resource (e.g. people) can react or dynamically change the process execution to adapt the goal of the process. Under the notion of location-aware BPM, we suggest location as an attribute in each business process that can be tracked anytime and anywhere, elevating sensitivity to any behaviour in process. As a trigger in process, it defines an ability to react the information intelligently, which makes an important decision for the resources. It makes process change, process redesign or process improvement automatic or semi-automatic possible.

Recent location-aware services and applications are used in the fields such as Navigation and Travel (e.g. WHERE, Yelp, Zagat, Google Places etc.), Tracking and Geosocial Networking (Foursquare, Facebook Places, Twitter, Glympse, TomTom Buddies, Google Latitude, Centrl, Google Latitude, Pocket Life and Porket Life Lite[17]etc.), Gaming and Entertainment (Flickr, iPhone Camera, scrabble etc.), Retail and Real Estate (Google Shopper, Hungry Jack, Woolworths, Commonwealths Bank etc.), Advertising (go2 Media, skyhook etc.), News and weather (ABC News, Scanner911), Device Management (Find My iPhone, Lookout etc.) and Public Safety (Amber Alert etc.)[18]. These and other technology-enabled services can be categorized as location-based services, location-assisted network optimization, location-assisted transceiver optimization and location-assisted environment identification [19-21].

The connection of location services to principles of business process management, however, is scarce in the literature. Many researchers focus on spatial-favoured optimization connecting location such similar resources (georeferenced data, geographical information, localization, positioning etc.) with scientific workflows, but not business workflow. For example, they utilize scientific workflow knowledge to optimize spatial decision-supporting system[22] such as GEO-WASA [23], Geo-Opera [24], Environmental management Information system (EMIS)[25], Workflow-based spatial

Decision Support System (WOODSS) [26], Kepler scientific workflow system [27-29] and so forth. However, different from the research we propose, those studies probe to use workflow knowledge but not optimize them. The existed research focusing on business workflow with location services is very limited. Until now, [30] being a notable exception uses Map metaphor to visualize work items and resources in process-aware information systems (e.g. YAWL).

3 A Conceptual Framework for Location-Aware BPM

In order to better understand the question how the idea of location awareness to connect to the lifecycle of business process management, a conceptual framework is proposed. It contributes to introduce a meta-location-aware process model into the traditional BPM lifecycle (see Fig.1.). A meta-process defines the objects (e.g. recourse, activity, data, application etc.), their relationship and attributes within a process definition[31]. The notion of location-aware mechanism embedded inside, which consists of location sensing, location awareness and location adaptation. This mechanism describes the common, generic features of a meta-location-aware process model allowing new rules of reconfiguring a meta-process in a more dynamic and automatic way.

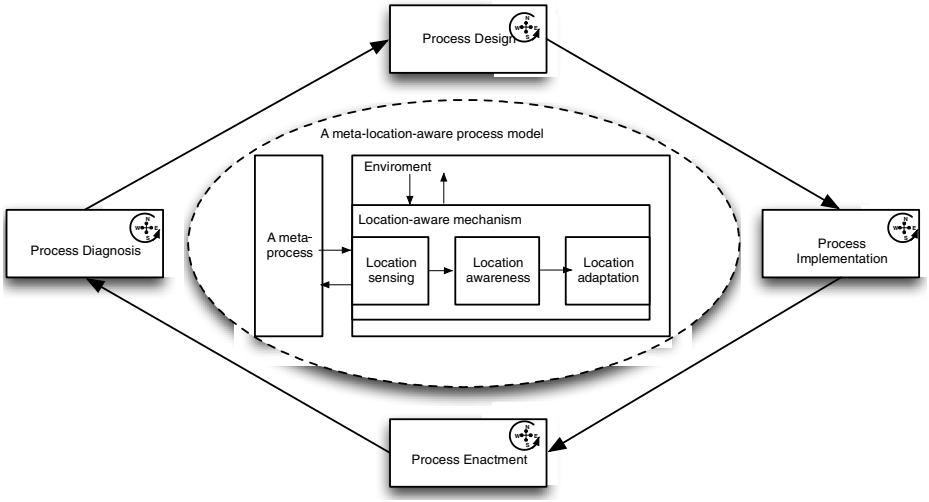


Fig. 1. Conceptual framework of location-aware BPM lifecycle

In process design phase, embedding the idea of meta-location-aware process model improves 'to be' processes location-focus ensuring a more accurate theoretical design. In the process implementation phase and process enactment phase, this helps to automate processes with location services in the form of application. In the forth phase, process diagnosis, identifies the potential actual bottleneck caused by location-related issue or potential opportunities for location-based improvement. In conclusion, it not only derives a location-aware process model, but also forms a basis for a flexible case-specific workflow instantiation.

4 Evaluating Location-Awareness in Business Processes

We suggest four main categories of evaluation criteria, these being reliability, adaptability, flexibility and efficiency explained below. And each category is further classified into several factors (see Table 1).

- Reliability measures the degree of location dependency in each process at any situation. It's a necessary condition of location-aware processes for making valid inferences, which impacts the quality of the processes.
- Adaptability measures capabilities to enable modifications in a location-aware workflow system on conditions that internal context changes in the processes.
- Flexibility measures location capability for accurately, rapidly reconfiguring tasks on conditions that external context changes in the processes. It's explored as an identifier of bottle necks and as a trigger for change actions[32].
- Efficiency: measures business value of processes under the idea of location-awareness.

The first three categories, reliability, adaptability, flexibility, are main means to measure its functionality. Theses criteria factors consider influence and scalability of location variables in processes. A good qualification of location services should enable processes to become more adaptable and flexible in light of location changes. The final category, efficiency, can be used to evaluate a location-aware process from the perspective of business value[11]. It's conducted to serve as the primary guideline to assess the financial value emanating from location-awareness.

Table 1. The hierarchical framework of location-aware business process evaluation model

Category	Factors	Description
Reliability[33]	Location acquisition	Evaluate ability to enable location service for acquiring location information
	Location accuracy	Evaluate the range of the accuracy of location information in a stable and changing environmental condition
	Location coverage	Evaluate the range of tasks related to a sound location
	Correlation to locations	Evaluate correlation between workflow and locations
Adaptability[34]	Modifiable	Evaluate ability of location-based scalability, location-based response accuracy, response time, exceptional handling
	Location sensibility to associated environment	Evaluate ability to identify location-based environment

Table 1. (*Continued*)

	Security	Evaluate security level of location-aware service
Flexibility[11, 32, 35]	Location-based predicate	Evaluate a series of predicating parameters related to the location area such as in or out location area, distance, velocity, density etc.
	Assisted decision-making ability	Evaluate ability of spatial decision support to related recourses' during the performance of tasks and its accuracy
	Dynamic Tracking ability	Evaluate ability of tracking a location-based task
Efficiency[11]	Satisfaction	Evaluate participants' satisfaction to the location-aware workflow
	Process cost	Evaluate location-aware process direct and indirect cost comparing with normal process
	Workflow lost	Evaluate location-aware process direct and indirect lost comparing with normal process
	Time cost	Evaluate location-aware process running time comparing with normal process
	Workflow quality	Evaluate quality of location-based task and workflow

The framework of location-aware business process evaluation model provides a means of addressing various location capabilities in processes. Basically, it's an effective way to evaluate location-dependent control-flow patterns. A set of theories (e.g. fuzzy theory) can be used in the evaluation, one of which suitable depending an evaluation objective.

5 Conclusions and Outlook

In this paper, we introduced the notion of location-aware BPM as a new perspective to extend context-awareness in BPM research. As a first step in this area, we sought to discuss how location to influence in a business process. The results of our initial exploration yielded an optimized conceptual framework of location-aware BPM lifecycle and illustrative evaluation criteria.

4 perspective and 14 factors are presented as a starting point for evaluating location-aware business processes. Each factor includes several criteria as suggested dependent measures. Some of parameters can be directly derived from the location-aware business process cases. Some can be acquired from a questionnaire survey. One basic benefit of the evaluation is the determination of a location-aware business process or not. Another important purpose is to test whether a location-dependent

workflow pattern to be a typical and common characteristic of a location-aware process.

One issue concerns the experiment to realize the idea of location-aware business process. Such trials (experiment identification) would provide evidence supporting the acceptance of location-aware business process patterns. In the future experiment, efficiency under the idea of location-aware business process modelling and normal business process modelling will be tested and compared. To achieve this, quantitative and qualitative evaluation methods both required. However, a much more extensive and longer lasting study would be needed to provide real assessment of the acceptance of location-aware BPM.

References

1. Hammer, M., Champy, J.: Reengineering the Corporation: A Manifesto for Business Revolution. Harper Business, New York (2003)
2. Dumas, M., van der Aalst, W.M.P., ter Hofstede, A.H.M.: Process-aware Information Systems: Bridging People and Software through Process Technology. John Wiley & Sons, Inc., New York (2005)
3. Jablonski, S., Bussler, C.: Workflow Management. Modeling Concepts, Architecture and Implementation. International Thomson Computer Press, London (1996)
4. Rosemann, M., Recker, J., Flender, C.: Contextualisation of business processes. International Journal of Business Process Integration and Management 3(1), 47–60 (2008)
5. Rosemann, M., Recker, J.C.: Context-aware process design: Exploring the extrinsic drivers for process flexibility. In: The 18th International Conference on Advanced Information Systems Engineering. Namur University Press (2006)
6. Ploesser, K., Recker, J.C., Rosemann, M.: Building a methodology for context-aware business processes: insights from an exploratory case study. Association for Information Systems (2010)
7. van der Aalst, W.M.P., ter Hofstede, A.H.M., Weske, M.: Business Process Management: A Survey. In: van der Aalst, W.M.P., ter Hofstede, A.H.M., Weske, M. (eds.) BPM 2003. LNCS, vol. 2678, pp. 1–12. Springer, Heidelberg (2003)
8. Zur Muehlen, M., Rosemann, M.: Multi-paradigm process management. In: Proc. of the Fifth Workshop on Business Process Modeling, Development, and Support-CAiSE Workshops (2004)
9. Green, P., Rosemann, M.: Integrated process modeling: an ontological evaluation. Information Systems 25(2), 73–87 (2000)
10. Recker, J., et al.: Business Process Modeling: A Comparative Analysis. Journal of the Association for Information Systems 10(4), 333–363 (2009)
11. Vom Brocke, J., Recker, J., Mendling, J.: Value-oriented process modeling: integrating financial perspectives into business process re-design. Business Process Management Journal 16(2), 333–356 (2010)
12. Mendling, J., Strembeck, M., Recker, J.: Factors of Process Model Comprehension — Findings from a Series of Experiments. Decision Support Systems 53(1), 195–206 (2012)
13. Wynn, M.T., Low, W.Z., Nauta, W.: A framework for cost-aware process management: generation of accurate and timely management accounting cost reports. In: Proceedings of Conferences in Research and Practice in Information Technology (2013); Conforti, R., Fortino, G., La Rosa, M., ter Hofstede, A.H.M.: History-aware, real-time risk detection in business processes. In: Meersman, R., et al. (eds.) OTM 2011, Part I. LNCS, vol. 7044, pp. 100–118. Springer, Heidelberg (2011)

14. Ghose, A., et al.: Green business process management: A research agenda. Australasian Journal of Information Systems 16(2) (2010)
15. Hoesch-Klohe, K., Ghose, A., Lê, L.S.: Towards Green Business Process Management. IEEE (2010)
16. Iqbal, R., et al.: Peripheral Display for Multi-User Location Awareness. Journal of Advances in Information Technology 1(3), 116–126 (2010)
17. Commission, F.C.: Location-based Services: an Overview of Opportunities and Other Considerations (2012)
18. Yarkan, S., Arslan, H.: Exploiting location awareness toward improved wireless system design in cognitive radio. IEEE Communications Magazine 46(1), 128–136 (2008)
19. Celebi, H., Arslan, H.: Utilization of location information in cognitive wireless networks. IEEE Wireless Communications 14(4), 6–13 (2007)
20. Celebi, H., Arslan, H.: Enabling location and environment awareness in cognitive radios. Computer Communications 31(6), 1114–1125 (2008)
21. DeMers, M.N.: Fundamentals of geographic information systems. John Wiley & Sons (2008)
22. Medeiros, C.B., Vossen, G., Weske, M.: GEO-WASA-combining GIS technology with workflow management. IEEE (1996)
23. Alonso, G., Hagen, C.: Geo-Opera: Workflow concepts for spatial processes. In: Scholl, M.O., Voisard, A. (eds.) SSD 1997. LNCS, vol. 1262, pp. 238–258. Springer, Heidelberg (1997)
24. Günther, O.: Environmental information systems. ACM SIGMOD Record 26(1), 3–4 (1997)
25. Seffino, L.A., et al.: WOODSS—a spatial decision support system based on workflows. Decision Support Systems 27(1), 105–123 (1999)
26. Altintas, I., et al.: Kepler: An extensible system for design and execution of scientific workflows. IEEE (2004)
27. Ludäscher, B., et al.: Scientific workflow management and the Kepler system. Concurrency and Computation: Practice and Experience 18(10), 1039–1065 (2006)
28. Jaeger, E., et al.: A scientific workflow approach to distributed geospatial data processing using web services. Citeseer (2005)
29. de Leoni, M., et al.: Visual Support for Work Assignment in Process-Aware Information Systems: Framework Formalisation and Implementation. Decision Support Systems (2012)
30. Hollingsworth, D.: Workflow management coalition: The workflow reference model. Document Number TC00-1003, 1995(1.1)
31. Georgoulias, K., et al.: Evaluation of flexibility for the effective change management of manufacturing organizations. Robotics and Computer-Integrated Manufacturing 25(6), 888–893 (2009)
32. Meyer, J.P.: Reliability. Oxford University Press, Oxford (2010)
33. Edmond, D., ter Hofstede, A.H.M.: A reflective infrastructure for workflow adaptability. Data & Knowledge Engineering 34(3), 271–304 (2000)
34. Ross, A.M., Rhodes, D.H., Hastings, D.E.: Defining changeability: Reconciling flexibility, adaptability, scalability, modifiability, and robustness for maintaining system lifecycle value. Systems Engineering 11(3), 246–262 (2008)

Establishment of Chlorophyll-a Concentration Distribution Model in Dahuofang Reservoir Based on HJ-1 Satellite

Qi Wang, Wei Meng, Yunfeng Ma, and Zhihong Sun

School of Energy and Environment,
Shenyang Aerospace University, Shenyang, 110136
mengwei012345@sina.com

Abstract. With the latitude and longitude coordinates and same place, 26 field monitoring data of Dahuofang was associated with HJ-1 satellite multispectral CCD data. This paper analyses the correlation between chlorophyll-a concentration and four bands, band combination of multispectral data. The Pearson's correlation coefficient was calculated by MATLAB software to find band combination T1, T27, T32 which have higher Pearson coefficient and then establish the linear model with these independent variables. Due to the complex optical characteristics of the water body, the relationship between chlorophyll-a concentration and the spectrum can't simply be described with a linear model. So in the paper we use BP neural network for modeling and prediction. The results show that BP neural network is better than the linear model and it can be used for the prediction of chlorophyll-a concentration.

Keywords: HJ-1, Dahuofang reservoir, chlorophyll-a, establishment of model, BP neural network.

1 Introduction

Earth's water resources are very rich, but 98% is salt water. Freshwater resources for human production and living are extremely valuable. As the population growth and the acceleration of urbanization our country's lakes and rivers are polluted to varying degrees, which aggravated water crisis [1,2], Dahuofang Reservoir is a large-scale water conservancy projects for utilization and the reservoir water quality is directly related to drinking water safety of residents in Fushun and Shenyang. Eutrophication of water body is an important measure of water quality parameters. When plank tonic algae plants are more and widely distributed, chlorophyll-a is an indicator to monitor the degree of eutrophication of water bodies. Detecting chlorophyll-a concentration needs sampling, filtering, and extraction process and instrument analysis [3,4], with large human and financial cost, low efficiency, and limited monitoring scope. Remote sensing monitoring becomes an important way of water quality monitoring due to short cycle and wide range of advantages. HJ-1 satellite by independent research and development in our country can achieve a wide range, all-weather, all-day dynamic

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 257–265, 2013.

monitoring. This paper use HJ-1 satellite data to build models to predict the chlorophyll-a concentration in water, in order to monitoring the water quality of Dahuofang reservoir eutrophication.

HJ-1 satellite launched on September 6, 2008, with a CCD camera single view data coverage of 185×185km , the spatial resolution of 30 m, including blue wave, green, red and near infrared wave four band [3,4], CCD image spectrum information for quantitative values of radiation uses MATLAB to read the CCD data and the data structure is m×n×4 three dimensional array, which m is the matrix rows, n is the number of columns of the matrix and number 4 represents the band. Assumes that the array is W, the data structure of four bands is as follows:

$$W(m,n,1)=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad W(m,n,2)=\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix}$$

$$W(m,n,3)=\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mn} \end{bmatrix} \quad W(m,n,4)=\begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ d_{m1} & d_{m2} & \cdots & d_{mn} \end{bmatrix}$$

2 The Inversion Regional Overview

Dahuofang reservoir is in North east of Liaoning province, Hun River, upstream of Fushun city, which is 18 km away from the center of Fushun, 68 kilometers from Shenyang and the geographical coordinates is 123° 39′ 42″ ~ 125° 28′ 58″ east longitude, 41° 41′10″~ 42° 38′ 32″ north latitude. Reservoir is about 35 kilometers long, and the water at its widest point is up to 4 km, the narrowest place of ribbon valley is 0.3 km, with a area of 5420.2 square kilometers and a maximum depth of 37 m. During 1990 ~ 2011, the water quality of Dahuofang reservoir is generally good, in addition to the total nitrogen and total phosphorus, and the rest of the water quality indicators of water quality monitoring project are accord with national II water quality standards [5].

3 The Collection and Processing of Data

3.1 The Measured Data

According to the time and space distribution of Dahuofang reservoir water quality monitoring for many years, increasing sampling points to 26 on the basis of the existing 6 station. Water samples were analyzed and tested by Fushun Dahuofang

reservoir monitoring station, sampling time: September 6, 2012 and the weather was well. The distribution of sampling points is shown in figure 1, the sampling points of the chlorophyll-a density is shown in figure 2.

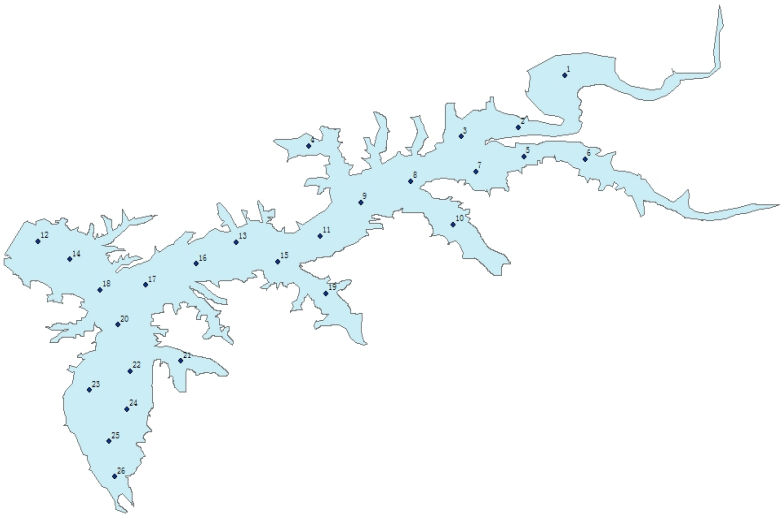


Fig. 1. Distribution of sampling point

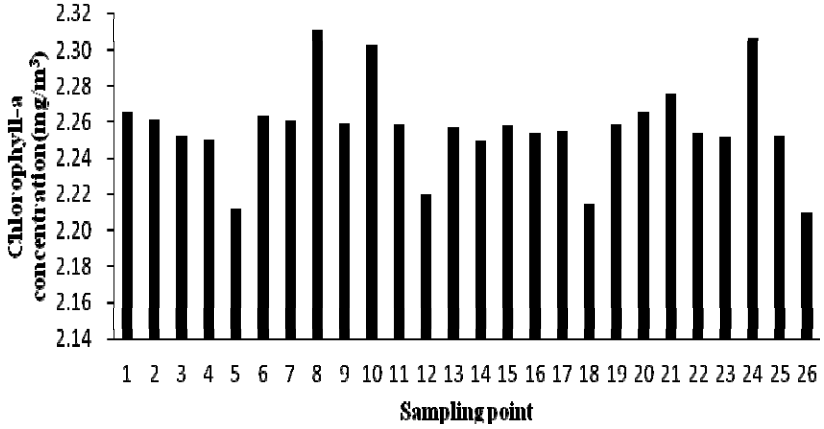


Fig. 2. Chlorophyll-a concentration of sampling point

3.2 CCD Data

Select the CCD secondary data which is after the radiation correction and geometric correction. The work was completed by the satellite environment Center, ministry of environmental protection. What we need to do is geometric precision correction,

atmospheric correction and the inversion of the chlorophyll-a concentration. The transit time of satellite is September 6, 2012.

3.2.1 The Geometric Precision Correction of CCD Data

With the help of ENVI 4.8 software, having geometric precision correction for the ground control points of HJ-1 satellite CCD image, which takes mathematical simulation for the process of geometric distortion of the original graphics via GCP data, establishes the corresponding relationships between original distortion image space and geographical standard image space and transforms the elements of image space into calibration graph space [3].

Correction method is: Image to Image. Base image is Landsat TM image whose geometric precision correction have been made. With the support of ENVI 4.8 software, having geometric correction for the remote sensing image, looking for the same point between HJ - 1 satellite Image and TM Image, controlling the total error within 0.5 pixel, adopting quadratic polynomial correction method and resampling with double linear differential method [6].

3.2.2 The Atmospheric Correction of CCD Data

Atmospheric correction is divided into absolute calibration and relative correction. The absolute calibration is what uses atmospheric radiation transmission theory model and transforms the DN value of remote sensing images into the surface reflectivity, surface emissivity, and surface temperature parameters and so on. This paper adopts the 6s (second simulation of the satellite signal in the solar spectrum) model for absolute atmospheric correction [3,7]. Assume that the target object whose reflectance is uniform lambert surface, then the 6S radiation transfer model is as follows:

$$\rho_{toa}(\theta_S,\theta_V,\phi)=\rho_a(\theta_S,\theta_V,\phi)+\frac{\rho_S(\theta_S,\theta_V,\phi)}{1-\rho_S(\theta_S,\theta_V,\phi)S}T(\theta_S)T(\theta_V) \quad (1)$$

$\rho_{toa}(\theta_S,\theta_V,\phi)$:Reflectance which was observed at the top of atmosphere by sensors; $\rho_a(\theta_S,\theta_V,\phi)$:Reflectance of the atmosphere radiation; $\rho_S(\theta_S,\theta_V,\phi)$: Reflectance of the real surface; *S*: Hemisphere reflectivity which is down at the bottom of atmosphere; θ_S : The solar zenith angle, θ_V : The satellite zenith angle, ϕ :The relative azimuth angle between the sun and moon; $T(\theta_S)$: The total atmosphere transmittance of incoming light from top level to the surface; $T(\theta_V)$:The total transmittance which goes upward into the direction of the satellite sensor; The main input parameters for atmospheric correction by using the ENVI software are shown in table 1.

Table 1. Parameters of atmospheric correction

Input parameter	Parameter value	Input parameter	Parameter value
Latitude	41, 53, 1.03	Flight Date	Sep, 6, 2012
Longitude	124, 13, 19.76	Flight Time GMT	02, 08, 19.94
Aerosol Retrieval	None	Atmospheric Model	Mid-Lat Summer
Title size(Mb)	100	Initial Visibility(km)	40
Water Retrieval	No	Ground Elevation(km)	0.175
Pixel Size(m)	30	Sensor Altitude(km)	650
Aerosol Model	Rural	Sensor type	UNKNOWN—MSI

3.3 Pearson Correlation Coefficient Method

Pearson correlation coefficient is used to describe linear correlation degree between two interval variable, usually expressed in r, the value of r is in [-1, +1]. If r>0, suggested that the two variables are linear positive association. If r<0, suggested that the two variables are linear negative association. If r=0, suggested that the two variables are linear independence. The calculation formulas:

$$r = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}} \tag{2}$$

N: Number of samples, $\overline{X}$: Mean value of X, $\overline{Y}$: Mean value of Y.

After atmospheric correction, the spectral matrix W is turned into Q, let A= Q(m,n,1) , B= Q(m,n,2) , C= Q(m,n,3) , D= Q(m,n,4). Band combination T1=D/C, that is to say T1*C=D, T1 is a matrix of m×m dimension. Select 17 sample values of X from m order matrix with the same name point method, which corresponds to sample chlorophyll-a concentration values of Y. Then use the formula to calculate correlation coefficient.

4 The Construction and Analysis of Model

4.1 The Best Band Combination

The date of HJ-1 satellite multispectral CCD data is September 6th, 2012. With the MATLAB software, it calculates the Pearson coefficients between 65 band combination and chlorophyll-a concentration and the higher band combinations of Pearson coefficient are T1, T27, T32.They are 0.9112, 0.9064, 0.8987 and T1=B4/B3, T27=(B4-B3)/(B4+B3),T32=(B1+B4)/(B1+B3).

The ratio of near infrared band and red light has the highest correlation and the two bands are reflection peak and absorption valley respectively [8,9], They are two sensitive wave band of chlorophyll-a.

The correlation degree was divided according to the Pearson coefficient and as shown in table 2. The calculation results show that the band combination T1, T27 and T32 has a strong correlation with chlorophyll-a concentration.

Table 2. Classification of correlative degree

Number	Pearson coefficent	Relevance class
1	0.8-1.0	Highly relevant
2	0.6-0.8	Strong correlation
3	0.4-0.6	Intermediate relevant
4	0.2-0.4	Weak correlation
5	0-0.2	Very weak relevant or irrelevant

4.2 Establishment of One-Dimensional Model

Through the calculation, it found out that he band combination which has the highest Pearson coefficient is T1=B4/B3.The Pearson coefficient of T1 and chlorophyll-a is 0.9112. So with the random 17 sampling data, takes the band combination T1 as independent variable to establish one-dimensional model. The rest of nine points as the validation data. Models and their comparison as shown in table 3.

Table 3. One-dimensional models and comparison

Model	Establishment of model	R^2	RMSE(mg/m^3)
Linear model	$y = 0.024x + 2.209$	0.8304	0.0382
Exponential model	$y = 2.2095e^{0.0106x}$	0.8308	0.0382
Polynomial model	$y=0.0065x^2-0.002x+2.2345$	0.8326	0.0384
Logarithmic model	$y= 0.047\ln(x) + 2.2246$	0.8244	0.0381

It can be seen from table 3, according to the degree of fitting R^2, polynomial model is better and according to the root mean square error of the validation data, the logarithmic model is better. It showed that the accuracy of the model which a single band established needs to be improved.

4.3 Establishment of Multi-Dimensional Linear Model

Combined with the above models, it showed that the model which certain single variable established could not have an ideal effect. Water chlorophyll-a concentration is in connection with multiple band combination in some extent. Some contribution rate is big, however some is small. So choose several band combinations which have a good correlation with chlorophyll-a concentration to establish multidimensional model.

$$y=2.1909+0.034*T1-0.076*T27+0.0156*T32. \quad (3)$$

$$R^2=0.8383, F=22.4595, P=0<0.05, RMSE=0.0382.$$

Compared with one-dimensional model, fitting R^2 has an evident rise and root mean square error has little changed. Both comprehensively, multidimensional model is better to reflect the chlorophyll-a concentration of Dahuofang reservoir.

4.4 Establishment of BP Neural Network Model

Due to the complex optical characteristics of case 2 waters, the linear model between chlorophyll-a concentration and spectral values is not precise. However, BP neural network has better adaptability, self-learning ability and fault-tolerant performance, which can simulate and forecast complex relationship [10,14].

The input layer of BP neural network has four nodes which are four band reflectance values (already atmospheric correction) of HJ-1 satellite respectively. The number of hidden layer nodes is important to the BP neural network, however the selection criteria of nodes is uncertain, so the number of hidden layer nodes ranged from 2 to 15 in this paper [13]. The transfer function is hyperbolic tangent function which is $y = \tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}$. The output layer has one node and the transfer function is linear function which is y=x. The training function is traingd. 17 sets of data were selected as training sample; the rest 9 groups as the validation data of the model.

In this paper, the minimum of hidden layer nodes is 2 and the maximum is 15. Calculate the fitting degree between predicted values which is obtained by BP neural network and the measured values. At the same time, calculate the root mean square error. The results as shown in table 4.

Table 4. The R^2 and RMSE of different hidden layer nodes

Hidden layer node	R^2	RMSE(mg/m^3)
2	0.9625	0.0339
3	0.9546	0.041
4	0.9667	0.0301
5	0.9657	0.031
6	0.9738	0.0237
7	0.9707	0.0265
8	0.962	0.0344
9	0.9671	0.0298
10	0.9728	0.0246
11	0.9648	0.0318
12	0.9585	0.0375
13	0.9724	0.0249
14	0.9637	0.0328
15	0.9567	0.0392

According to shown in table 4, when the number of hidden layer nodes of BP neural network is 6, the R^2 is the maximum, the RMSE is the minimum and the network error is the minimum, so the number of hidden layer is 6. Fitting degree R^2 =0.9738 which is greater than the linear model whose is 0.8304 and multidimensional model whose is 0.8383. At the same time, the RMSE=0.0237 which is less than the linear model and multidimensional model. Taken together, BP neural network can produce a better prediction result.

5 Conclusion

The multispectral band values of HJ-1 satellite have certain correlation with chlorophyll-a concentration. Compared to the established one-dimensional model, dimensional model and BP neural network model, it found out that the BP neural network has good prediction effect. Due to the training and learning of BP neural network need much sample data, if the measured data point is less, then network performance is not stable. So it should increase the measured data in the future.

Acknowledgments. This research was supported by Major Science and Technology Program for Water Pollution Control and Treatment (2012ZX07202-004-02).

References

1. Yang, Y., Wang, Q., Xiao, Q., et al.: Quantitative Remote sensing inversion methods of chlorophyll-a concentration in Taihu Lake based on TM data. Geography and Geo-Information Science 22(2), 5–8 (2006)
2. Du, C., Wang, S., Zhou, Y., et al.: Chlorophyll-a retrieval by concentrations classification in Taihu Lake based on remote sensing. Remote Sensing Information 1, 41–47 (2009)
3. Lu, H., Jiang, N., Luo, L.: Quantitative retrieval of chlorophyll-a by remote sensing in Taihu Lake based on TM data. Scientia Geographica Sinica 26(4), 473–476 (2006)
4. Liu, K., Huang, J., Zhang, Q.: Study on hyperspectra l monitoring of concentrations for chlorophyll-a of alga in Taihu Lake. Journal of Nanjing Normal University(Natural Science) 28(3), 97–101 (2005)
5. Wang, Q., Wei, B., Wang, C., et al.: Remote sensing monitoring of ecological environment based on satellite HJ-1, pp. 85–92. Beijing Science Press (2010)
6. http://www.secmep.cn/secPortal/portal/index.faces
7. Zhang, S.: The eutrophication status and its development trend of water quality of Dahuofang reservoir in Liaoning province. Heilongjiang Environmental Journal 36(1), 25–27 (2012)
8. Liu, W., Li, Y., Zhang, X., et al.: Image fusion and recognition, pp. 167–169. Publishing House of Electronics Industry, Beijing (2008)
9. Ghulam, A., Qin, Q., Zhu, L.: 6S model based atmospheric correction of visible and near-infrared data and sensitivity analysis. Acta Scientiarum Naturalium Universitatis Pekinensis 40(4), 611–618 (2004)

10. Kuang, D., Han, X., Liu, X., et al.: Quantative estimation of Taihu Chlorophyll-a concentration using HJ-1A and1B CCD imagery. China Environmental Science 30(9), 1268–1273 (2010)
11. Li, Y., Song, Y., Song, X., et al.: Modeling of chlorophyll-a retrieval based on remote sensing data-a case of Taihu Lake. Sichuan Environment 28(3) (2009)
12. Lu, H., Li, X.-G., Cao, K.: Quantitative retrieval of suspended solid concentration in lake Taihu based on BP neural net. Geomatics and Information Science of Wuhan University 31(8), 683–686 (2006)

Dyna-CLUE Model Improvement Based on Exponential Smoothing Method and Land Use Dynamic Simulation

Minghao Liu[1,2], Yaoxing Wang[1], Donghong Li[1], and Baobao Xia[1]

[1] School of Computer Science, Chongqing University of Posts and Telecoms, P.R. Chongqing, 400065, China
[2] Spatial Information System Research Center, Chongqing University of Posts and Telecommunications, Chongqing 400065, P.R. China
liumh@cqupt.edu.cn, {823618636,179128381,1036542372}@qq.com

Abstract. Response variables and their driving factors often vary with time in the process of land use dynamic simulation; however, there are few existing literatures mentioned it.In order to evaluate the impact of time factor on urban land use simulation and to improve simulation accuracy, exponential smoothing method was adopted to improve Dyna-CLUE model, as an example of Chongqing Metropolitan Area. Results showed that accuracies of simulation results vary from smoothing coefficients. When the smoothing coefficient was 0.4, the accuracy of the simulation was highest, and the accuracy of simulation results has improved 3.51%, compared to the simulation results without smoothing. Results indicated that time factor has impact on simulation accuracies of urban land use dynamic changes to a certain extent. The possible reason may be that the exponential smoothing method updates the regression parameters dynamically. Exponential smoothing method was introduced into Dyna-CLUE model in this study and results show that simulation accuracy can be improved which reflects that time series plays an important role in the simulation of land use dynamic.

Keywords: Exponential Smoothing, Dynamic simulation of Urban Land Use, Dyna-CLUE model, Main District of Chongqing.

1 Introduction

Land use dynamic simulation is designed to answer the questions such as when, where and why land use changes will occur and what outcomes will lead to. Urban system is an important part of the land system, so it is an urgent task to simulate land use dynamic in spatiotemporal scale. Dyna-CLUE version 2.0 (Dynamic Conversion of Land Use and its Effects model) is a widely used spatiotemporal model to test land use conversions and its effect. It is mainly by means of logistic regression equations to establish the relationship between land use conversion probability and driving factors. It combines 'top-down' approach and 'down-top' approach to complete dynamic simulation of land use under the constraints of land demands. From the point of view of the space simulation, Dyna-CLUE model itself does not propose solutions

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 266–277, 2013.

to spatial autocorrelation problems. The basic premise of logistic regression analysis is that spatial data are statistically independent and uniformly distributed, which violates Tobler's First Law of Geography :everything is related to everything else, but near things are more related than distant things" [1] (Tobler 1970).To solve this problem, Ms WU etc.(Gui-ping, WU,etc.2010) introduced space-related factor to solve autocorrelation effects of spatial data in the spatial statistical analysis[2].Regression analysis method is based on the Total Least Squares Method, which ignores spatial instability and spatial differentiation characteristics of driving factors. To overcome this deficiency, regression analysis method was replaced by Geographically Weighted Regression method (Yi-xi, Shao; Man-chun Li, etc. 2011) which take spatial differentiation effect of driving factors on land use pattern into consideration[3].In addition, Mixed Geographically Weighted Regression method[4] (Zhang Jie.2012) and spatial filtering method [5] [6] (Yang Yun-long,2012 ;Wu Kai-zhao,2010) have also been introduced to deal with spatial autocorrelation. Research on time factors on Dyna-CLUE model, existing literature mainly focus on the most appropriate time scale [7-9] (Ming Huang, etc., 2012; Miao Liu, etc., 2009; Shi-chao Feng, etc.,2013). The following method is usually adopted: Different time scales were set and then simulation precisions were compared under different scales to study the most appropriate time scale of Dyna-CLUE. However, as for the specific role of the time factors playing in the simulation process, there are few literatures mentioned. In fact, response variables and their driving factors vary with time in the process of land use dynamic simulation. Existing literature in which Dyna-CLUE model were used often assumed that driving factors used in land use dynamic simulation do not change over time or because of lack of data ,directly used driving factors in the base year to predict the land use changes in the target year.

Exponential smoothing method was proposed by Robert G. Brown. He argued that the trend of time series is stable or systematic， so time series can be reasonably homeopathic projections. More recently an event occurred; more possible it is to continue to the future. So greater weight will be given to the data occurred most recently[10](Robert G. Brown).

Exponential smoothing method calculates the exponential smoothing values, combined with a certain models of time series to predict the phenomenon in the future. Because of easy to understand, simply to calculate and excellence of performance exponential smoothing method has been widely used in the field of economic and natural sciences [11-15] (Xiao-si Shang, etc., 2011; Feng-xiao Wang, etc.,2012; Qiang Tong, etc.,2013; Wang M,, etc.,2012; Koehler A B, etc.,2012).

For example， Yun-long Yang etc., used exponential smoothing methods to forecast urban expansion in Zhangzhou city[16] (yun-long Yang, etc.,2011)， however there are also some deficiencies :1) binary data was used in the target variable of the model, which 1 indicated the appearance of urban land, while 0 indicated not. Land classification is too simple. 2) Do not integrate with existing land use dynamic models. 3) Despite probability distribution maps are obtained by exponential smoothing method, it was subjective to decide whether or not land use will change according to probability threshold in the process of simulation.

2 Study Area

Chongqing is located in the southwest of China, the upriver of Yangtze River and the southeast of Sichuan Basin, between 105°11' to 110°11' east longitude and 28°10' to 32°13' north latitude. Chongqing has an undulating terrain and various physiognomies which mainly include mountains and hills. So, Chongqing is also called “Mountain City”. Chongqing has various kind of physiognomy distinguish from other region obviously. In Chongqing, staggered distribution between urban and rural areas, which finally lead to forming a "multi-center, group-style" layout structure. The metropolitan area is about 5473 square kilometers, accounting for less than 9% of the whole city. Its altitude ranges from 141 to 1343m. In the region, "plain", "hill", "valley" and "flat" staggered dispersedly. The permanent resident population in metropolitan area is about 7.458 million, accounting for one quarter of the whole Chongqing. Metropolitan area is the center of the political, economic and transportation in whole Chongqing. After Chongqing became a municipality, economic develops fast, income gap between urban and rural increases. Statistical data shows the GDP of metropolitan area accounted for 92% of the whole Chongqing in2008, with less than 9% of the total area of Chongqing, Which not only indicates that there is a huge wealth gap between urban and rural, also reflects that the metropolitan area is still key areas of urban development in the future. Chongqing City Master Plan (1996-2020) puts forward that new space for urban development will be expanded, new transport system will be established to adapt mountain city’s character, and continue to controlling the population size in the old city and to promote the population growth in the new district and small towns. Urban land use patterns in the region will show different patterns in time and space as results of urban space expansions, improvements of transport facilities, demographic and other factors change in metropolitan area, etc. Unique topography and geomorphology in the study area are natural elements to determine urban form in the future, while rapid economic development, transportation and some other factors are social and economic elements to determine urban form in the future. Metropolitan area in Chongqing is an ideal zone to study the land use transformation.

3 Flowchart and Methods

3.1 Data Sources and Data Preprocessing

Dyna-CLUE model requires land use data and driving factors data at least for two periods. The amount of each land use type must account for more than 1% of total land use. Driving factors are divided into stable factors and unstable factors. Stable factors are factors that remain constant, while unstable factors changes with time [17].

According to the data availability and correlation, land-use data from the year of 2002, 2006, and 2009 was used. Eight driving factors data were chosen, including: 1) Topographical factors such as elevation, slope, waviness and roughness, etc.2) Accessibility factors such as the distance to roads, railways, water bodies, rivers, and the center of markets, etc.3) Socio-economic factors, mainly population density

The sources of Data: ①Land use data for 2002, 2006 were extracted from the fusion results of TM Remote Sensing Image Data, while land use data for 2009 was extracted from second national land survey data. ②Data of road got from open street map, with website http：//www.vdstech.com/osm-data.aspx. ③Demographic data as unstable driving factor from the Chongqing Statistical Yearbook was made by spatial interpolation. DEM image data got from the global land cover Research Institute: University of Maryland Earth Science Data Download Interface, with spatial resolution of 30m.

Data preprocessing: Firstly, land use data for 2009 was reclassified to six categories, namely urban land, arable land, forest land, pasture land, water body and other land as land use types studied in the paper according to the amount and distribution characteristics of land use type in the study area refer to the National Land Classification (Trial) (2002) on the platform of ArcGIS9.3.

Secondly, land-use data in the year of 2002 and 2006 was obtained after a series operation such as reclassification, unified geographic coordinates, map projection, unified scope of the study and resolution, etc.

Topographical factors such as elevation, slope, waviness and roughness, were made by spatial analysis function of ArcGIS.

Location factors such as distance to the road, water, the distance to the markets were made by buffer analysis.

3.2 Flowchart

This paper takes Chongqing Metropolitan Area as research area. Firstly , land use data for 2002, 2006 and 2009 and driving factors that affect land use change were select to build a quantitative relationship in order to get logistic regression parameters between the probability of urban land use conversion and driving factors by Logistic regression equations. Secondly, improved logistic regression parameters in every period were calculated by exponential smoothing method. Finally, land use pattern in the target year was simulated by the improved regression parameters(see Figure.1.).

3.3 Method

Allocation Principles of Dyna-CLUE Model. Dyna-CLUE model was used to simulate the dynamic changes of land system in this paper[18](Verburg P H, etc., 2009). The space allocation module of Dyna-CLUE model uses complexity system theory to deal with the competitive relationship between different land use types and simulates land use change synchronously.

$$Ptot_{i,t,lu} = ploc_{i,t,lu} + pnbh_{i,t,lu} + Elas_{lu} + Comp_{t,lu} \tag{1}$$

$Ptot_{i,t,lu}$ is the total probability of grid-cell i for land use type lu at time t ; $ploc_{i,t,lu}$ is the location suitability of grid-cell i for land use type lu at time t ; $pnbh_{i,t,u}$ is the neighborhood suitability of grid-cell i for land use type lu at

time t. $Elas_{lu}$ is the conversion elasticity, whcih is a measure of the cost from one land use type lu to another. The bigger $Elas_{lu}$ is, the harder it is convert land use type lu into another land use type, and vice versa. $Comp_{t,lu}$ is competitive advantage ,which is iteratively determined for all land use type during an iterative procedure. if the allocate area is smaller than area demand, values of $Comp_{t,lu}$ are increased while values are decreased when allocated area exceeds the demand. According to the order of $Ptot_{i,t,lu}$, Dyna-CLUE model constantly adjusts values of $Comp_{t,lu}$ and allocate land use type lu to spatial location in the study area after several iterations to achieve the purpose of space simulation.

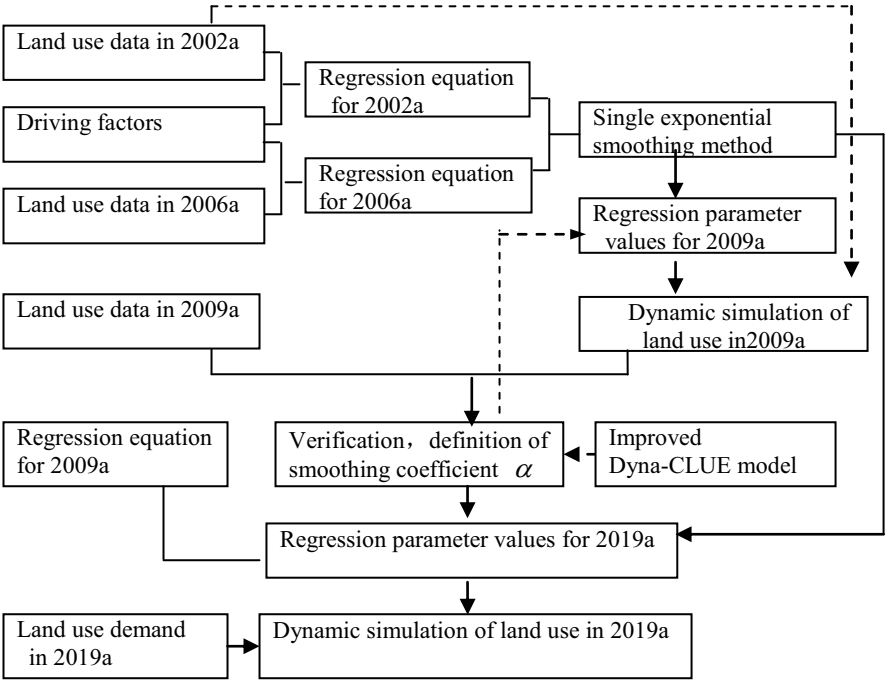


Fig. 1. Flowchart of land use change in main district of Chongqing

Location suitability and neighborhood suitability can be determined by empirical or expert knowledge. The interaction between neighborhoods is similar to constraint cellular automaton. Logistic regression was used in this study to calculate the quantitative relationship between land use type and driving factors in order to obtain land suitability and make evaluations to land use suitability.

Logistic regression equation as following：

$$\log(\frac{\mathbf{P}_{iu}}{1-\mathbf{P}_{iu}}) = \beta_0 + \beta_{1,}\mathbf{X}_{1,i} + \beta_{2,}\mathbf{X}_{2,i} + \ldots + \beta_{n,}\mathbf{X}_{n,i} \quad (2)$$

$P_{i,u}$ is the spatial distribution probability which is obtained by Logistic regression; $\mathbf{X}_i$ is the driving factor of i ; β_i are the regression coefficients of each factor, which need to save in the installation directory of Dyna-CLUE model with a name of allloc.reg .

In the process of simulation, contents in the allocreg.txt were read to access to land suitability parameters. Usually, only the quantitative relationship of land use and driving factors in the initial year were taken into consideration, as a result, effects of time factors on land use suitability were ignored. At a certain time interval period, the same driving factors have different effect on different land use types, and the same driving factors have different effect on a certain land use at different time intervals. Therefore, it is necessary to take time factors in to consideration so as to reasonably and accurately characterize land use change in the process of simulation.

Exponential smoothing method. There are some shortage in the traditional Dyna-CLUE model that directly using quantitative relationship between land conversions and driving factors in the base year to simulate the land use change in the future. Targeted with the imperfection of traditional Dyna-CLUE model, single exponential smoothing method was used to smooth the effect of driving factors on land use so as to improve spatial allocation module of Dyna-CLUE.

Exponential smoothing method is evolved from exponentially weighted moving average method. According to the difference of smoothing coefficient, exponential smoothing method can be divided into single exponential smoothing method, double exponential smoothing method and cubic exponential smoothing method. Single exponential smoothing method was used in this paper, which was suitable for short-term prediction of relatively stable time series in which growth or decline is not obvious. Smoothing coefficient α is critical in the use of exponential smoothing method.

α defines the weight of the most recent data in the predictive value and represents the response speed of forecast model to changes of time serials. The larger α is, the greater weight recent data account for, and vice versa. In addition, α also reflects the amplitude to be modified and determines the model's ability to correct errors.

Different smoothing coefficients were used to simulate land use for 2009. And the smoothing coefficient with the best simulated accuracy was select to smooth quantitative relationship between land use and driving factors in the future.

4 Results Analysis

4.1 Logistic Regression Analysis

First of all, driving factors and land use data were converted from initial format of raster into ASCII format.

Secondly, land use data and driving factors with ASCII format were merged into one file-stat.txt, in which the null value was removed by the file conversion software owned by Dyna-CLUE model.

Finally, the stat.txt files were Imported into SPSS(Statistical Product and Service Solutions, SPSS) 19.0 software, and the quantitative relationship between each type of land use conversion probability and driving factors were obtained respectively by Logistic regression model for 2002 year, 2006 year. ROC (receiver operating characteristic curve, ROC) was used to test the relationship between driving factors and conversion probabilities of land use type.

(1) According results of ROC test, distribution pattern of land use can be basically explained by driving factors selected in the paper. ROC values of urban land, forest land and water body are relatively high, while ROC values of the Pasture land and other land use were relatively low. The main reason is that pasture land other land use are scatter in distribution, so can not be described by general natural environment and economic features.

(2) The regression equations which were used to explain the relationship between conversion probability of land types and driving factors are different in different years.

Taking arable land as an example, ROC value in 2002 was 0.700, while it was 0.578 in 2006, which showed that the driving factors have better explanation ability in 2002 than in 2006.

As to pasture land, the ROC value in 2002 was 0.633, while it increased to 0.716 in 2006. The results showed that the explanation ability of driving factors to land use change is not fixed. The ROC value changes with time, which reflected land use conversion probabilities and their driving factors, had obvious characteristics of the times.

(3) Results of the value β indicated that different driving factors had different impact on a certain land type. As to the same driving factors, their influence to the conversion probabilities of same land type will vary with time. Table 2 shows that the driving factors such as roughness, waviness height have different valueβin different years.

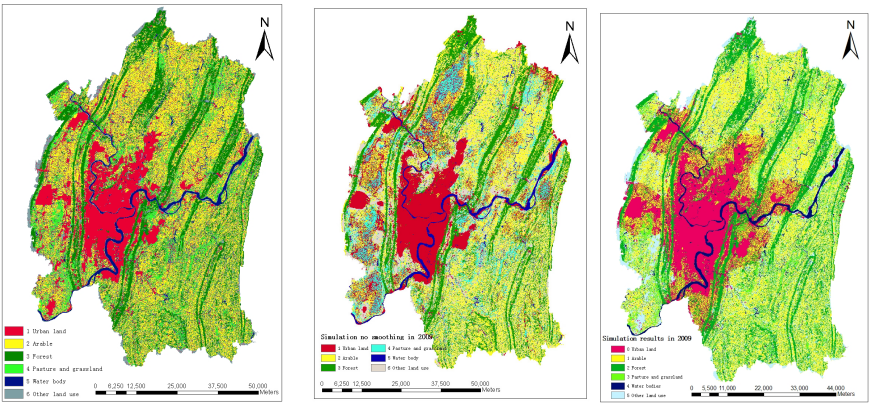
4.2 Parameter Verification

Simulation results show that the number of correct simulation grids were 558417 and Kappa index was 79.22% when exponential smoothing method was not adopted, while the simulation results were improved when single exponential smoothing method was adopted, with any smoothing coefficient (see Table 1), which reflected that in the process of simulation time characteristics owned by spatial data can not be ignored.

Table 1. Kappa Value of simulation results with different smoothing coefficients α

Smoothing coefficient α	0	0.1	0.2	0.3	0.4	0.5
Kappa value	79.22%	87.16%	87.11%	86.80%	86.27%	85.42%
Smoothing coefficient α		0.6	0.7	0.8	0.9	1
Kappa value		84.98%	83.76%	82.37%	80.76%	87.02%

When smoothing coefficient was 0.1, the number of correct simulation grids was 603131 and Kappa index was 87.16%. Compared with result without smoothing, the overall accuracy was improved by 7.94%.So 0.1 was identified the best smoothing coefficient as a result of best performance of Dyna-CLUE model. (see Table 2).



A_1 A_2 A_3

Fig. 2. Difference of simulation results in the spatial distribution
(A_1 is actual map of land use in 2009,A_2 is simulation results no smoothness, A_3 is the simulation results of smoothing coefficient 0.1)

Table 2. Comparison of accuracy simulation results

Land use type	Amount of 2009 actual land grids	Correctly simulated amount of land grids with unimproved model	Correctly simulated amount of land grids with improved model	accuracy of unimproved	accuracy of improved model
Urban land	117502	77354	84328	65.83%	71.77%
Arable	205756	192268	203242	93.44%	98.78%
Forest	138635	115421	135577	83.26%	97.79%
Pasture	60430	59146	59919	97.88%	99.15%
Water bodies	27231	26707	26714	98.08%	98.10%
Other land use	125816	87521	93351	69.56%	74.20%
Total	675370	558417	603131	82.68%	89.30%

Simulation results showed that the temporal characteristics of driving factors can not be ignored. It also showed that it was necessary to dynamically update the regression parameters when Dyna-CLUE model being used.

The number of grids misclassified is larger in other land and forest land when exponential smoothing method is not adopted. In the northwest or northeast regions, the misclassified grids were allocated properly when exponential smoothing method was adopted. Similarly, the forest land misclassified in the middle of the area was rectified when exponential smoothing method is adopted.

From the view of land use, simulation accuracy of each land use types were improved in a certain degrees when exponential smoothing method was adopted. Among them, the simulation accuracy for forest increased from the original 83.26% to 97.79%, increased by 14.53%. The improvement in water body is not obvious, only increased by 0.02%. The overall simulation accuracy was improved from 82.68% to 89.30%, increased by 6.62% (Table 2).

4.3 Land Use Change Simulation in 2019

The improved Dyna-CLUE model simulated land use change for 2009 in Chongqing Metropolitan Area well. Parameters obtained by using existing data were used to simulate land use change in future was reasonable. The exponential smoothing method were used to simulated and predict the land use pattern in the study area the next 10 years holding the hypothesis that to the study area maintain the current development.

Exponential Smoothing Again. Firstly, driving factors data and land use data in 2009 were analyzed by logistic regression methods and the regression coefficient-βvalue were acquired. Secondly, βvalue were smoothed again according to the smoothing coefficient determined above (Table 3). Dynamic change of land use was simulated for 2019 in the study area based on the smoothed results.

Table 3. Results of driving factors and land use smooth again（βvalue）

	Urban land	Arable land	Forest land	Pasture land	Water body	Other land use
DEM	-2.21*10-3	-1.064*10-3	3.59*10-3	-1.251*10-3	7.72*10-3	-4.06*10-4
Slope	-5.19*10-2	1.584*10-2	1.61*10-1	7.622*10-2	-1.83*10-1	2.592*10-2
waviness of surface	-2.09*10-3	-8.0*10-5	1.32*10-4	7.27*10-4	6.72*10-4	-1.45*10-4
Distance to markets	-1.17*10-4	3.40*10-5	-2.4*10-5	1.70*10-5	-2.0*10-6	2.30*10-5
Distance to roads	-9.00*10-5	2.5*10-5	1.10*10-5	-4.20*10-5	1.0*10-6	-1.00*10-6
Distance to water	1.47*10-4	-4.000*10-6	5.0*10-6	-1.00*10-5	-4.81*10-4	4.5*10-5
Population density	2.57*10-3	-3.261*10-3	-5.8*10-4	-1.889*10-3	-7.59*10-4	-7.3*10-5
roughness of surface	-6.45	-1.583*10	-1.160*10	-1.295*10	2.577*10	-7.173
constants	6.85	1.511*10	7.432	1.048*10	-2.524*10	5.550

Dyna-CLUE Dynamic Simulation. Land demands of each land use type for 2019 were predicted using the linear extrapolation method. Cov_all.0, cov_*.0, demand.in*, region, *.fil, sclgr*.fil, Main, alloc.reg and other files were prepared according to data format fit for Dyna-CLUE model. These files can be modified directly in the installation directory.

Land Use Pattern Analysis in the Future. Simulation results showed (Figure 3) that the construction land in Chongqing Metropolitan Area expand rapidly. With a large number of arable land, garden and pasture land were replaced by urban land.Urban expansion basically was constrained between Jinyun Mountains, Zhong liangshan Mountains, Tongluo mountains and the Mingyue mountains. The expansion trend of urban land is the most obvious to northward and westward, while southward and eastward of the area were followed.

Land use conversion matrix was established by overlay analysis of land use data in 2009 and 2019. The results showed the number of arable land, garden and pasture land use and other land use significantly reduced in this period, and the number of urban land and forest land increased, while the number of water body unchanged. Source of construction land is the garden and pasture land and the other land, followed by woodland. In the period of 2009 -2019, there are 14957.46 hectares of garden and pasture land use will be transferred into the construction land. There are 16607.43 hectares of arable land; 3137.13 hectares of pasture land use, 3790.80 hectares of other land will be transferred into forest land.

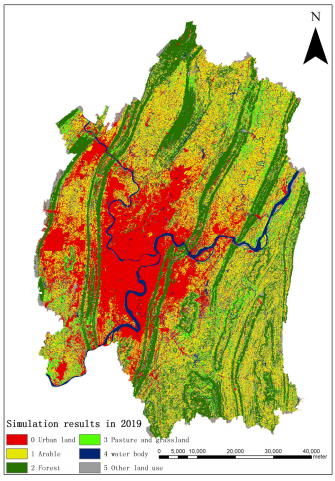


Fig. 3. Simulation of land-use of main district of Chongqing city in 2019

5 Conclusion and Discussion

Single exponential smoothing method was introduced in this paper to smooth the effects of driving factors on land transition probabilities, aiming at improving the simulation accuracy of Dyna-CLUE model. The paper taken Chongqing Metropolitan

Area as study region, logistic regression equation was established and improved Dyna-CLUE model was used to forecast land use pattern in the next 10 years. Results showed that:

(1) Result of ROC analysis showed that the ROC value and the β value in different land use pattern change with time. It indicated that the probability of land conversion and its driving factors have obvious time characteristics. The model simulation accuracy improved obviously when exponential smoothing method was used. It indicated that temporal characteristics of driving factors can't be ignored. It also showed that it was necessary to dynamically update the regression parameters.

(3)The regression parameters were updated dynamically by single smoothing method in the way of "strengthen the nearness and weaken the farness". The method improved the simulation accuracy and overcame the shortage of traditional Dyna-CLUE model which predicted dynamic land use in the target year by using driving factors in the base year since the driving factors is not always available in the target year.

(4) The simulation accuracy of the Dyna-CLUE model was improved by introducing exponential smoothing method in this study, with the Kappa index increased from 87.16% to 79.22%, and the overall accuracy of the model increased by 7.94%. If the time interval is short, or the change of the driving factors is not obvious, it is not necessary to use the exponential smoothing method; when the time scale is extended, such as 5 or 10 year, the accuracy of the model will be significantly improved by using the exponential smoothing method. The essence of Exponential smoothing is to alter the quantitative relationship between land use change and its driving factors which will inevitably exclude some valid information along with removing the influence of random factors. Therefore accuracy of a certain land use maybe reduced, but the overall simulation accuracy is improved which reflects that time series plays an important role in the simulation of land use dynamic.

Acknowledgment. This work was financially supported by Natural Science Foundation Project of CQ CSTC (Grant NO.2011jjA30014), and National Social Science Foundation of China (Grant NO. 11GBL051).

References

1. Tobler, W.R.: A computer movie simulating urban growth in the Detroit region. Economic Geography 46, 234–240 (1970)
2. Wu, G.-P., Zeng, Y.-N., Feng, X.-Z., et al.: Dynamic simulation of land use change based on the improved CLUE-S model: A case study of Yongding County, Zhangjiajie. Geographical Research 39, 460–470 (2010)
3. Liao, Q., Li, M., Shi, Y.-Q.: The Research on Land Use Pattern Simulation Using Geographically Weighted Regression and Improved CLUE-S. Shanghai Land & Resource 32(4), 31–37 (2011)
4. Zhang, J.: Research on the Spatial Variation of the Urban Housing Land Prices Based on Geographically Weighted. Zhejiang University (2012)

5. Yang, Y.-L., Zhou, X.-C., Wu, B.: Spatial distribution simulation of urban construction land expansion in Zhangzhou city. Journal of Fuzhou University(Natural Science Edition) 40(1), 63–69 (2012)
6. Wu, K.-Z., Wu, B.: Simulation of Urban Expansion Based on Logistic Regression with the Consideration of Spatial Correlation. Journal of Henan University(Natural Science) 40(3), 267–273 (2010)
7. Huang, M., Ru, H., Zhang, J., et al.: A Multi-Scale Simulation of Land Use Change in Luoyugou Watershed Based on CLUE-S Mode. Resources Science 34(4), 769–776 (2012)
8. Liu, M., Hu, Y.-M., Chang, Y., et al.: Analysis of temporal predicting abilities for the CLUE-S land use model. Acta Ecological Sinica 29(11), 6111–6119 (2009)
9. Feng, S., Gao, X., Gu, J., et al.: Land use spatial distribution modeling based on CLUE-S model in the Huangshui River Basin. Acta Ecological Sinica 33(3), 985–997 (2013)
10. Brown, R.G.: Exponential Smoothing for Predicting Demand, p. 15. Arthur D. Little Inc., Cambridge (1956)
11. Shang, X.-S., Lin, W.-D., Tang, Y.-K.: Development and Application of a Combined Water Quality Prediction Model Based on Exponential Smoothing and GM(1,1)—A Case Study of Iron and Manganese Levels in Yongjiang River. Environmental Science & Technonogy 34(1), 191–195 (2011)
12. Wang, F.-X., Zhou, W.-P.: Statistical Combination Forecast Method Based on Exponential Smoothing and Grey Mode. Journal of Kashgar Teachers College l33(3), 1–3 (2012)
13. Tong, Q., Zhang, K.-G., Du, J.-L.: Exponential Smoothing Forecasting Method and Its Application in Economic Forecast. Economic Research Guide 4, 11–13 (2013)
14. Wang, M., Zuo, X., Zuo, W., et al.: A Novel Combine Forecasting Method for Predicting News Update Time. Journal of Software 7(12), 2787–2793 (2012)
15. Koehler, A.B., Snyder, R.D., Ord, J.K., et al.: A study of outliers in the exponential smoothing approach to forecasting. International Journal of Forecasting 28(2), 477–484 (2012)
16. Yang, Y.-L., Zhou, X., Wu, B.: Urban expansion prediction for zhangzhou city based on GIS and spatiotemporal logistic regression model. Journal of Geo-Information Science 13(3), 374–382 (2011)
17. Yu, T., Ke, C.-Q.: The simulation of the land use change of Nanjing based on CLUE-S model. Science of Surverying and Mapping 35(1), 186–188 (2010)
18. Verburg, P.H., Overmars, K.P.: Combining top-down and bottom-up dynamics in land use modeling: exploring the future of abandoned farmlands in Europe with the Dyna-CLUE model. Landscape Ecology 24(9), 1167–1181 (2009)

Analysis of Land Use Changes and Driving Forces in Wuhan City

Fangyuan Chen[1], Xinsong Chen[2], and DengChao Ma[1]

[1] School of Remote Sensing and Information Engineering, Wuhan University , 129 Luoyu Road,Wuhan, China, 430079
[2] Basic Surveying and Mapping Institute of Jiangxi, 503 Nanlian Road, Nanchang, Jiangxi, 330001
805950339@qq.com

Abstract. Based on the statistic data of land use and interrelated social economic factors from 2004 to 2008 about Wuhan City, this paper analyses the range and dynamic degreen of land use change and its degree during past five years, and analyses the driving forces of land use change by making use of the analytical function of principal components of SPSS. The conclusions indicate: first, the area of the cultivated land has a downward trend, however, the area of residential land grows generally; second, the main driving factors affecting the land use change can be summarized as population growth, economic development, urbanization and the development of transportation industry.

Keywords: land use change, driving forces, analysis of principal components.

1 Introduction

Land use refers to a series of manner, process and results of land protection and transformation, as well as productive or non-productive activities by virtue of certain attributes of the land, through this process human beings can obtain certain economic, environmental or political welfare (benefits) [1]. In recent years, with the rapid economic development, population growth and growing urbanization, the analysis on driving forces of land use change has become a hot issue of global research. The research of issues on land use involves many aspects, the most critical is the research on driving force of land use. The driving force is the motivating factors of land use change; it refers to the main biological and socio-economic factors which lead to changes on the patterns and purpose of land use patterns. The driving forces of land use change are determined by the driving forces of land use patterns change, the human activities and the natural environment in this region are the basic driving force of a regional land use change [2]. In addition, socio-economic factors also have a deep impact on land use change. Depth analysis of the driving forces of land use change provides decision support on scientific land use planning and prediction on land use change trend.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 278–285, 2013.

2 Study Area, Data Sources and Methods

2.1 Overview of Study Area

Wuhan is located in central China, it is the capital and the political, economic and cultural center of of Hubei province. Wuhan is not only an important industrial city in the Changjiang River region, but also the economic center of this region. It is also the China's important cultural and educational center and the country's major transport hub. Wuhan is located at longitude 113 ° 41' to115 ° 05 ', latitude 29 ° 58' to 31 ° 22',it borders on Ezhou and Huangshi east,borders on Xiantao and Honghu west, connects with Xianning in south and Xiaogan in north.It is likely a butterfly from west to east. The Changjiang river and its largest tributary Hanshui meet in this region, so, Wuchang, Hankou and Hanyangthe, which is the tripartite confrontation of the two rivers , are commonly known as the three towns of Wuhan. Wuhan is in the superior center of economic geography circles in China, it is China's heart geographically, so It is called "thoroughfare of nine provinces" . By the end of 2008, the total population of the city was 8,332,425 , while the agricultural population was 2,960,061 , GDP had achieved 411.551 billion yuan ,GDP per capita had reached 46,035 yuan.

2.2 Data Sources and Methods

The research involves the statistic data of land use from 2004 to 2008 in Wuhan City, and these data mainly the situation of land use change of each year provided by wuhan land resources and planning bureau. Population, economic data and so on mainly come from wuhan city statistical yearbook.

According to the land use data of each year, this paper analyzes the structure changes of various types of land use first. Second, it chooses 24 indicators which can reflect the information population, economic and agricultural as the driving force factors of land use change. Finally this paper extracts the main driving factors affecting land use change, and analyzes the various driving forces quatitatively by making use of the analytical function of principal components of SPSS.

3 Analysis on Land Use Change of Wuhan

Based on the statistic data of land use from 2004 to 2008 in Wuhan City, this paper analyses the range and dynamic degreen of land use change and its degree.

3.1 Analysis on Range and Dynamic Degree of Land Use Change

The range of land use change refers to the change of land use type area within a certain period of time. We can realize the situation of land use change and the change of land use structure through the analysis the range of land use change. Its mathematical expression is:

$$Rd = \frac{U_b - U_a}{U_a} \times 100\% \quad (1)$$

The dynamic degree of land use[3] shows the the number of land use type change within a certain period of time. Its mathematical expression is:

$$K = \frac{U_b - U_a}{U_a} \times \frac{1}{T} \times 100\% \quad (2)$$

U_a, U_b stands for the area of a certain land type at the beginning of the study and at the end of the study respectively. T indicates the time period of the study, taking years as the unit[3].

In this paper, the range and dynamic degree of land use change are calculated by the EXCEl. According to Table 1.,we can find out: from 2004 to 2006 the rate of Grassland reduces the sharpest, followed by unused land and arable land, and the traffic land has the largest increase; during next two years the Residential land and traffic land has a relatively larger increase, while the cultivated land, garden land, forest land, other cultivated land all showed a trend of decrease. It also shows that the farmland gradually transform to the land for the urban construction in Wuhan city.

Table 1. The change ectent and velocity of land use in wu han city from 2004-2008

Land use pattern	2004-2006		2006-2008	
	Rd	K	Rd	k
Cultivated land	-8.81%	-4.40%	-1.33%	-0.67%
Garden land	8.01%	4.01%	-2.69%	-1.34%
woodland	13.28%	6.64%	-0.37%	-0.18%
grassplot	-95.94%	-47.97%	1.89%	0.94%
Other agricultural land	10.48%	5.24%	-1.58%	-0.79%
Residential land	9.06%	4.53%	5.34%	2.67%
Traffic land	34.57%	17.29%	10.76%	5.38%
Water conservancy facility land	-0.43%	-0.22%	-0.42%	-0.21%
Unused land	-24.79%	-12.39%	-1.67%	-0.83%
Other land	7.32%	3.66%	-0.16%	-0.80%

3.2 Analysis on the Degree of Land Use Change

Land use degree mainly reflects the breadth and depth of land use. It not only reflects the natural properties of land itself, but also reflects the combined effects of the human factors and natural environmental factors [3].Changes of Land use degree are the results of comprehensive changes of land use categories. Land use degree and its changes can quantitatively express the comprehensive level of land use and trends in the area.

Based on Mr. Liu Jiyuan's comprehensive analytic methods of land use degree [4], land use degree in this essay is divided into several levels according to land natural complex under the natural equilibrium state in the influence of social factors, and graded index is given. Thus composite index of land use degree [3, 5] and the temporal evolution model expression of the degree of land use are given [5]. We Adopted Mr Liu Jiyuan 's diversity standard of land use degree proposed from the perception of ecology as our grading standards[6], as shown in Table 2.

Table 2. The classification and evaluation of land use degree

Grading type	Level of Unused land	Level of Forest/grass/water land use	Level of Grad agricultural land	Level of urban land settlements
land use pattern	Unused land	Woodland\ grassplot\ Water conservancy facility land	Cultivated land\ Garden land\ grassplot	Residential land\ Traffic land
Graded index	1	2	3	4

Land use degree index expression:

$$I = 100 \times \sum_{i=1}^{n}(A_i \times C_i) \qquad I \in [100,400] \tag{3}$$

Extent of land use change expression:

$$\Delta I_{b-a} = I_b - I_a = [\sum_{i=1}^{n}(A_i \times C_{ib}) - \sum_{i=1}^{n}(A_i \times C_{ia})] \times 100 \tag{4}$$

In the expressions ,A_i,C_i are graded index of i-class land use degree and graded area percentage respectively; n is the number of classification of land use degree; C_{ib},C_{ia} are graded area percentage of the first class land use degree of a time period and b , respectively. The meanings of extent of land use changes are: If ΔI_{b-a} is greater than zero, it shows that regional land use is in development stage; If ΔI_{b-a} is less than zero, it shows that regional land use is in a recession or period of adjustment. When ΔI_{b-a} is the same with zero, the quality of the ecological environment is not to be reflected.

We employed the land use data of Wuhan City as composite index and the extent of land use temporal evolution model of land use degree expressions. We got the results with EXCEL calculation, as shown in Table 3. Land use degree indexes are larger than 200 and are in an upward trend, indicating the degree of land use of Wuhan City will continue to maintain a high level in the coming years. Extent of land use changes is less than zero from the year 2004 to 2006, indicating that in this time land use of Wuhan City is in an adjustment period; extent of land use changes from the year 2006 to 2008 is greater than zero, indicating that in the next two years land use of Wuhan City is in development stage. It also shows that during different periods the degrees of changes in land use are different.

Table 3. The aggregate index and variable quantity of land use degree in Wuhan City from 2004 to 2008

year	I	ΔI_{b-a}
2004	257.36	
		-1.25
2006	256.11	
		1.29
2008	257.4	

4 Analysis on Driving Forces of Land Use Change of Wuhan

The driving factors of urban land use change mainly includes the social economic forces, natural driving forces and land use policies and other factors. The effect of land use policies on land use change is difficult to quantify, and natural driving force is small for the relatively small scale area in Wuhan. So the two factors should be considered in statistical analysis only. According to the requirement of the principal component analysis method and the situation of existing data about Wuhan, we select the following 24 variables as the analysis indicators(Table 4.).

The data have different units, in order to safeguard the availability and the reliability of statistical data, standardized processing of raw data is required,this processing is handled prior to the statistical analysis through the SPSS software. Then use the software for calculation of the principal component analysis, according to the requirements of the principal component analysis, the number of principal components is determined by the method using smallest factor eigenvalues greater than 1 ,and rotate principal components though varimax method. Ultimate, calculate the characteristic value,Proportion, Cumulative(Table 5.) and the principal component loading matrix(Table 6.).

Table 4. The Indicators System of Driving Force in Land Use Change

Influencing factors	Analysis index	Influencing factors	Analysis index
Population growth	X1：total population	urbanization	X14：total retail sales of consumer goods
	X2：the agricultural population		X15：urban per capita disposable income
	X3：labor force non-agricultural rate		X16：rural per captia net income
	X4：population density		X17：the total output of crops

Table 4. (*Continued*)

Economic Development	X5：gross domestic product(GDP)	Agricultural production	X18：total power of agricultural machinery
	X6：total cost of the first industry		X19：single output of grain
	X7：the second industry gross		X20：Sown Area of Farm Crops
	X8：the third industry gross		X21：Cultivated area
	X9：per capita gross domestic product		X22：Gross Output Value Agriculture
	X10：total industrial output value	traffic	X23：Freight amout in the whole society
	X11：total value of agricultural output		X24：Passengers in the whole society
	X12：social fixed assets investment		
	X13：fiscal revenue		

Table 5. Result of Principal Component Analysis for Driving Force Factors

Principal component	Characteristic value	Proportion(%)	Cumulative(%)
1	20.167	84.028	84.028
2	2.709	11.287	95.314

It can be seen from the Table 5., the he cumulative contribution rate of the first two principal components has reached 95.314%， Use them to represent the most information of the original factor. Compare the first two principal components variance contribution rate, the first principal component contribution rate has reached 84.028%, and among these, most of the variable load coefficients is larger. According to the Table 6., the load of the first principal component factor are generally more than 0.9,only the loads of the total number of agricultural population X2, agricultural output X17, single grain production X19,crops sown surface X20 are relatively small. So, the first principal component can be considered as the representation of population growth, economic development, urbanization and transportation. The second principal

Table 6. Componet Matrix of Driving Force Factors

	1	2
Zscore(X1)	0.969	0.228
Zscore(X2)	-0.681	0.496
Zscore(X3)	0.984	0.111
Zscore(X4)	0.969	0.226
Zscore(X5)	0.988	-0.153
Zscore(X6)	0.988	-0.140
Zscore(X7)	0.988	-0.149
Zscore(X8)	0.988	-0.156
Zscore(X9)	0.984	-0.176
Zscore(X10)	0.934	-0.356
Zscore(X11)	0.983	-0.160
Zscore(X12)	0.991	-0.126
Zscore(X13)	0.998	-0.051
Zscore(X14)	0.989	-0.147
Zscore(X15)	0.996	-0.082
Zscore(X16)	0.985	-0.167
Zscore(X17)	-0.187	0.931
Zscore(X18)	0.995	0.070
Zscore(X19)	0.115	0.989
Zscore(X20)	-0.668	0.133
Zscore(X21)	0.861	0.425
Zscore(X22)	0.988	-0.135
Zscore(X23)	0.951	-0.281
Zscore(X24)	0.981	0.094

component is closely related to X17and X19 ,and these two factors have a larger positive correlation with agricultural production, therefore the second main component may be considered as the representation of agricultural production. So, the main driving factors impacting the land use change in Wuhan are population growth, economic development, urbanization, transportation, followed by agricultural production efficiency.

5 Conclusion

Through the analysis of land use change and driving forces in Wuhan, we can find：the degree of land use in Wuhan for the past few years has a upward trend in the first post-recession, and will be at a high level of development in the next few years; Cultivated land, Garden land, woodland, Other agricultural land, Water conservancy

facility land and Unused land would continue to decrease, at the same time the residential land and traffic land would still increase. According to the result of principal component analysis, it is concluded that the main driving factors affecting the land use change in Wuhan are population growth, economic development, urbanization and the development of transportation industry. In order to maintain the sustainable utilization and development of land resources in Wuhan, the government should take corresponding measures to protect agricultural land, improve the efficiency of land use efficiency, optimize the structure of land use and attach importance to urban ecological environmental governance.

References

1. Li, P., Li, X., Liu, X.: The macro analysis of driving forces of land use change. Journal of Geographical Research 20(2), 129 (2001)
2. Bai, W., Zhao, S.: Land use change driving force system analysis. Journal of Resources Science 23(3), 39 (2001)
3. Wang, X., Bao, Y.: Dynamic change of land use research methods to explore. Journal of Geographical Science Progress 18(1), 81–86 (1999); Wang, X., Bao, Y.: Changes in land use methods of. Progress in Geography 18(1), 81–87 (1999)
4. Fan, Y., Liu, J.: Tibet autonomous region land use. Science Press, Beijing (1992)
5. Zhuang, D., Liu, J.: China's land use degree of regional differentiation model. Journal of Natural Resources 12(2), 105–111 (1997)
6. Zhuang, D., Liu, J.: The extent of land use model of regional differentiation. Natural Resources 12(2), 105–111 (1997)
7. Liu, J.: China's resources and environment of macroscopic investigation and dynamic research, pp. 158–188. China Science and Technology Press, Beijing (1996)

Spatial Analysis of Gymnasiums in Wuhan City

Qing Han[1], Chao Yin[1], and Jiangping Chen[2]

[1] Resources Environment and Tourism Management Department, Hengyang Normal University, 16 Henghua Road, Hengyang, Hunan, 421002
[2] School of Remote Sensing and Information Engineering,Wuhan University, 129 Luoyu Road, Wuhan,China, 430079
chenjp@163.com

Abstract. Spatial distribution of gymnasiums in Wuhan is displayed by density analysis and spatial auto-correlation. Creatively considering distribution of roads, spatial autocorrelation and spatial binary correction are being used in quantitative studying how the spatial distribution relates to economical factors social factors and nets of roads. According to the reports, distribution of gymnasiums is influenced by density of population, roads and GDP in different degree. Based on this motivation system with the development of non-central area and improvement of roads, the mount of gymnasiums will increase rapidly. And the whole gymnasiums distribution of Wuhan will transform into "intensive central city area and popular country" distribution pattern.

Keywords: spatial autocorrelation analysis, spatial correlation analysis, road nets.

1 Introduction

With the continuous development of the sports industry, the implementation of the National Fitness Program and the promotion of the National Fitness Activities, sports facilities improve continuously. Sports Industry itself is an industry which utilizes its functions and its radical effects to create value, is to provide the community collection of economic activities in the same sports products and economic sectors combined for the society [1]. Therefore, mastering the spatial characteristics and the development mechanism of the sports industry is of great importance, can lay the foundation of the regional sports planning, regional sport guidance and the People's Fitness Plans.

Spatial analysis focuses on the spatial distribution and characteristics of the object, is the cause and effective means of dynamic mechanism. Based on the regional economics, spatial economics, the new geographical economics discipline and urban planning disciplines, Lu Linfei, mainly uses the related location theory to discuss the space layout and location characteristic of the Ten city of the center of Shanghai. Meanwhile, through the analysis of the degree of aggregation, distribution patterns and the factors that influence layout analysis of the downtown station Arena, he sums up

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 286–295, 2013.

their respective characteristics in space, and dives into the location pattern of the billiard hall in the dense and sparse areas and its dynamic mechanism [2]. The paper aims to study the spatial distribution situation of Wuhan Stadium, utilizes the demographic, economic, roads and other factors [3], to establish the appropriate model which can serves for the analysis of relevance and regression on the distribution of Wuhan Stadium. With the mode, the characteristics and development mechanisms of Wuhan Stadium are analyzed, and the spatial distribution of the future stadium is predicted.

2 Research Data

2.1 The Situation of Study Area

Wuhan City located in the center of China, is the capital, political, economic and cultural center of Hubei Province and also an important industrial city and economic center of the Yangtze River Region. Wuhan is also the cultural and educational center of China, and the major transportation hub. For the sake of positively responding to the call--“sports should go the social and industrial development road" [1], Wuhan City, through the wide publicity of multi-angle, multi-channel, multi-level wide, improves the people's awareness of fitness and guides people to participate in sports activities actively, adheres to the idea which sports social science and natural science stands together, and strives to create a high-level sports.

2.2 Research Data and Its Source

2.2.1 Spatial Data

Spatial data is divided into two parts: vector data and raster data. The raster data of Wuhan is downloaded from Google earth (2010 year); the vector data of Wuhan includes the punctate distribution of the sports facilities in Wuhan, vector data of Wuhan urban area, the road network of Wuhan urban area.

2.2.2 Non-Spatial Data

The study attributes data include: the population density of Wuhan, the total output value of production industry, and construction, the road density of first-class road, the road density of second-class road, the road density of third-class road, and the district density of highway. Among them, the population density of Wuhan, the total output value of production industry, and construction data is mainly obtained through sorting and extracting from Wuhan Statistical Yearbook (2010 year); and other data comes from the operation using the overlay function module in ArcGIS. The study data is as Fig.1 showed.

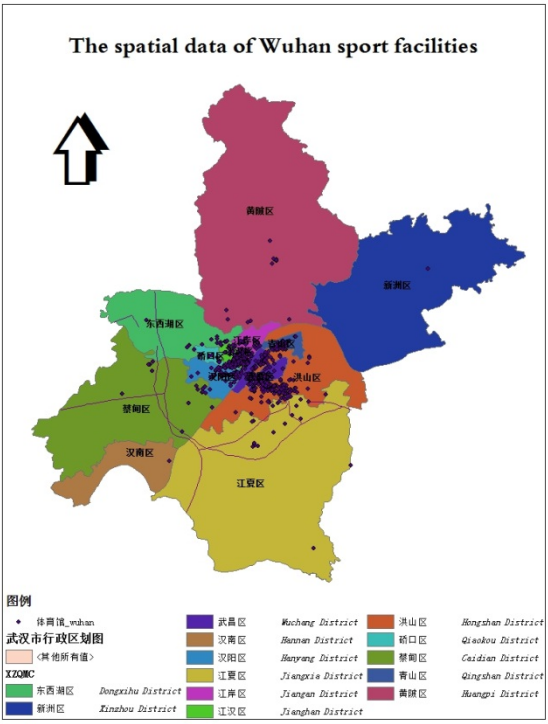


Fig. 1. The spatial data of Wuhan gymnasiums

3 Spatial Distribution Analysis of Wuhan Gymnasium Layout

3.1 Density Analysis in Spatial Distribution of Wuhan Gymnasium

The distribution density describes the spatial distribution of dots, lines and polygons, is the simplest and most commonly used method to describe the space distribution of dot pattern. It is the number of objects within one unit area, its molecular is the number of distributed objects, its denominator is the distribution area. We can get to know the sparse degree of density of Wuhan gyms through the spatial distribution of Wuhan gyms, so as to make further analysis and do related regulations.

Utilizing the rendering function in ArcGIS, the gyms are showed in five classes according to the distribution density [4]. Figure 2 is the distribution density figure of gym (13 districts), this simply can show the gym layout roughly, but the specific location of the gyms is unclear.

According to the above figure, we can see that the sports industry in the center of Wuhan distributes more densely, because these areas have a high population density, people live a high quality life and the transportation is well developed .for the sake of mastering dynamic mechanism of the distribution of gyms, it is of importance to make

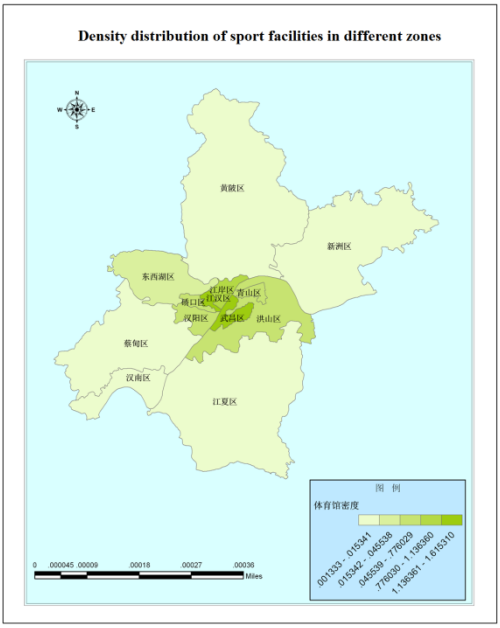


Fig. 2. Density distribution of gymnasiums in different zones

quantitative analysis on the factors which influence the distribution of gyms in the following study.

3.2 Spatial Autocorrelation Analysis

Use the technology of spatial autocorrelation to detect the gym spatial distribution pattern of Wuhan city circle. This paper uses the Moran's I index to measure the correlation's degree. Moran's I is the most commonly used global correlation index, used to detect the spatial distribution characteristics in the study area [5, 6]. Global Moran's I calculation formula is as follows:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}} \times \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{1}$$

In this formula, n stands for the number of the research object, x_i、x_j is the attribute values of the study area i and j respectively , $\bar{x}$ is all the attribute values mean, W_{ij} is the weight matrix measuring the spatial relationships among the study objects[7].

3.2.1 Autocorrelation Analysis of Gym Distribution in Wuhan City

Sum up the number of sports gyms within the grid in different scales [8] using ArcGIS, and then use Moran's I [9] of the space statistics module in ArcGIS to calculate the total autocorrelation index, the results are as follows:

Table 1. The results of Moran's I value

	Moran's Index	Z Score
500m grid	0.579083	118.101598
1000m grid	0.455736	64.373492
5000m grid	1.534497	48.529274
10000m grid	1.121208	26.755002

According to the size of the Z value, under the significant setting level, we can accept or reject the zero hypotheses. Take $\alpha = 0.05$,so when Z is less than -1.96 or greater than 1.96, reject zero hypothesis. We Can see from the table1, the Z value all is far greater than 1.96 and what's more the index of Moran's I values is greater than 0 and close to 1, showing that the overall distribution of the gyms is in positive spatial autocorrelation, in the agglomeration spatial pattern.

Under the four dimensions, Moran's I values are not the same, but all show the significant positive correlation. And the 5000 meters grid is strongest, while the grid of 1000 meters is weakest [10]. Given the above, we can come to a conclusion that different scales division influence the results of the auto-relation analysis. If the spatial relevance index is bigger in one scale, you can't obtain the same result in the other relevance. Therefore different size calculation should be made under different situations [11].

3.2.2 Distribution Autocorrelation Analysis in a Kilometer Grid of Stadium

Given the above analysis, we choose 1 km grid, and carry on the grid handling and spatial autocorrelation analysis for 13 areas respectively, the specific results are as follows:

Table 2. The spatial autocorrelation analysis after data being divided into grid

	Moran's Index	Z Score
Jiang Xia district	0.228090	11.671829
Cai Dian District	0.201864	8.612559
Han Yang District	0.446721	6.188522
Dong Xi Hu District	0.257819	6.836644
Qing Shan District	0.483644	4.792137
Hong Shan District	0.452225	13.470383
Jiang An district	0.398822	5.232148
Qiao Kou District	0.329358	3.388869

Table 2. (*Continued*)

Wu Chang District	0.280481	3.770263
Xin Zhou District	-0.001134	-0.030385
Huang Pi District.	0.050290	3.088442
Jiang Han District	0.438549	4.110656
Han Nan District	Lack of data	Lack of data

Take alpha as 0.01, when Z is less than - 2.33 greater than 2.33, reject zero hypothesis. According to the table2 : Jiang Xia, Cai Dian, Han Yang, Dong Xi Hu, Qing Shan, Hong Shan, Jiang An, Qiao Kou, Wu Chang, Huang Pi, Jiang Han, these 11 district's stadiums are in positive correlation, have the aggregation characteristics, and according to Z value judgment, they are in the significant positive correlation. While Xin Zhou District 'Moran I value is close to 0, Z value does not conform to the requirements of the significance, showing the random distribution. As to Han Nan district, because of a lack of data, Moran, I failed to calculate Moran I value and Moran Z values.

The comparison of correlation is as follows: Qing Shan District > Hong Shan District > Han Yang District > Jiang Han District >Jiang An district > Qiao Kou District > Dong Xi Hu District > Jiang Xia District > Cai Dian District > Huang Pi District. It shows again that Qing Shan District, Hong Shan District, Han Yang District and other downtown district are more obvious in the gathered characteristics of spatial distribution.

4 Correlation Analysis between Layout of Gymnasiums and Its Influential Factors

Lee (2001) proposed the measure L coefficient which combines the Pearson correlation coefficient and Moran's I autocorrelation coefficient, the coefficient has a good nature, is not only the Pearson correlation coefficient, and contains data space factors, has the advantage of Pearson coefficient and Moran's I coefficient, can quantify lag in 2D space, which can be used to measure spatial separation and difference. Specific formula is:

Supposing $\tilde{x}_i = \sum_{j=1}^{n} W_{ij} x_j$, $\tilde{y}_i = \sum_{j=1}^{n} W_{ij} y_j$, $\tilde{x}_i$ 、 $\tilde{y}_i$ stands for the weighted average of neighboring units , $\bar{x}$ 、 $\bar{y}$ stands for its vector's mean value, W_{ij} stands for the standardization of spatial weight matrix elements. Lee has gotten the binary variables of spatial correlation coefficient [12].

$$L_{X,Y} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}} \tag{2}$$

Based on the above analysis, this paper selects Lee's L parameter as the measuring index between the distribution of gyms and its possible factors; its significant test forms the foundation on the normal distribution. Make analysis on these variables: population density, gross domestic product, gross value of industrial output, construction industry output value, first class roads, second class roads, third class roads and highways, using the Pearson coefficient and Lee's L coefficient to calculate the bivariate correlation of the gym distribution with these variables respectively, so that reasons can be found to analyze and interpret the gym space distribution.

4.1 Pearson Correlation Analysis between the Gymnasiums Partition Distribution and Its Factors

The result is shown in figure 3.

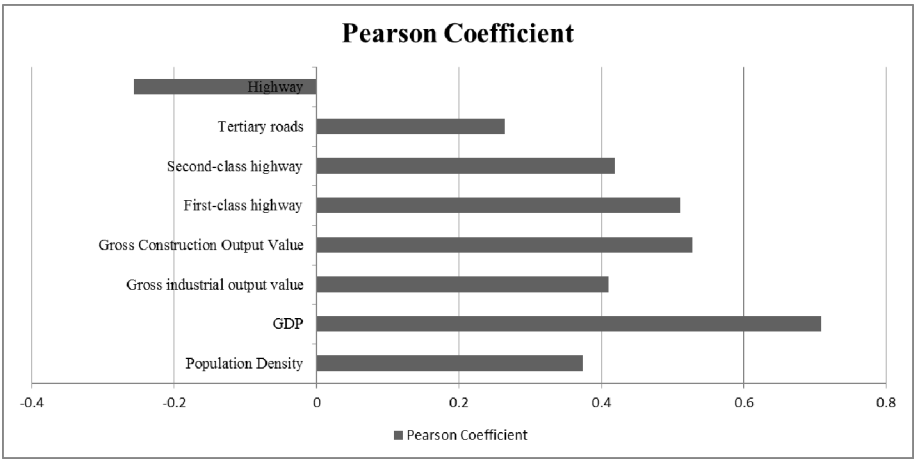


Fig. 3. Correlation of gymnasiums distribution and other factors

We can see from the table, GDP and the distribution of gyms are the most relevant, reflecting that the regional economic development and people's income level have played a positive role in promoting the development of the sports industry; And level of secondary roads also have high correlation, explains regional transportation infrastructure stock of wealth affects the demand for sports venues in the region; Among them, the highway and sports bureau's distribution have a high negative correlation; one possible reason is that the highway populated places often are in the suburbs or the countryside, which are not suitable for constructing and developing gyms.

4.2 Lee's L Correlation Analysis

Based on the coefficient of Lee' L, this paper study the spatial correlation between gym distribution and Population, GDP, Gross industrial output value, Gross Construction Output Value and road nets impact factors, and the result shown in table 3.

Table 3. Lee index of Wuhan gymnasiums distribution and other factors

Influential factors	Lee index	Z statistics	Correlation significant level
Population	0.4867*	0.4807	Significant correlation
GDP	0.7082**	1.4017	High linear correlation
Gross industrial output value	0.4084*	0.2396	Significant correlation
Gross Construction Output Value	0.5262*	0.6961	Significant correlation
First-class highway	0.4996*	0.5930	Significant correlation
Second-class highway	0.2024	-0.5586	Low linear correlation
Tertiary roads	0.2679	-0.3047	Low linear correlation
Highway	-0.2143	-2.1738	Low linear correlation

In the table 3, * indicates that the correlation is significant under the significance level of 0.05, and ** indicates that the correlation is significant under the significance level of 0.01 [13, 14].

From table 3, we can find that there exists strongest correlation between gymnasiums amount and GDP, and the correlation coefficient reach 0.7082. Reasonable explanation is that GDP is an important part of economic factors, which directly determines the resources and also the supply ability of gymnasiums in a region. So, the region with high economic level always has intensive distribution of gymnasiums. In addition to GDP, at significance level $\alpha = 0.05$, the factors which has significant spatial correlation with physical layout was as follows (from high to low in turn according to the significance level):

- Gross Construction Output Value
- First-class highway
- Population
- Gross industrial output value

It indicates that the surrounding buildings, population density and road conditions are all active factors which influencing the gymnasiums construction. The correlation of Second-class highway, Tertiary roads and gymnasiums is not significantly. And among this, the highway is negative correlation to gymnasiums distribution, but the specific reasons requiring further analysis.

5 Summary

Sport gymnasiums in Wuhan distribute intensively over central cities such as Jiang han, Wu chang, Hong shan and so on, presenting an inner to outer distribution pattern. This paper selected economic factors, social factors and traffic factors to analysis the driving

force for the distribution pattern. According to the results of spatial binary correlation analysis, we found that GDP is essential to the construction of gymnasiums, directly determining the economic supply of the preparation, maintenance and utilization of sport gymnasiums. Therefore sport gymnasiums gathered at the region with high economic level. Population density, the representative of social factors, has strong correlation with the amount of gym in Wuhan, for the reason that gymnasiums construction ultimately depended on the demand and small population density was difficult to form concentrated requirements for gymnasiums construction. Meanwhile traffic factors as the main factors influencing the economic development and population flow, indirectly affect the sports gymnasiums layout.

According to the spatial distribution of Wuhan sports gymnasiums and the dynamic mechanism behind it, this paper try to forecast the future distribution of gymnasiums in Wuhan city. With the continuous development and improvement of people's living standards in the surrounding area of central city, he demand for sports entertainment increases. Combined with the promotion for national fitness by government and the popularity of sports fitness facilities, the gymnasiums distribution pattern in Wuhan will transform from "Gathered in the central city and Spread outward" to "Concentrate in central city and Popularize in village". All in all, as the center of central china, the ancient capital of world civilization, international tourist city, Wuhan has a wealth of human resources and huge potential consumer groups. Consequently, Wuhan has great potential to establish the sports leisure entertainment market with local characteristics.

References

1. Chen, Y., Zeng, C.: The Status and Countermeasures of Leisure Sports in Wuhan. Hubei Sports Science 6, 628–630 (2011)
2. Lu, L.: Empirical Research on Spatial Distribution of the Billiard Halls in Shanghai Center Districts. Shanghai University of Sport (2010)
3. Yuan, Y.-J., Jin, Y.-H.: A Study on Influence Factors of Sports Facility Distribution in Beijing. Journal of Beijing Sport University 30, 1490–1492 (2007)
4. Zhu, Y., Zhang, J.: Analysis of Liaocheng road network based on ArcGIS density analysis. In: International Conference on Remote Sensing (2010)
5. Cliff, A.D., Ord, J.K.: Spatial processes: Models and applications. Pion, London (1981)
6. Liu, X., Huang, F., Wang, P.: Principle and Method of GIS spatial analysis. Science Press, Beijing (2005)
7. Wang, J.: Spatial Analysis. Science Press, Beijing (2006)
8. Sun, Q.-X., Li, M.-T., Lu, J.-X., Guo, D.-Z., Fang, T.: Scale Issue and Its Research Progress of Geospatial Data. Geography and Geo-Information Science 4, 57–60, 84 (2007)
9. Guo, R.: Spatial Analysis. Higher Education Press, Beijing (2001)
10. Xie, H., Liu, L., Li, B., Zhang, X.: Spatial autocorrelation analysis of multi-scale land-use changes: A case study in ongniud banner, inner Mongolia. Acta Geographica Sinica 4, 389–400 (2006)
11. Hu, Y., Wang, Q., Liu, Y., Li, J., Ren, W.: Index system and transferring methods to build the national society and economy grid database. Journal of Geo-Information Science 13, 573–578 (2011)

12. Lee, S.: Developing a bivariate spatial association measure: An integration of Pearson's r and Moran's I. Journal of Geographical Systems 3, 369–385 (2001)
13. Wei, Y.-C., Chen, S.-Z.: Principles and methods of geographic modeling. Science Press, Beijing (2007)
14. Han, D.-F., Wu, Q.-B.: Numerical computing methods. Zhejiang University Press, Zhejiang (2006)

A Cartographic Labeling Method in Chinese Characters

Boyan Cheng[1,2,*], Qiang Liu[3], and Xiaowen Li[4]

[1] School of Resources and Environment, University of Electronic Science and Technology of China,
No.2006, Xiyuan Avenue, West Hi-Tech Zone, Chengdu, Sichuan, China, 611731
myou_cheng@163.com
[2] 95007 Troops, Guangzhou, China, 510071
[3.] School of Resources and Environment, University of Electronic Science and Technology of China,
No.2006, Xiyuan Avenue, West Hi-Tech Zone, Chengdu, Sichuan, China, 611731
liuqiang_em@sina.com
[4.] School of Resources and Environment, University of Electronic Science and Technology of China,
No.2006, Xiyuan Avenue, West Hi-Tech Zone, Chengdu, Sichuan, China, 611731
lix@bnu.edu.cn

Abstract. To be understood correctly, the features of a map must be labeled properly. In this paper, rules for three kinds of cartographic labeling were analyzed, and a method of cartographic labeling for point, polyline, and area features in Chinese characters is designed and implemented efficiently. This paper implements two ways of point-feature labeling. One is that labels are placed at the point sites; the other is that labels are placed on one of eight positions relative to each corresponding point-feature. For polyline-feature labeling, an algorithm for searching for a number of discrete candidate positions to place the Chinese labels one character by one character along polyline-features is implemented. For a very long polyline-feature, the label is placed more than once. If an area-feature is sufficiently large, based on skeletonisation technique, the skeleton line of the area-feature, that is candidate support lines for labeling, is found inside the area-feature. Then the area-feature label is placed along the skeleton line, and covers most of the area-feature. Experiments showed that the dynamic labeling results are satisfactory.

Keywords: Cartographic labeling, Point, Polyline, Area.

1 Introduction

A map is a medium of communication, and the effectiveness of spatial information which it conveys depends on the quality that the features within an illustration are labeled to a large extent. A map should render the information of interest clearly,

* Corresponding author.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 296–303, 2013.

rapidly, and without ambiguity. Cartographic labeling is a convenient solution to transmission of map-information and makes map to be more legible and interpretable.

Cartographic labeling refers to the text insertion process on maps [1]. Cartographic labeling is an essential aspect of maps. It is necessary to label roads, landmarks, and other important features to help users understand their location and the environment. The labels should be easy to read and should follow basic cartographic rules [2]. Cartographic labels can tell us the names of the geo-graphical entities represented [3]. Finally, feature name labels serve also to enhance the aesthetic appearance of a map. It has been said that 'good form and placing of type makes a good map' [4].

Map features can be divided into three types, point-features, polyline-features and area-features. Therefore, there exist three different cartographic labeling issues: labeling of point features (such as cities, schools, hospitals), labeling of polyline features (rivers, highways, canals, railroads, streets, ship courses etc.), and labeling of area features (countries, states, oceans, lakes, bays etc.) [5].

2 Basic Cartographic Rules for Cartographic Labeling

To be understood correctly, the features of a map must be labeled properly. That is, a cartographic label should be placed onto or next to the features bearing the name of town, street, state, etc. represented. To communicate well, a cartographic label should make the spatial relationships among the features perceivable and understandable.

The labeling methods of point, polyline and area features are different. Some cartographic labeling rules can't be broken no matter what feature it is. The common rules include the following aspects [4]:

(1) Names should be easily read, easily discriminated, and easily and quickly located.

(2) The name and the object which it belongs to should be easily recognized.

(3) Avoid covering, overlapping, and concealment.

(4) Names should assist directly in revealing spatial situation, territorial extent, connections, importance, and differentiation of objects.

(5) Type arrangement should reflect the classification and hierarchy of objects on the map.

(6) Names should not be evenly dispersed over the map, nor should names be densely clustered.

In addition, according to the reading custom of the Chinese readers, Chinese labels may permit a left-to-right of texts and a top-to-bottom of texts.

2.1 Cartographic Rules of Point-Feature Labeling

Each point-feature label is always positioned close to the point object which it is associated with, and horizontal placement is generally preferred. It is convenient to identify 9 primary locations, as illustrated in Fig. 1. Usually, each point-feature label is preferably positioned to the right of a given point, while other positions in

connection to the feature can also be considered [6]. Location 9 (see Fig. 1 (c)), centered at the point coordinate, would be placed at the point location. Given a point-feature, a label can be placed on one of four positions relative to it (illustrated in Fig. 1 (a)), and also on one of eight positions relative to it (shown in Fig. 1 (a) and (b)). According to [7], it mainly considers the close relationship between the label and the point symbol and whether the label overlaps other labels or has conflicts with other cartographic objects.

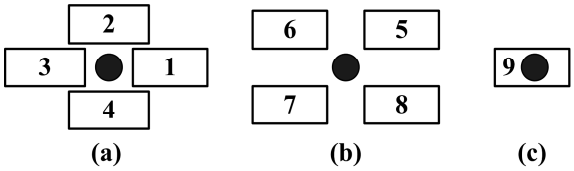


Fig. 1. Point-feature labeling. (a): 4 candidate positions for point-feature labeling (the best is location 1, and the worst is location 4). (a) and (b): 8 candidate positions for point-feature labeling (the best is location 1, and the worst is location 8). (c): A candidate position for point-feature labeling, centered at the point coordinate (location 9). The black dots are the point-features to be labeled.

2.2 Cartographic Rules of Polyline-Feature Labeling

A polyline-feature label is typically placed centrally along the corresponding polyline-feature, and is preferably horizontal. If the polyline-feature is curved, the polyline-feature label is made to conform to the polyline-feature's curvature. In addition, polyline-feature labels in Chinese characters may be spread out, as illustrated in Fig. 2. A label may be placed above the corresponding polyline-feature (shown in Fig. 2 (a)) or below the corresponding polyline-feature (see Fig. 2 (b)). The baselines of

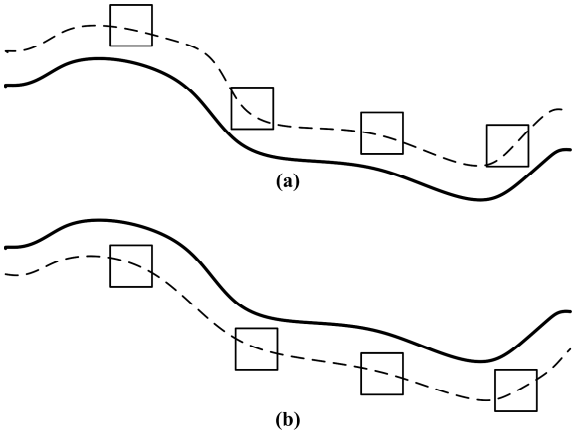


Fig. 2. Polyline-feature labeling. (a) The baseline is a top parallel line (dotted line); as well as (b) the baseline is a bottom parallel line (dotted line). The solid lines are the polyline-features to be labeled.

polyline-feature labels are the parallel lines that are a little distant from the polyline-feature. For a very long polyline-feature, the label may need to be placed repeatedly along the associated the polyline-feature more than once (12-15cm once) [8].

2.3 Cartographic Rules of Area-Feature Labeling

Because of the complexity of area-feature shapes, area-feature labeling is more difficult than point-feature labeling and polyline-feature labeling. There are varieties of area-feature labeling method such as labeling as point-features, labeling horizontally or vertically in area-features, labeling along the skeleton line of the area-feature and so on. The labeling of an area-feature is the same as a point-feature, if the area of the area-feature is smaller than that of one Chinese character. When the area of an area-feature is sufficiently large, the label should be placed inside, spread out and curved to conform to the shape and extent of the area-feature [9] [10]. This is illustrated in Fig. 3.

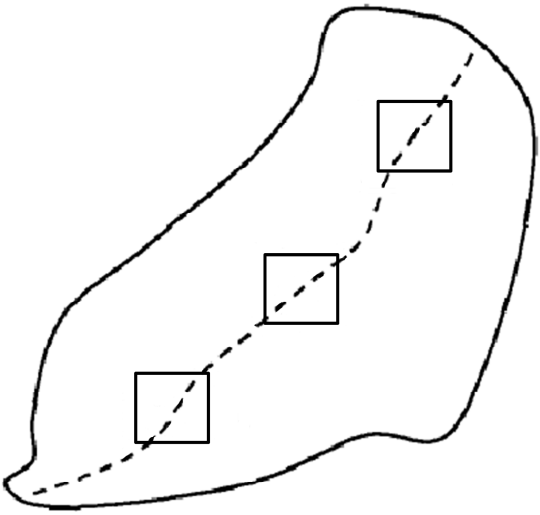


Fig. 3. Area-feature labeling. The dotted line is the skeleton line of the area-feature.

3 Method and Implement of Cartographic Labeling

A number of candidate positions for labeling one feature are defined. After this, these candidate positions are classified according to whether there is any cartographic object that might be hidden. The label should be placed on the most suitable location and then a next label could be conducted. If the position of a label would hide a suitable position of another label, the related labels should be adjusted to conform to rules of cartographic labeling.

3.1 Point-Feature Labeling

For point-feature labeling, this paper implements two ways of point-feature labeling. One is that labels are placed on the point locations (see Fig. 4 (a)); the other is that labels are placed on one of eight positions relative to each corresponding point-feature (shown in Fig. 4 (b)).

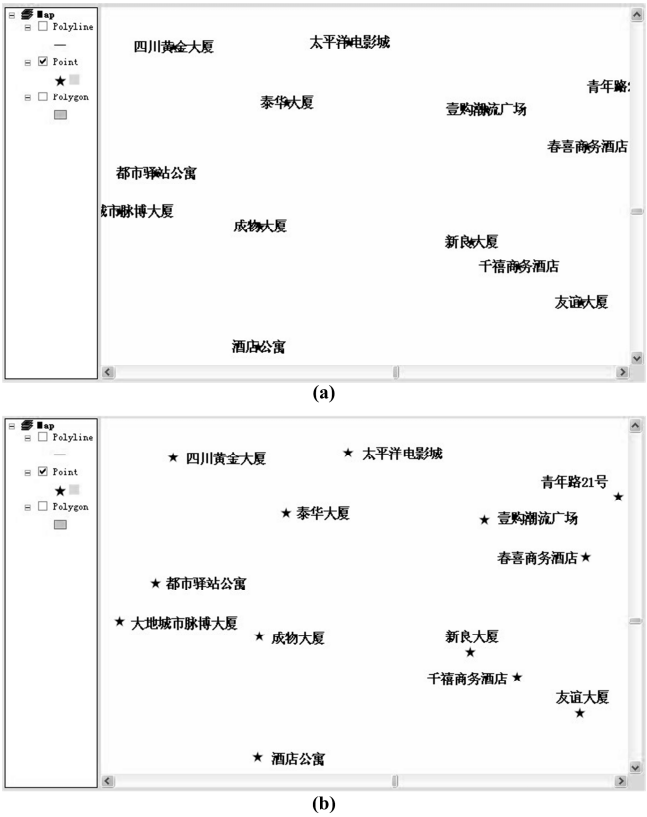


Fig. 4. Point-feature labeling. (a) Labeling at the point location; (b) labeling on one of eight positions relative to the point.

3.2 Polyline-Feature Labeling

For polyline-feature labeling, methods for either discrete or continuous search spaces have been studied. An algorithm for searching for a number of discrete candidate positions to place the Chinese labels one character by one character along polyline-features (illustrated in Fig. 5 (a)) is implemented in this paper. For a very long polyline-feature, the label is placed more than once (see Fig. 5 (b)).

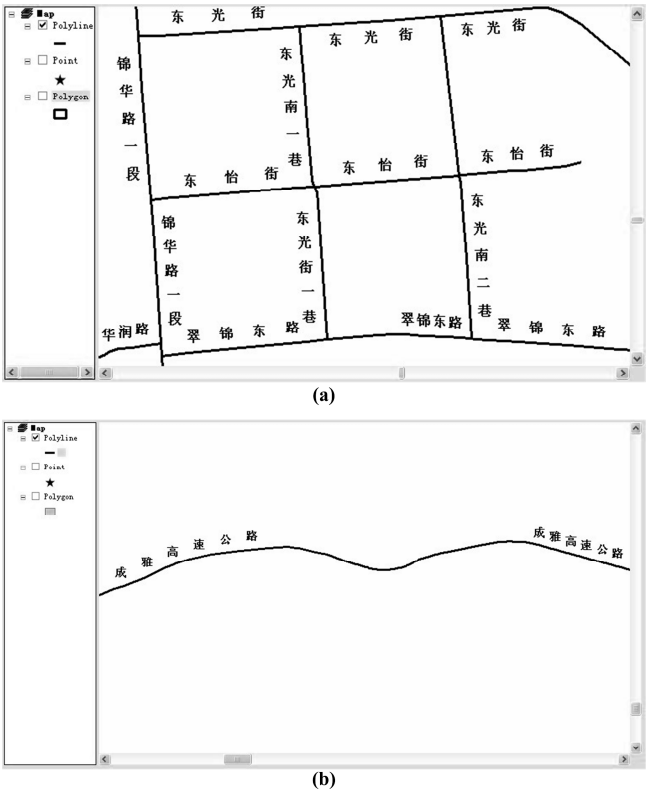


Fig. 5. (a) Polyline-feature labeling; (b) the polyline-feature label maybe need be placed more than once

In this paper, polyline-feature labels are placed above or below the polyline-feature, following the polyline-feature and conforming to the polyline-feature's curvature (illustrated in Fig. 5 (a)). The baselines of polyline-feature labels are the parallel lines that are a little distant from the polyline-feature. The parallel lines are divided into top parallel lines and bottom parallel lines if the polyline-feature is horizontal, and similarly, they are divided into right parallel lines and left parallel lines if the polyline-feature is vertical.

Given a polyline-feature, if the polyline-feature label need be placed above the polyline-feature, the top parallel line must be found. Then, discrete candidate positions are searched along the top parallel line. Finaly, each Chines character of the polyline-feature label is placed in the most suitable location in accordance with rules of polyline-feature label.

3.3 Area-Feature Labeling

If the area of an area-feature is sufficiently large, based on skeletonisation technique, the skeleton-line of the area-feature, that is candidate support lines for labeling, is

found inside the area-feature. Then the area-feature label is placed along the selected support line, and covers most part of the area-feature. This method is similar to that of polyline-feature labeling. This is illustrated in Fig. 6.

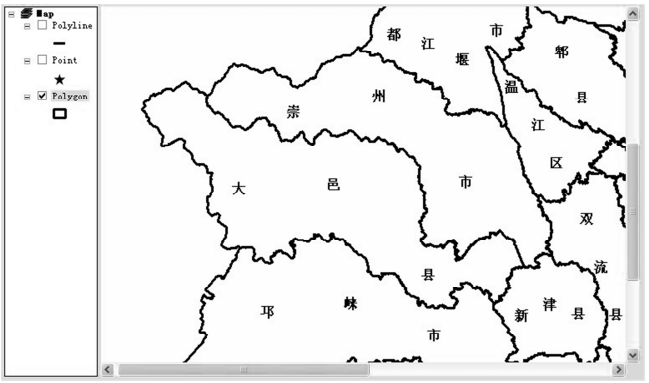


Fig. 6. Area-feature labeling

The algorithm of labeling along the skeleton-line of area-feature works as follows:

(1) Split the area-feature based on the theory of parallel line splitting.
(2) Get the center-points of the split-line inside the area-feature.
(3) The center-points are lined as the skeleton-line of the area-feature.
(4) Place a label along the skeleton-line by the label constraints on the area-feature.

4 Conclusion

The cartographic labeling is of importance for geoprocessing and automated cartography. With the increasing demand for electronic map on Internet, the automated cartographic labeling problem must be solved as quickly as possible. This paper presented the methods of cartographic labeling, and showed that the dynamic labeling results are satisfactory.

Acknowledgments. This research was partially funded by the Natural Science Foundation of China under Grant No. 41071222 to the University of Electronic Science and Technology, China. The authors are grateful to the anonymous reviewers whose comments helped to improve the paper.

References

1. Yamamoto, M., Camara, G., Lorena, L.A.N.: Tabu search heuristic for point-feature catrographic label placement. GeoInformatica 6(1), 77–90 (2002)
2. Zhang, Q.-N.: Real-time map labeling for mobile applications. Computers, Environment and Urban Systems 30(6), 773–783 (2006)

3. Ahn, J., Freeman, H.: A program for automatic name placement. Cartographica 21(2/3), 101–109 (1984)
4. Imhof, E.: Positioning names on maps. The American Cartographer 2(2), 128–144 (1975)
5. Christensen, J., Marks, J.: An empirical study of algorithms for point-feature label placement. ACM Transactions on Graphics 14(3), 203–232 (1995)
6. Du, W.: Point Feature Label Placement Research Based on Simulated-Annealing Algorithm. PhD dissertation. Wuhan University, Wuhan, Hubei, China (2005)
7. Roy, S., Goswami, P.P., Das, S., Nandy, S.C.: Optimal algorithm for a special point-labeling problem. Information Processing Letters 89, 91–98 (2004)
8. Yang, L.: Electronic map dynamic labeling research and implementation. PhD dissertation. China University of Geosciences. Wuhan, Hubei, China (2007)
9. Du, R.-Y., Liu, J.-N.: A research on automatic placement of Geo-Name in Area feature. Acta Geodaetica et Cartographica Sinica 28(4), 365–368 (1999)
10. Zhang, X., Li, L., Shu, Y., Wang, H.: Intelligent Automated Cartographic Text Placement of Area Features. Geomatics and Information Science of Wuhan University 33(7), 762–765 (2008)

Ecological Suitability Assessment of Construction Land Based on Spatial Information Technology – The Case Study of Xiaonan District

Hai Liu[1], Huiqiong Xia[1,*], Han Luo[1], Wenting Mo[1], Shenghua Yan[2], and Xiaoqiang Feng[3]

[1] Faculty of Resources and Environmental Science, Hubei University, Wuhan 430062, China
xhqiong2003@163.com
[2] Hubei institute of urban planning and design, Wuhan 430071, China
[3] Pingdingshan Municipal Bureau of Housing and Urban-Rural Development, Pingdingshan 467000, China

Abstract. Ecological suitability assessment for the construction land is to guide urban development on the basis of Health Sciences. Using GIS technology, from the evaluation index system build, the single factor ecological suitability assessment, evaluation factor weights confirmation, the ecological suitability comprehensive evaluation, ecological suitability partition and others aspects, the construction land ecological suitability assessment method was supplied. Xiaonan district was taken as an example and applied to assess the ecological suitability of its construction land. The method can provide reference method for other urban construction land to assess ecological suitability.

Keywords: Spatial information technology, Construction land, Ecological suitability, Xiaonan district.

1 Introduction

With the acceleration of urbanization and growing environmental problems, explore land suitability have become an important research topic for resolving the contradiction of urban development and ecological environmental development, and construction of ecological civilization [1]. Without loss of quality of the environment or circumstances, through a comprehensive evaluation of various factors to determine the ecological as urban and rural construction land suitability is defines as construction land ecological suitability assessment. Construction land ecological suitability is stronger, the pressure on the outside and changes is more sensitive, and regulate their own situation is faster. It is the core issue of urban ecological planning, and the important basis for urban master planning [2-3].

Many scholars made a variety of methods to study the land ecological suitability. Such as fuzzy comprehensive evaluation, ecological comprehensive assessment act, maps

* Corresponding author.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 304–312, 2013.

overlap method. However, these methods are mostly focused on one aspect. Therefore, the result is more one-sided, or data acquisition is difficulty, or the actual operability is poor [4-8]. In this paper, taking the Xiaonan district as the study area, where located in Wuhan city circle that is the comprehensive reform pilot building zone of resource-saving and environment-friendly society. A new method of urban construction land ecological suitability assessment is present based on space information technology. In this study, one hand, the method can provide reference method for other urban construction land to assess ecological suitability. On the other hand, it can provide support to Xiaonan urban planning.

2 Study Area and Technical Route

2.1 Study Area

Xiaonan located in northeastern part of Jianghan Plain, southern part of Dabie mountain and Tongbai mountain, and Wuhan with the north. The total land area is 1035km^2. It includes 13 townships, one development zone, and four district offices. It has good transportation of land,water and air, where it from Wuhan city center is about 50km, from Tianhe International Airport is 32km.Beijing-Zhuhai Expressway, 107 National Road, Beijing-Guangzhou railway running north and south. Han shi railway, Han shi highway, 316 national highway and railway across from northern to eastern, and inland shipping through the Yangtze River[9].

2.2 Technical Route

Firstly, combining with Xiaonan land use characteristics and the truth, the sensitivity factors that affect Xiaonan construction land ecological suitability assessment is extracted from the natural ecological factors, socio-economic factors and human factors extraction. Then evaluation index system is built, and the factors are quantified though corresponding mathematical theory, and sensitivity factor weights be determined. Finally, Xiaonan construction land ecological suitability is analysis based on space weighted superposition.

3 Establishment of Index System

3.1 Ecological Assessment Factor Selection

There are many factors that affect region ecological environment. Therefore, in allusion to topographic features, considering the status land use of Xiaonan situation, future development and construction goals, the nature of land use and other factors, some factors that affect Xiaonan ecological environment and ecological safety were selected. The natural ecological elements included ground elevation, vegetation and water. The socio-economic factors included transportation accessibility, and the town's location.

The human element included flood control and cultivated land protection. Those factors will be used to suitability evaluation.

3.2 Grading Standards of Single Factor Ecological Suitability

Single ecological suitability level is divided into five according to the actual situation, which is very suitable, suitable, basic suitable, basic unsuitable, unsuitable, while empowering values 1,3,5,7 and 9 respectively. Numerical size is inversely proportional to the size of ecological suitability, and is proportional to the size of ecological sensitivity.

Very suitable means the land can be sustained for a particular purpose without important limitations, and will not damage the ecological environment, reduce productivity or efficiency. It is generally empowered value of 1.

Suitable means the land be limited, and when there will be lasting for planning purposes moderate adverse, so that destruction of the ecological environment and reduce benefits. It is generally empowered value 3.

Basic suitable ranged between suitable and unsuitable. It is generally empowered value 5.

Basic unsuitable ranged between basic suitable and unsuitable. It is generally empowered value 7.

Unsuitable means it has a severe restriction, the continued use of certain uses to their impact is serious, it will seriously damage the ecological environment, and the use is basic reasonable. It is generally empowered value 9.

4 Ecological Suitability Analysis of Single Factor

4.1 Elevation

One of the key factors of the ecological environment heterogeneity is the elevation. Elevation changes are small in Xiaonan, and construction land elevation within the region is mainly concentrated in between 10m and 50m. Therefore, the elevation of 10m, 50m are used as the critical point of this single factor evaluation. Gradient grading between 10m to 50m are divided into 10m ~ 25m, 25m ~ 40m, 40m ~ 50m, its single-factor assessment scores are 1,3,5,7and 9. Generated single factor evaluation is shown in Figure 1.

4.2 Vegetation

Vegetation is an important natural resource that composed by forest, shrub and meadow community. Forest was considered as representative of ecologically sensitive factor in our study, because it's the major regulator of nature. According to the distribution of resources and woodland species in Xiaonan, forest has been divided into four categories: shrub land, forest land, other forest land and non-forest land, and

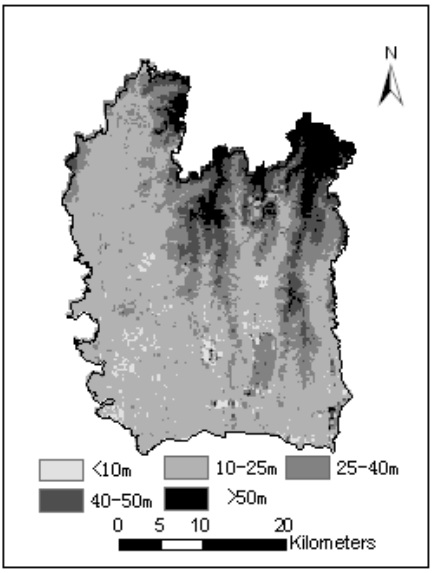


Fig. 1. Elevation

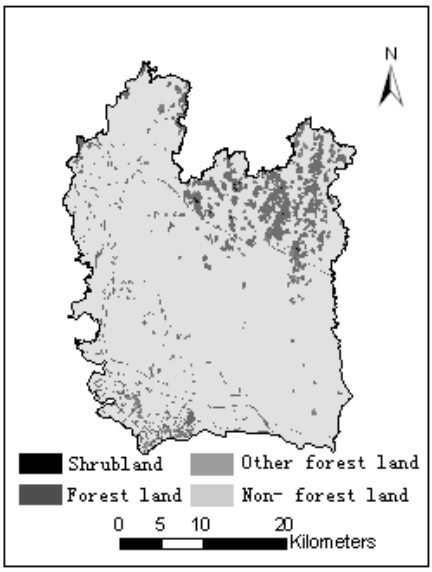


Fig. 2. Vegetation

is empowered 9,7,5,1 to characterize the ecologically sensitive degree of relative strength. Vegetation evaluation is shown in Figure 2.

4.3 Water

Surface water plays an important role in improving the cities and towns landscape quality, space environment and other aspects. In this study, surface water ecological suitability divided into major rivers, lakes, ponds and its extension 100 m buffer zone, 100m to 500m, 500m to 1000m and a range of more than 1000m buffer zone, and were assigned to 9,5,3and 1. Water evaluation is shown in Figure 3.

4.4 Transportation Accessibility

According to the basic theory and practical experience of geography, generally, urban layout has the characteristics that layout to the convenient transportation and high accessibility region. Traffic in Xiaonan developed. The expressways, national roads, highway, provincial road and other roads cross the town. The value of different levels road and the elevation is shown Table 1 and Figure 4.

Table 1. Assessment factor weights

Road level	Expressway	National road, highway	Provincial road	County road	Other road
Value	9	7	5	3	1

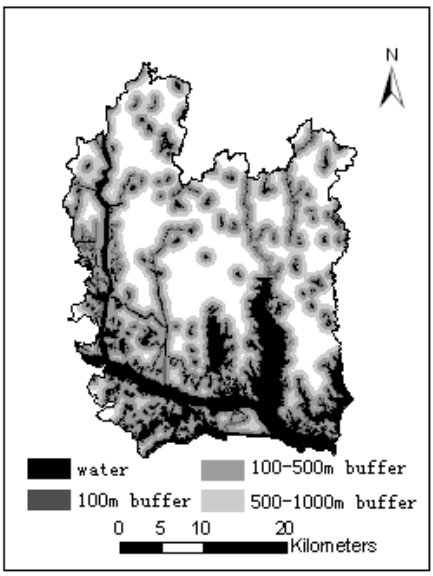


Fig. 3. Water

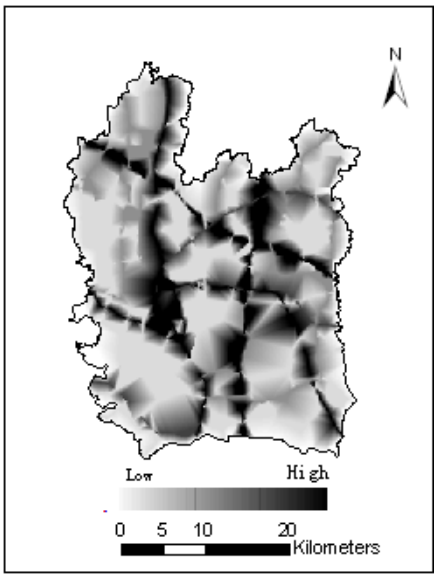


Fig. 4. Transportation accessibility

4.5 Location

Town is the area that economic, technological, capital and population are most concentrated. The development of new town must first consider the location relationship with the original towns. The region that has good geographical location with the existing urban can often give priority to develop. Based on the level of Xiaonan towns, it is empowered value of 9,7,5,3 and 1. Location evaluation is shown in Figure 5.

4.6 Flood Control

Flood control related to national and social stability. Impact on flood safety factor, the distance of the river 100 meters, 200 meters, 300 meters, and more than 300 meters of the dam, it were divided into the ban foot land, engineering retained land, security of protected areas and other areas. And based on their importance extent that its assignment as 7, 5, 3,1. Flood control evaluation is shown in Figure 6.

4.7 Cultivated Land Protection

The speed of soil loss that results by the unreasonable in agricultural production can faster than update. It can cause the superior natural environment to damage with changes in vegetation destruction. In this study, based on recovery of damaged farmland, the cultivated land was divided into basic farmland, cultivated and non-cultivated. It's assigned as 7,5,3,1 to characterize the relative strength degree of ecologically sensitive. Cultivated land protection evaluation is shown in Figure 7.

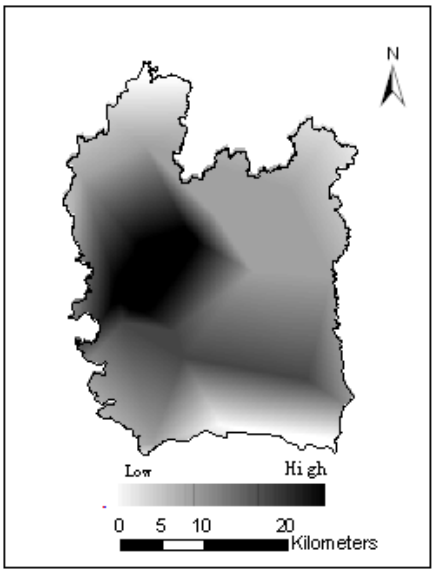


Fig. 5. Location

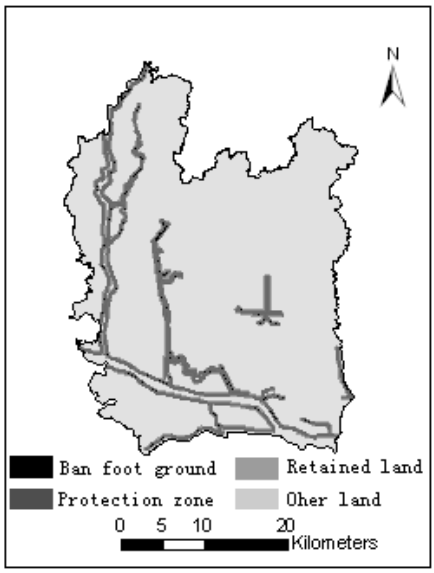


Fig. 6. Flood control

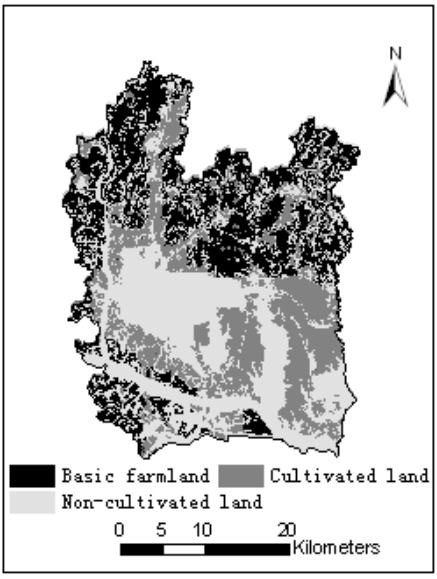


Fig. 7. Cultivated land protection

5 Assessment Factor Weights Determination

The data sources is more broadly of construction land ecological suitability assessment, and the methods of evaluation factors quantitative approach are differences due to the

different data. In determining the evaluation factor weights, if the pure mathematical methods be used, the weighting will be impact greatly by the quantitative analysis, and its objectivity and practicality is limited. In the present study, using the Delphi method, it cannot be impacted by pre -processing method in determining the weight process, and with a wealth experience and knowledge of expert. Therefore, the weight is more objectivity and practicality [10-11]. The weight values as follows:

Table 2. Assessment factor weights

Assessment factor types	Assessment factor name	Weight
Ecological factors	Ground elevation	0.15
	Vegetation	0.15
	Water	0.20
Socio-economic factors	Transportation accessibility	0.15
	The town's location.	0.15
Human factors	Flood control	0.10
	Cultivated land protection	0.10

6 Construction Land Ecological Suitability Comprehensive Assessment

Based on the above analysis and evaluation, the use of GIS technology in spatial overlay analysis, each individual participating element is analysis separately by weight overlay. Each individual participating element was normalized before the weighted overlay analysis, so that each individual element values were between 0 and 1. Treatment formula is as follows:

Is positively correlated with the construction land ecological suitability：

$$Yi=Xi/(\max(xi)) \tag{1}$$

Is negatively correlated with the construction land ecological suitability：

$$Yi=1-Xi/(\max(xi)) \tag{2}$$

Yi：Each individual contestant element normalized values

Xi：Each individual contestant element original values

There are five single factors with a positive correlation about seven single factors that are used to Xiaonan construction land ecological suitability. Those factors are ground elevation, vegetation, water, flood control and cultivated land protection. There are two factors with a negative correlation. Those factors are transport accessibility and location.

After a treatment with the regulations of the individual factors, combined with the weight of each factor, Xiaonan construction land ecological suitability was evaluated by the weighted superposition. The map is shown in Figure 8.

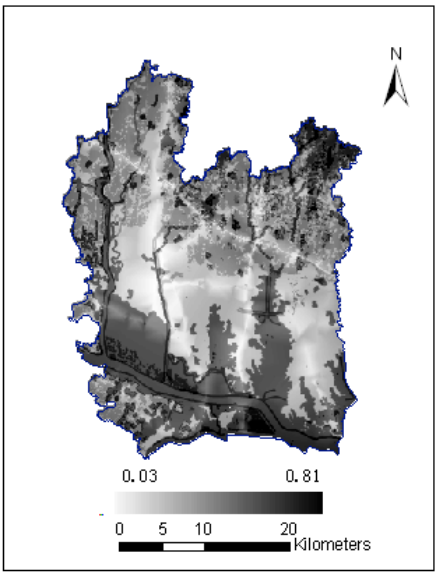


Fig. 8. Ecological suitability assessment of construction land

Be seen from the above chart, The area suitable for building mainly distribute in urban areas, and the surrounding the urban areas where locate in Midwestern region, and Xiaonan airport economic zone where locate in southeast region, and high-speed Highway, State Road surrounding areas. These regional has good traffic and location conditions, and the ecological sensitivity is poor. Therefore, it not only conducive to the city's construction, but also reduce the damage to the ecological environment construction. The ecological suitability of poor areas mainly locates in the southern and western where distribute many lakes and rivers, and locates in the northeastern region that covers dense woodlands. These regional has good natural ecological environment and strong ecological sensitivity, and it is the key areas to maintain the healthy development of regional ecosystem. So these areas should be minimized in the construction and enhanced ecological protection.

7 Summary

Based on Spatial Information Technology, the new methods of construction land ecological suitability evaluation was proposed. Firstly, construction land ecological suitability evaluation system was built. Then, quantitative criteria and weights of the various indicators were determined. Finally, through space weighted stacking, the suitability score of the study was obtained. The new method of construction land suitability evaluation can provides reference for others city.

The ecological suitability situation of Xiaonan construction land was evaluated by the proposed method. The suitable area of Xiaonan mainly locates central and southeast regions where has good traffic and location conditions. The unsuitable area

of Xiaonan mainly locates in southern, western and north-eastern regions. Those areas have fine natural environment. The evaluation can provide technical support for the city's development.

Acknowledgements. This research was supported by an Open Research Fund (No.(11)key03) of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (Wuhan University).

References

1. Guo, Y.-T., Liao, H.-P., Xu, J.-G.: Evaluation on suitability of rural residential land in Three Gorges Reservoir region. Transactions of the Chinese Society of Agricultural Engineering 28(5), 252–259 (2012)
2. Zhou, J.-F., Zeng, G.-M., Huang, G.-H., et al.: The ecological suitability evaluation on urban expansion land based on uncertainties. Acta Ecological Sinice 27(2), 774–783 (2007)
3. Liu, J., Ye, J., Yang, W., et al.: Environmental impact assessment of land use planning in Wuhan city based on ecological suitability analysis. In: International Conference on Ecological Informatics and Ecosystem Conservation, vol. 2, pp. 185–191 (2010)
4. Fan, C.-J., Shen, S.-G., Wang, S.-H., et al.: Research on urban land ecological suitability evaluation based on gravity-resistance model: A case of Deyang city in China. In: International Conference on Green Buildings and Sustainable Cities, vol. 21, pp. P676–P685 (2011)
5. Liang, T., Cai, C.-X., Liu, M., et al.: Study on methodology of ecological suitability assessment of urban landuse: An example of Pingxiang. Geographical Research 26(4), 782–788 (2007)
6. Malczewski, J.: Gis based land-use suitability analysis: A critical overview. Prog. Plan. 62, 3–65 (2004)
7. Xu, K., Kong, C.-F., Li, J.-F., et al.: Suitability evaluation of urban construction land based on geo-environmental factors of Hangzhou. Computers & Geosciences 37(8), 992–1002 (2011)
8. Liu, X.-F., Shu, J.-M., Zhang, L.: Research on applying minimal cumulative resistance model in urban land ecological suitability assessment: As an example of Xiamen city. Acta Ecological Sinice 30(2), 421–428 (2010)
9. Yun, J., Yang, G.-Q., Zhu, J.-B.: Driving mechanism of rural residential area change-based on investigation into farmers in Xiao nan district of Hubei province. Economical Geography 28(6), 991–994 (2008)
10. Lin, S., Ren, Z.-P.: Fuzzy Delphi Method and Its Application. Forum on Science and Technology in China 5, 102–103 (2009)
11. Tian, J., Zhang, P.-Z., Wang, K.-L., et al.: The Integrating Model of Expert's Opinion Based on Delphi Method. Systems Engineering-Theory & Practice 1, 57–62 (2004)

The Design of Water Resources and Hydropower Cloud GIS Platform Based on Big Data

XiChun Wang[1,2] and Zhiyu Sun[2]

[1] Changjiang Spatial Information Technology Engineering Co., Ltd,
Changjiang Institute of Survey Planning Design and Research, Wuhan, 430010
wangxc73@sina.com
[2] Department of S & T and Environment Protection,
China Three Gorges Corporation, Beijing 100038

Abstract. The paper introduces the definition and the characteristic of big data. The paper also introduces the internal and external situation, the development tendency of big data. Based on the service-oriented architecture, GIS and cloud computation, the paper issues the water resources and hydropower cloud GIS platform. The purpose of the platform is to manage the various and massive data efficiently based on the construction of the big data fundamental framework of survey, design, construction, environment, immigrant and equipment and supplies. The information provided by the advanced analysis method such as predictive analysis, data mining, statistic analysis SQL, data visualization, etc can be acquired in the platform. After the construction of the well-rounded, decision-intelligently and universal service cloud GIS platform of water resources and hydropower based on big data, the decision support for the owner, the designer and the construction company in the process of planning, design, construction, operation, overhaul and maintenance can be provided efficiently. The paper elaborates the water resources and hydropower big data and the application. The paper discusses some key technologies such as integrating traditional data warehouse tools technology and real-time data processing technology and so on. In the end of the article, the paper discusses the prospective application of big data in water resources and hydropower projects.

Keywords: big data, water resources and hydropower, Cloud GIS Platform.

1 Introduction

There is an old adage in USA, "In God we trust, all others bring data" [1].It is obviously that we live in an information age based on data, Lots of commercial opportunity will be created by the data playing the main role in the new information war. Our daily life is changed by the multi-resources information. The data is new petroleum. Human civilization trend and our cognition about the world and life will be changed and rebuilt by the big data. In the past years, we could not talk with each other face to face without

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 313–322, 2013.

Twitter, and had to go shopping in the real street. Nowadays with the popularization of social network, stream media and broad band commercial service, all the necessities of life can be provided and all the real-time information can be obtained remain within the door. According to the IDC [2], the processing of big data is the necessary ability in 2012 and more than 85% enterprises of top 500 will lose advantage in big data competition in 2015.

2 Brief Introduce about Big Data

2.1 Definition about Big Data

From Wikipedia, the free encyclopedia, in information technology, big data is a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The challenges include capture, storage, search, sharing, analysis, and visualization. The trend to larger data sets is due to the additional information derivable from analysis of a single large set of related data, as compared to separate smaller sets with the same total amount of data, allowing correlations to be found to "spot business trends, determine quality of research, prevent diseases, link legal citations, combat crime, and determine real-time roadway traffic conditions."

Big data is closely and inseparably to cloud computing. They develop side by side through mutual cross-fertilization. Cloud computing should satisfy the capture of data and on the other hand, the big data is the promotion of the data analysis ability. The tremendous data storage and computing ability is the mainstay of big data and cloud computing. The integration of the computing resources to realize the scale effect is the inevitable outcome of the information industry development. Those enterprises without data center can save cost efficiently through the utilization of the cloud computing resources to store and compute the big data. Those enterprises with large scale data center can enhance the utilization of the data center through the sale of the storage and computing resources.

2.2 The Characteristic of Big Data

Normally the four characteristics of big data are described as, Volume、Variety、Value、Velocity [3, 4].

(1) The immense data volume (Volume). The size of data is about ten times as big as before, from the level of TB to PB and even to ZB. In the real world, millions of data capture sensors are plugged in all kinds of instruments. In the digital world, the daily life of consumer (communicate, browse, shopping, share, search in the internet) produces enormous data. The research of IDC indicates that there are 1.8 billion GB data in the digital field and the size is increasing 55% per year.

(2) The various data type(Variety). The data can be classified as structured data, half-structured data and non-structured data. The 80% of the data is non-structured data such as sound, video, picture and the position information, is. The difficulty of

extracting the valuable information from the data is raising as the percentage of non-structured data increases gradually.

(3) The low density of value(Value). The density of value tends to vary inversely with the data size. Take video as an example, the valuable data in an one-hour-long video is only one or two seconds.

(4) The high velocity(Velocity). This is the marked feature of big data distinguish from the traditional data mining. The value of data is directly proportional not only to the data size, but also to the data processing period. That is, the more promptly the data been dealt with, the more valuable it is.

2.3 The Status and Trends of Big Data at Home and Abroad

On Dec 8th, 2011, the data processing technology is proposed as one of the 4th key technology innovation projects in the 12th five years planning issued by the Ministry of Industry and Information Technology of PRC. The big data consists of the technology such as data storage, data mining, intelligent analysis of video images. The big data is also closed related to the other three key technology innovation projects consisting of information apperceiving, information transmission and information security.

In 2012, the Obama administration announced the Big Data Research and Development Initiative, which explored how big data could be used to address important problems facing the government. The initiative was composed of 84 different big data programs spread across six departments. The initiative planned to spend more than 200 million dollars. The Department of Energy tried to spend 25 million dollars to build the extensible institute of data management, analysis and visualization. The U.S. Geological Survey (USGS) provide places and time to make deep analysis, the most advanced computing ability and comprehend data set cooperative tool, for the scientists to catalyze the innovation of geographic system science.

The big data brings advantage for IT market. The data of IDC predicts that the scale of big data market will grow from 3.2 billion dollars of year 2010 to 16.9 billion dollars of year 2015. The additional size of big data in year 2020 will be about 45 times more than in year 2009. Only 15% of these data is traditional structured data. About 85% of the data is half-structured data and non-structured data coming from the domain of the enterprise interior information activity, electronic commerce, or external social network. The big data begins to cut a figure in the prediction of social economic development. Google surpassed the estate economists in the precise forecast of the trend of housing market according to the variety of housing search activity. Recently the international magnates such as IBM, Oracle, EMC, SAP pioneered to purchase the data management and analysis companies, and the sum of purchase is over 40 billion dollars. As the world biggest company of search engine and internet market, Google earned over 37.9 billion dollars revenue, 96% of which came from advertisement in 2011. In March 2012, the daily search activity proceed by Google was over 3 billion dollars. And 34 thousand questions were answered per second. Google managed to create the most valuable information for the user at the first time, and ascribed all these achievement to the big data computing.

3 The Big Data and Water Resources and Hydropower Cloud Gis Platform

3.1 The Framework of Water Resources and Hydropower Cloud Gis Platform

As the PC era been taken placed by the cloud computing era, the innovation of information technology will bring new industrial revolution, the cloud computing has obvious advantage in the big data processing, massive computing, reducing the system equipment and system maintenance. Confronting all kinds of water resources and hydropower application requirement and big data, a normal application service platform could not satisfy the demand of water resources and hydropower GIS in an all-round way. A powerful basic structure to support all categories of application service is demanding. The structure can be seen as the standard application service provided in the network and can supply the method of basic storage and computing. In order to meet all the demands of water resources and hydropower GIS, it is significant to build the modern water resources and hydropower cloud GIS platform based on the service-oriented architecture and study big data and other the relation key technology.

The framework of water resources and hydropower cloud GIS platform is illustrated in figure 1. From the bottom up, the architecture consists of physical layer, virtual layer, data layer, big data supporting platform and service component layer, service layer, application layer and the services crossing the layers such as service discovery, service monitor, resources allocation and statistic. Here is the detail function of six layers.

(l) Physical layer is the lowest layer of the framework, consists of physical equipment such as computer hardware and switch.

(2) Virtual layer consists of kernel of operation system, virtual machine and virtualization tool.

(3) Data layer consists of database server and file server which store all kinds of thematic data and the index to support the real application. Those thematic data consists of fundamental geographic information data etc, environment data, project management data and immigrant management data. The accession of database server can be accomplished either through middleware such as SDE, or through direct interface. Distribute storage is adopted for massive data.

(4) Big data supporting platform and service component layer consists of service component to fulfill service or function such as powerful computing, virtual realization, spatial processing The supporting platform of the layer realize high-performance computing through distribute computing. The layer contains running environment supporting GIS, reflecting the spatial feature of cloud GIS platform. The service component reflects connotation of service in the function of service and quality of service (QOS). It complies with the criterion of SCA (service component architecture) and SDO (service data object).

(5) Service layer consists of all the service defined in the service component layer. The water resources and hydropower cloud GIS platform register all the geographic resources of hydropower in the web and provide the web service network. The service provided by the platform consists of resources register service, spatial data service,

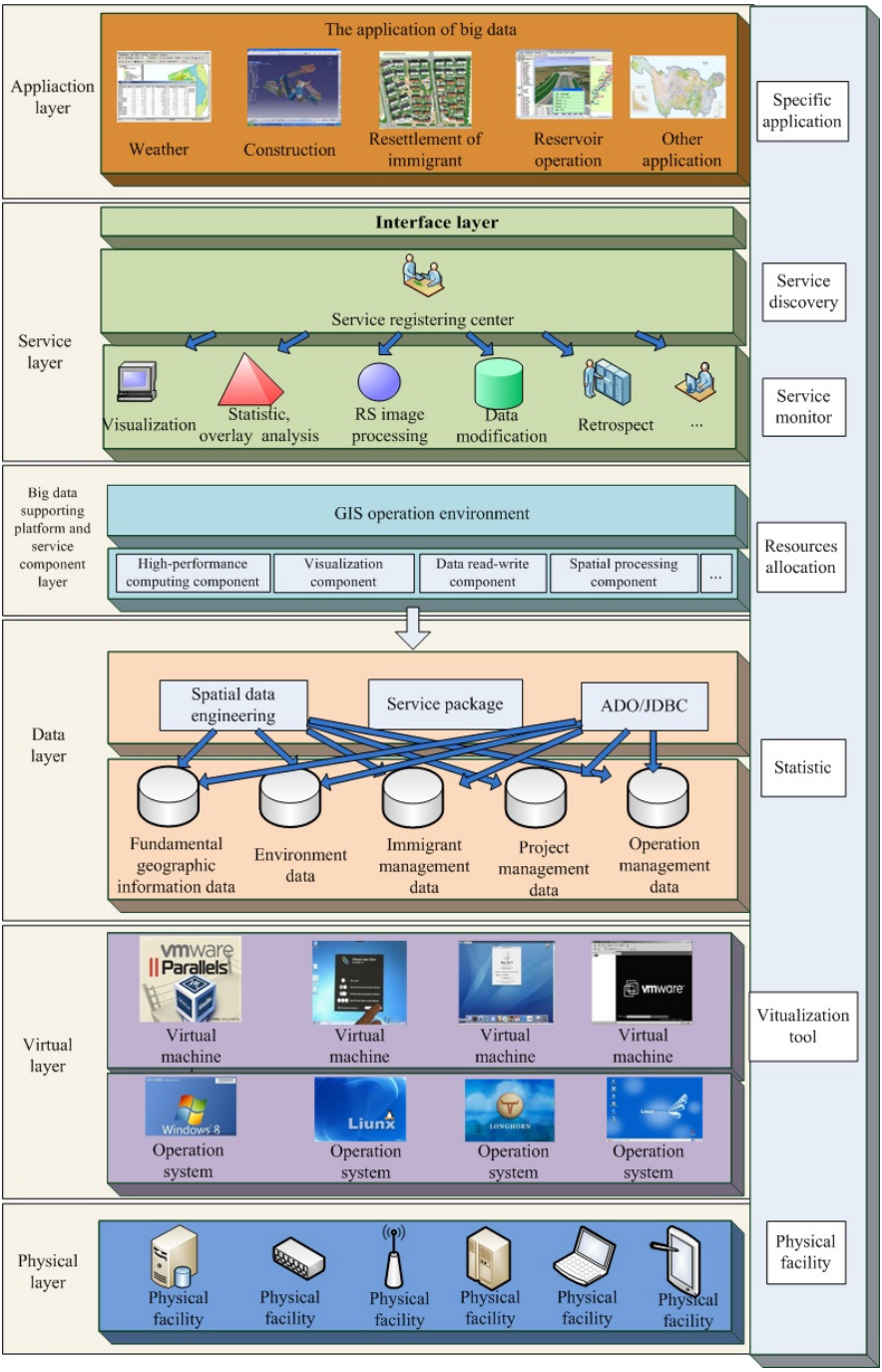


Fig. 1. The framework of water resources and hydropower cloud GIS platform

spatial information service etc. All these services can be discovered, invoked and arranged to provide the combination service. The Service discovery means supporting all the services to distribute its own functional metadata, to discover other service and to acquire the address information of the other service instance. The Service monitor means providing the information of node status of required service based on the service discovery and ensure the service is available. The service allocation means allocating all the services reasonably according to the service node status based on the service monitor. With the above services and all the network resources, computing resources and storage resources provide by the cloud GIS platform network, all kinds of the user such as the owner, the designer, the local governor and the construction company can obtain all categories of hydropower spatial data resources according to their own different requirement. The platform can provide all kinds of service handled by professional software and transfer the spatial data resources into spatial information resources. All the services realize the access function of platform through the programmable service interface defined by the service layer. Some interface of the specific operation component and the specific item is declared in form of description of the service.

(6)Application layer is the detailed GIS application which on the top of the framework. The lifecycle cover all the stages of water resources and hydropower such as survey and planning, project construction and operation management etc. The construction of hydroelectric station, the resettlement of immigrant, the hydrologic sediment, the reservoir operation and power generation is good example of various GIS application. All the application, including detailed operation process and user client, is implemented through service interface provided by service layer. In detail, operation process encapsulate the specific water resources and hydropower operation in the service component layer, assemble the operation process dynamically through the technology such as service arrangement and assemble, service chain, work flow etc. The operation process is implemented through the service interface provided in service layer. The user clients provide the graphical and non-graphical interface for the user to interact with the various services.

3.2 The Water Resources and Hydropower Big Data and Its Application

The water resources and hydropower cloud GIS platform is build to cover all the lifecycle of water resources and hydropower project such as survey and planning, design, construction and operation management. The digital management covering all the procedure such as survey, planning, design, bidding, purchase, construction management, immigrant management, security monitor, facilities maintenance, archive management and operation management can be implemented in the platform. The purpose of the platform is to manage the various and massive data efficiently based on the construction of the big data fundamental framework of survey, design, construction, environment, immigrant and equipment and supplies. The information provided by the advanced analysis method such as predictive analysis, data mining, statistic analysis SQL, data visualization, artificial intelligence, and natural language processing can be acquired in the platform. After the construction of the well-rounded, decision-intelligently and universal

service cloud GIS platform of water resources and hydropower based on big data, the decision support for the owner, the designer and the construction company in the process of planning, design, construction, operation, overhaul and maintenance can be provided efficiently. The application model of water resources and hydropower big data is illustrated in figure 2.

In the case of fundamental geographic information, with the integrated management of the fundamental geographic information of water resources and hydropower project along the basin, such as the Digital Orthophoto Map (DOM), the Digital Elevation Model(DEM), the Digital Raster Graphic (DRG), the Digital Line Graphic (DLG), the fundamental geographic information is provided for the application such as water resources and hydropower project management, immigrant management and operation management, after the construction of the fundamental geographic information sharing platform.

In the case of environmental issues, based on the fundamental geographic sharing platform, the decision support service is supplied for the disaster prevention and

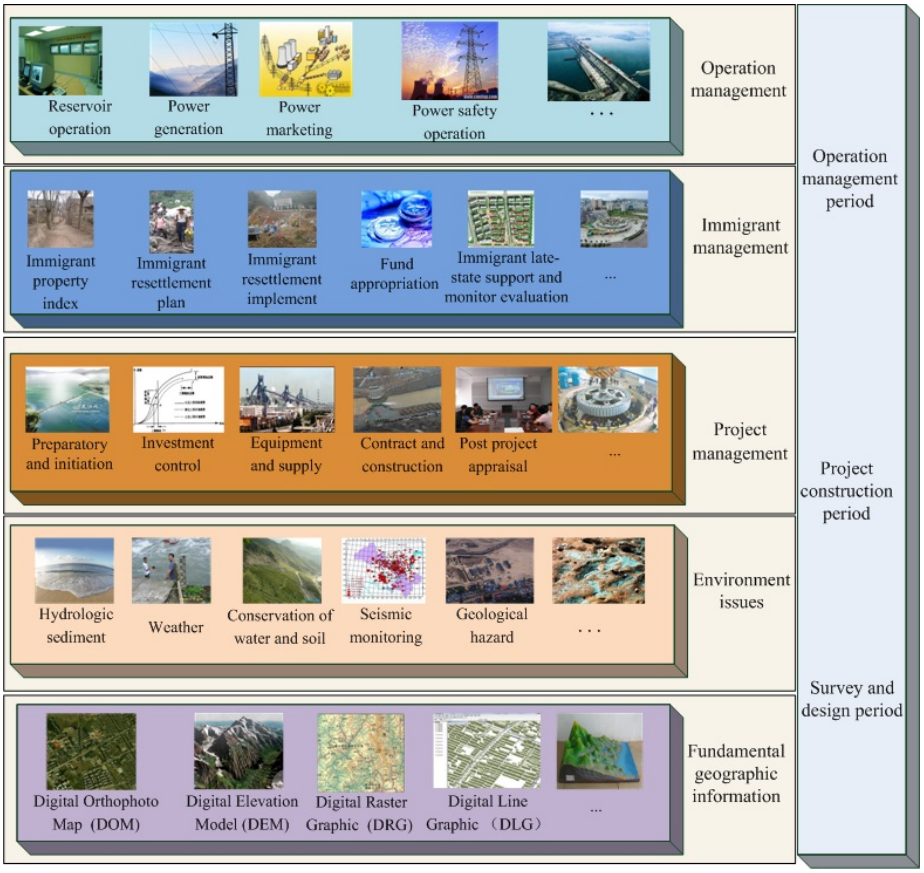


Fig. 2. The application model of water resources and hydropower big data

environmental management with the integration management of environmental information of all hydroelectric station, reservoir and basin such as hydrologic sediment, weather, conservation of water and soil, seismic monitoring. After the construction of hydrologic sediment subsystem, weather subsystem, water and soil conservation subsystem, geological hazard management subsystem, environment monitoring subsystem and seismic monitoring subsystem, the value of big data is appreciated by the user completely with the advanced analysis method such as natural language processing and supporting analysis to make decision with the environment data and feedback rapidly, reduce the cost of management with the query analysis thoroughly.

In the case of project management, after the construction of multi-project information management platform of the basin, all the information such as initiation, investment control, equipment and supply, contract and construction, post project appraisal are managed integrally. The decision supporting service of the elaborate management and control of multi-project of the basin is provided with the platform. After the construction of preparatory and initiation management subsystem, contract and construction subsystem, equipment and supply management subsystem, quality and security management subsystem, schedule planning management subsystem and financial management subsystem, the all-rounded progress management of design, planning, construction and quality control is implemented with the query and analysis, the feedback and demonstration of the various and massive data. The scientific forecast and early warning along the progress is carried out with the digital imitation and analysis computing reflecting construction progress.

In the case of immigrant management, after the construction of immigrant management platform, all the information such as the index of immigrant property, immigrant resettlement planning, immigrant resettlement implement, fund appropriation, immigrant late-state support and monitor evaluation etc, is integrally managed to provided decision supporting service for the prosperity and stability of immigrant. The index of immigrant property can be checked, the fund of immigrant appropriation can be traced, the effect of immigrant resettlement can be evaluated with the construction of immigrant resettlement planning result management subsystem, the immigrant resettlement implement management subsystem, the immigrant fund appropriation management subsystem, the immigrant late-stage support management subsystem and the management subsystem of the index of immigrant property.

In the case of operation management, after the construction of the operation management platform, all the information such as reservoir union-operation, power generation, power marketing and power safety operation, is integrally managed to provide decision supporting service for the intelligent operation. The powerful ability of data mining, data analysis and decision of cloud GIS platform is implemented after the construction of reservoir union-operation subsystem, the power generation subsystem, the power marketing subsystem and the power safety operation subsystem, with the production data and operation management data provided by the technology such as cloud computing, new energy incorporation and mobile internet etc. The elaborate operation management of enterprise, the dynamic optimized operation of the basin, the maximum of comprehensive benefit of the cascade reservoirs, the provision

the high value-added service, and the innovation of management mode is implemented with the platform.

3.3 The Key Technology

Considering the architecture layer model and the application-oriented characteristic of the cloud GIS platform of water resources and hydropower, the efficient integration, the efficiency of operation supporting and the intelligence of big data is the confronting challenge. There is lots of key technology related to the cloud GIS platform of water resources and hydropower and only some of them are discussed in the paper.

(1) The integration technology with traditional data warehouse tool

The increase of storage ability of is far behind the increase of the data, current technology of data center can not satisfy the application requirement of big data. The reconstruction of the IT structure is unavoidable. The design of the rational layer-storage structure is the key of cloud GIS platform based on big data. The ability to extend externally and upward is vital for the structure because of the different database and different analysis environment of the individual user. The new storage must be compatible with other database and other data management environment, including the standard relationship data and analysis data warehouse. The technology of creative compress and the duplicate data deletion is the key of resolution of big data problem. The management ability and efficiency of big data is enhanced with the large-scale parallel processing of the economic server cluster, and the investment of IT resource is saved.

(2) The processing technology of real-time data

The decrease cost of data promotes the increment of data volume. The advent of the new data resource and technology of data acquisition promote the increase of data type. With the advent of all kinds of non-structured data, the complexity of big data is growing up. The big data still carry some risk. Statistician and scientist point out that the risk of error-find will increase with the high density survey of the big data cluster. It is challengeable of big data to utilize the value embodied in the data efficiently. The future research plan of big data should focus on the data engineering and not on the data science. The analysis method of the big data and the efficiency of the system is most important. The future mission of big data research is not to obtain more and more data, but to get rid of the redundancy and extract the essence. The problem needed to be resolved is the high-extendable and high-practicable data analysis technology, the new representation method of dada and high-flux computer etc.

(3) The registering service technology of information resources

The technology is to resolve the problem of the distributed register and management of the spatial data and service resources of water resources and hydropower. In order to help the user to understand the service and which service can meet the demand, the service provider register the description and access interface of the information service in the register center. The user search the service wanted in the service center. The service register center of information resource consists of many global register central database and cluster register central database. The global register central database is the superior database and the cluster register central database is the subordinate database.

The data of the global register central database and the cluster register central database is synchronized in the process of distributed register and management.

(4)The security of management and data

The key element restricts the development of cloud GIS platform is administrable and credible. The good extendable and plug and play self-management of component is necessary to reduce or exclude the human intervene for the transparent cooperation or amalgamation of cloud GIS platform of water resources and hydropower in the heterogeneous network environment, especially in the case of the large scale self-organization ubiquitous nodes and terminal self-management. The platform should response rapidly with the variety of the network and operation environment, guarantee the information security of the network transformation, sensor information collection and the peer-to-peer operation authentication, build the administrable and credible network support architecture of the user-oriented application of water resources and hydropower.

4 Conclusion

The big data is an important link between human society, physical world and information world. The construction of the information system that merge human, machine and object is significant. IDC predict that in 2015, the scale of the global big data market will reach 17 billion. More and more prospective technology companies pay attention on the business opportunity triggered by the big data. The big data epoch of water resources and hydropower is approaching with the smaller grain size, the increasing categories of data and the more accurate data. The capability of the analysis staff is vital for the competitive power of nation and enterprise.

In the end, the theory and the key technology of the cloud GIS platform of water resources and hydropower need to be improved, the business mode and the laws need to be exploded. For the sake of the development of water resources and hydropower, it must be driven by elaborate and executable business mode. With the exploration of the value of data, the resolution and guidance in all the lifecycle of water resources and hydropower, such as design, construction, immigrant and operation is provided. For the sake of the promotion competitive power of water resources and hydropower, the model and method of big data confirmation in the process of data mining and analysis is vital during the spread of information construction of water resources and hydropower.

References

1. Zhang, G.: The inspiration and reflection of big data epoch—Review of “The big data” of Xu Zipei. The Water Resources World, 40 (September 2012)
2. Li, L.: The forecast of trend in 2012. The Telecommunication Industry of China, 72 (February 2012)
3. He, B.: The four historical change of big data. The Technology of Telecommunication Network, 45–46 (July 2012)
4. http://blog.softwareinsider.org/2012/02/27/mondays-musings-beyond-the-three-vs-of-big-data-viscosity-and-virality/

Application of Fuzzy Logic in Prediction of Fire in João Pessoa City - Brazil

André Oliveira* and Marcelo Nero

Universidade Federal de Pernambuco (UFPE), Technology and Earth Sciences Center, Cartographic Engineering Department. Recife – Pernambuco – Brazil
{andreyoliver,marcelo.nero.ufpe}@gmail.com

Abstract. Fires cause economic damage, reaching productive areas, such as agriculture and livestock. This article aims to draw up a fire risk maps in João Pessoa City, Brazil. Using fuzzy logic integrated with a Geographic Information System (GIS), elaborated a thematic map based on land use, slope, density, hydrography, roads, substandard housing and rainfall, through methodology of multicriteria decision making. The results provided subsidy for public policy.

Keywords: GIS, fire, risk maps.

1 Introduction

In recent years, disasters caused by natural and human effects have emphasized the need to analyze, from a broader perspective, many natural processes.

The global monitoring tends to consider environmental problems from this perspective, favoring the understanding of this complexity in natural phenomena [1]. In a timely manner, this approach requires the use of techniques for obtaining, processing, and display of spatial information. Thus, the main goal of Geographic Information System (GIS), which has the capacity to store and process large volumes of data [2]. Therefore, it becomes possible to update or retrieve spatial information as well as produce cartographic themes in different ways through the levels of information contained in the geospatial database. GIS applications are quite appropriate for the fire dynamic study. The variety of factors that contribute to the onset and spread of fire towards an approach of integrated analysis.

Considering the intrinsic dynamism of this phenomenon, the present article shows the fuzzy logic application integrated GIS, aiming to support the prevention and control of fires (rural and urban environment), with the final product a fire risk map to the João Pessoa city, Brazil.

1.1 Study Area

Joao Pessoa city, Paraiba State capital, located in Brazil northeastern, in the eastern portion of the state, coordinates -34º47'30",-7°09'28". According to data from the

* Corresponding author.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 323–334, 2013.

2010 Census, conducted by IBGE - Brazilian Institute of Geography and Statistics [3], the city has a population of 723,515 people in an area of about 214.052 square kilometers, resulting in a density of 3421.30 inhabitants / Km ².

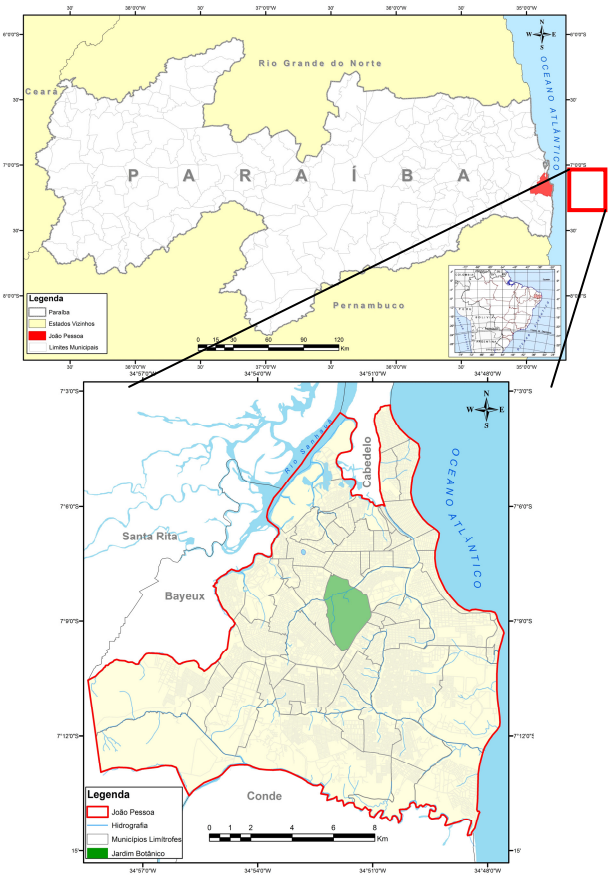


Fig. 1. Location of Joao Pessoa city, Brazil

2 Materials and Methods

Was obtained basemap municipal in shapefile format, UTM 25S projection, SIRGAS 2000, updated in 2010. The software used to generate the maps was ArcGIS 10 and fuzzy modeling was done in MatLab.

The first step is to create fuzzy rules using the software Matlab, using the tool Toolboxes - Fuzzy Logic - FIS Editor.

The variable "Land Use" evaluates the weight of the contribution of the predominant vegetation cover in the region for potential fire [4]. Therefore, it is necessary to rank the areas covered by vegetation, as well as the urban area evaluation

and the availability of water resources in the region, to identify risks for which systems are subject. The degree of relevance of that criterion comprises three classes of risk, as illustrated in the following figure.

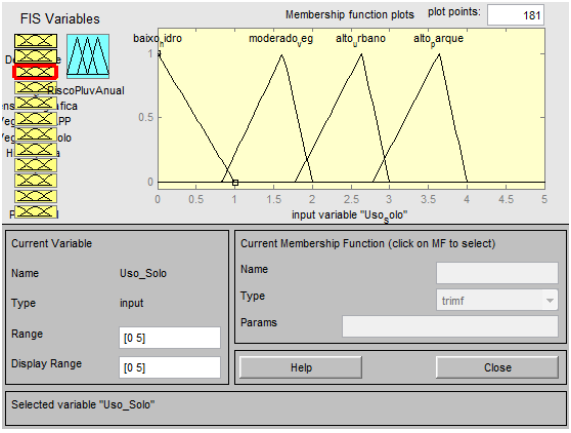


Fig. 2. Degree of Relevance to the variable "Land Use"

In evaluating of rainfall effect, one should consider not only the amount of rain that it falls, but also its seasonal distribution. If the rainfall distribution in a given location is uniform throughout the year, without a dry season, the potential for the occurrence and spread of fires is lower than a place where the rainy season is concentrated in a few months, with extended periods of drought during the other months. Thus, the distribution of rainfall is a key factor in defining the beginning, end and the length of the station fire danger [5]. Thus, the degree of relevance of this variable is shown in the figure below.

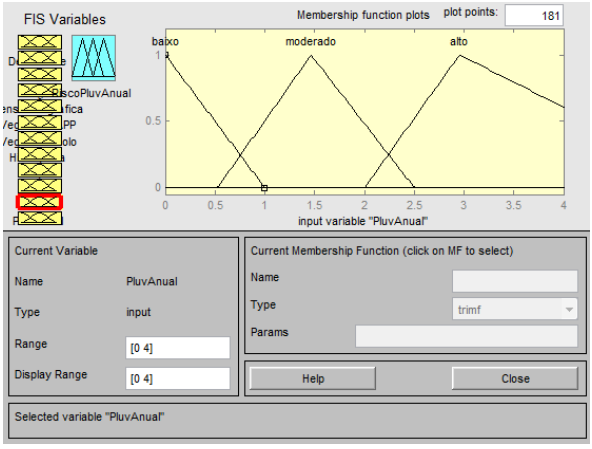


Fig. 3. Degree of Relevance to the variable "Pluviometry"

The road network allows visibility to people moving on the streets, allows access to the combat vehicles and can also function as cuts fires [4]. The access of people and vehicles increases the risk of fire. The degree of relevance of that criterion comprises three classes of risk, as illustrated in the following figure.

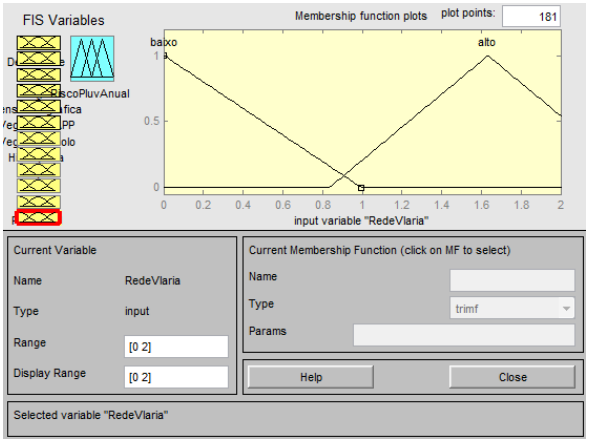


Fig. 4. Degree of Relevance to the variable "Road Network"

The more sloping terrain, bend over the flames in the direction of propagation. For the slope, the model was exemplified relevance of this variable illustrated in the following figure.

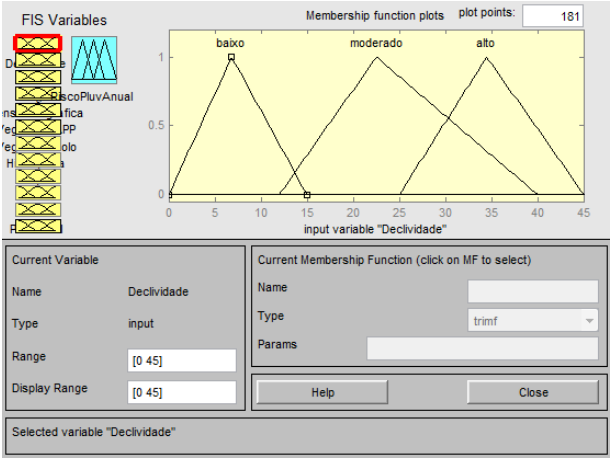


Fig. 5. Degree of Relevance to the variable "Slope"

The high population density increases the risk of fire due to possible oversights or action caused. The absence of such features as a factor that enhances the risk,

associated with abandoned property. Definition of fuzzy modeling for the density observed the degree of relevance this criterion, using the following template.

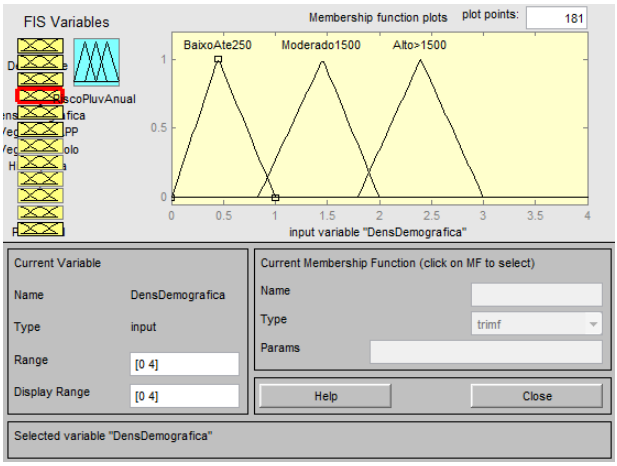


Fig. 6. Degree of Relevance to the variable "Density"

For hydrography, areas bounded by the buffers are considered to risks due to greater proximity to bodies of water, while the others are considered no risk, as illustrated by the following fuzzy model.

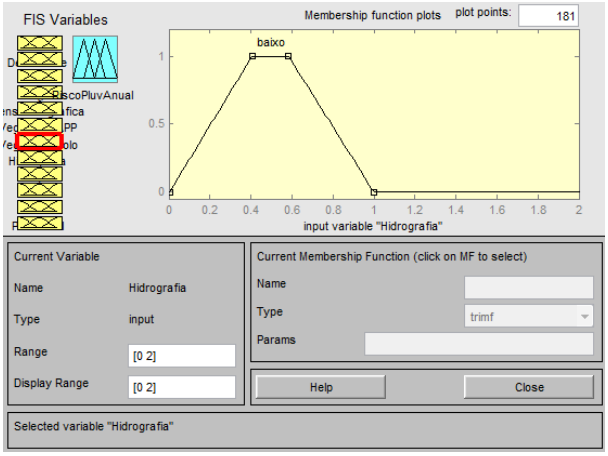


Fig. 7. Degree of Relevance to the variable "Hydrography"

The frequency of fires in slums generates understandable concerns and even suspicion, not necessarily justified. It is preferable to place the issue initially more appropriate in the context of adverse physical conditions and improper location of subnormal agglomerates. Dry weather undoubtedly favors these disasters because the houses are built with easily combustible materials such as wood and cardboard are

glued together, the hacks have exposed wires, gas cylinders, stoves lit and matches the reach of children whose parents, not infrequently, are absent at work. For subnormal agglomerates was attributed to high risk areas defined by this variable. Therefore, the degree of relevance of this criterion consists of a risk class only, as illustrated in the following figure.

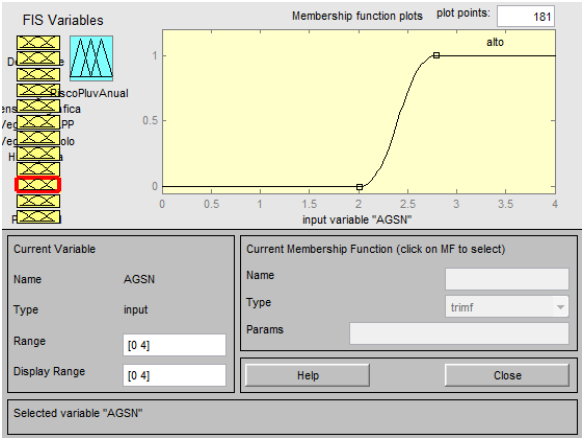


Fig. 8. Degree of Relevance to variable "Cluster Subnormal"

2.1 Calculation Modeling for Fire Risk

Each risk level defined according to their impact on the increased risk of fire. Weights were assigned in dimension to the final map between 0 and 25 (Fig. 9) taking into account the relative importance of each variable as a factor of risk of fire.

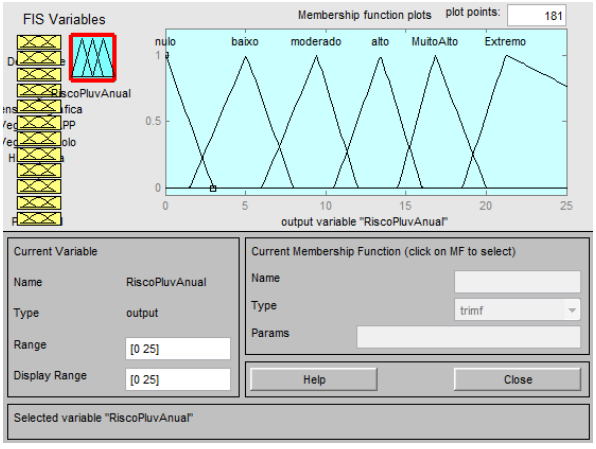


Fig. 9. Definition for calculating fire risk

2.2 Construction Rules

The rules were defined according to the degree of vulnerability of the variables used for this model, so that the final result in the risk factor also defined by linguistic variables: Null, Low, Moderate, High, Very High and Extreme. It was used the boolean operator "AND", as shown in the figure below.

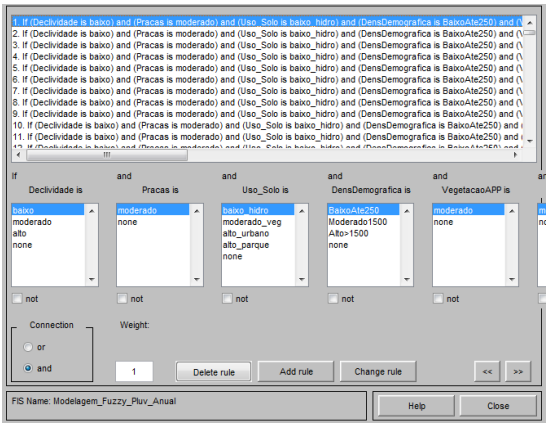


Fig. 10. Defining Rules

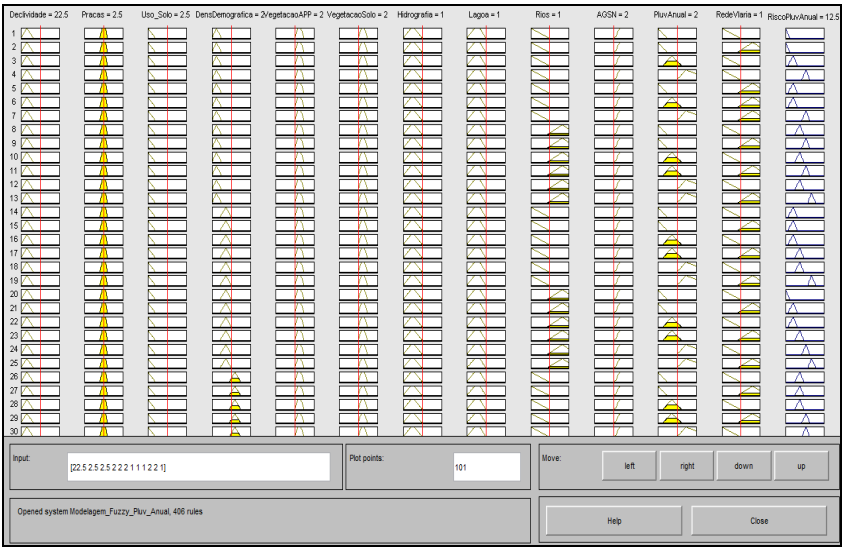


Fig. 11. Calculating the Risk of Fire

2.3 Visualization of Fuzzy Model for the Initial Values (Input)

After drafting the rules that define the risk was exported attribute table in ArcGIS with the values defined in this modeling. These values were inserted into Matlab

(input) and fuzzy value calculated was inserted in the attribute table and imported back into ArcGIS by joining tables, thus enabling the generation of a risk map.

Shape_Risco_Fuzzy

FID	Shape *	OBJECTID	PESO_DEC	PESO_B	PESO_L	PESO_H	VALOR_FUZZY
686	Polygon	687	4	1	3	0	0.998
687	Polygon	688	4	1	3	0	0.998
688	Polygon	689	4	1	3	0	0.998
689	Polygon	690	4	1	3	0	0.998
690	Polygon	691	4	1	2	0	0.998
691	Polygon	692	4	1	3	0	0.998
692	Polygon	693	4	1	1	0	0.998
693	Polygon	694	4	1	1	0	0.998
694	Polygon	695	4	1	1	0	0.998
695	Polygon	696	4	1	1	0	0.998
696	Polygon	697	4	1	1	0	0.998
697	Polygon	698	4	1	2	0	0.998
698	Polygon	699	4	1	3	0	0.998
699	Polygon	700	4	1	2	0	0.998
700	Polygon	701	4	1	3	0	0.998
701	Polygon	702	4	1	1	0	0.998
702	Polygon	703	5	1	3	0	1.03
703	Polygon	704	5	1	3	0	1.03
704	Polygon	705	5	1	3	0	1.03
705	Polygon	706	5	1	3	0	1.03
706	Polygon	707	5	1	3	0	1.03
707	Polygon	708	5	1	3	0	1.03
708	Polygon	709	5	1	3	0	1.03
709	Polygon	710	5	1	3	0	1.03
710	Polygon	711	5	1	3	0	1.03
711	Polygon	712	5	1	3	0	1.03
712	Polygon	713	5	1	3	0	1.03
713	Polygon	714	5	1	3	0	1.03
714	Polygon	715	5	1	2	0	1.03
715	Polygon	716	5	1	3	0	1.03
716	Polygon	717	5	1	3	0	1.03
717	Polygon	718	5	1	3	0	1.03
718	Polygon	719	5	1	3	0	1.03
719	Polygon	720	5	1	2	0	1.03

Fig. 12. Table imported into ArcGIS

3 Results and Discussion

To generate the fire risk model, the variables taken as input data were the land use, rainfall, population density, slope, roads, hydrography and subnormal agglomerates as mentioned earlier. The rainfall modeling was presented in four maps due to variation in rainfall throughout the year, with a thematic map for annual rainfall, one for rainfall in the rainy season, one for the first dry period and a pair in the second period of drought.

Analysis map and land use (Map 1), it can be concluded that although there are several classes of occupancy, there is a predominance of classes relating to urban areas (57%) and areas of environmental preservation (31%), such as mangroves and native vegetation, and hydrography (12%). Thus, interpreting the map as this variable falls as area fire risk from Low to Moderate.

The rainfall map was prepared using the Kriging interpolation where maps were generated in 2010 in the annual and monthly for the rainy season (April-August), first dry season (January to March) and second dry season (September to December). Looking at the map of annual rainfall (Map 2), it is observed that 67% of the study area had low risk areas distributed in the neighborhoods of the north, northeast, west and southwest of the county and 33% of the area showed moderate risk areas distributed throughout east and southeast south of the county. The thematic map generated for the rainy season (Map 3) had only one class of risk (low) for the entire municipality, where all the values resulting from the interpolation results showed less than 375 mm. The first dry period corresponding to the months of January, February and March, resulted in a thematic map illustrating two risk classes where 82% of the land area had low risk while 18% had moderate risk distributed in northern and

southeastern portions of the county (Map 4). The thematic map generated for the second drought period corresponding to the months of September, October, November and December (Map 5) resulted in two risk classes where 99% of the land area had moderate risk while only 1% had high risk distributed in a small area in the western portion and another in the southwest portion of the county. This result is justified by the fact that in the period from September to December 2010 the shortage of rainfall was higher than the first dry period, with a cumulative average rainfall of 1,238.7 mm in the last four months of 2010 while in the first three months of 2010, for the first period of drought presented a cumulative rainfall of 2719.0 mm.

Through the thematic map of population density (map 6), it was found that 41% of the land area have Low risk factor, 18% and 41% risk Moderate risk High.

Based on the map of hydrography (Map 7), visualize that 12% of the municipal area are influenced by the hydrographic network.

Elaborated a thematic map of the road network (Map 8) generating a catchment area of 100 meters from the axis of playgrounds. Thus, one can see that the area of influence road occupies 58% of the municipal area under the influence and 42% area without influence.

According to the thematic map of influence of slope to the risk of fire (Map 9), it can be concluded that the area presents quite varied and the slope f sister that this parameter corresponds to the variation of low risk to the extreme.

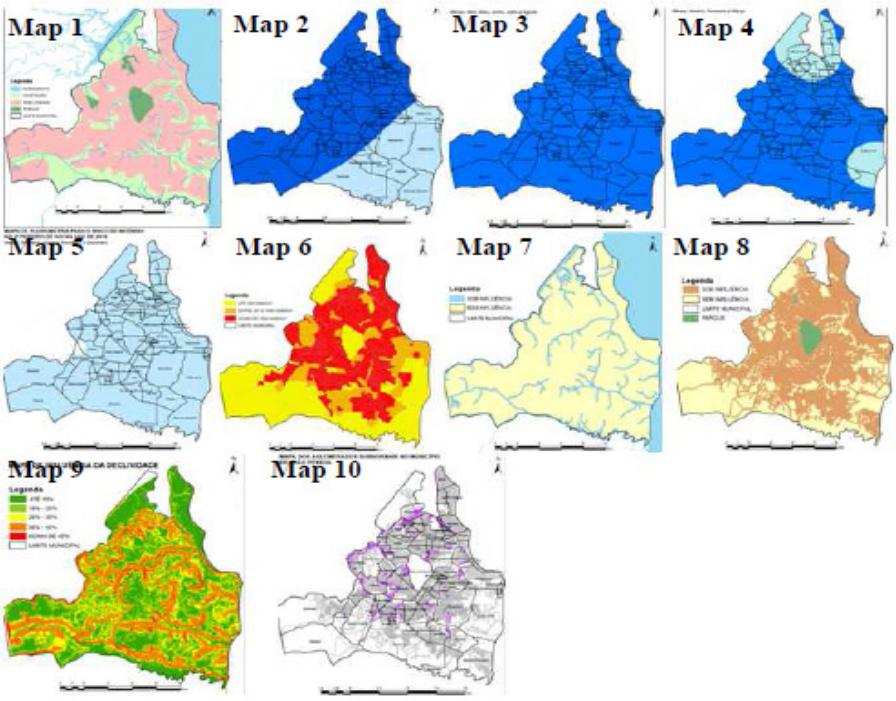


Fig. 13. Thematic Maps of the variables used in fuzzy modeling

Analyzing the thematic map of subnormal agglomerates (Map 10), it is clear that the areas classified in this category occupy a discrete area of approximately 4.8 square kilometers, which corresponds to 2% of the study area and according to the weight class proposal polygons paths of these areas are classified as high risk, according to the degree of fuzzy membership illustrated above.

3.1 Fire Risk Maps

Four maps were generated final risk, using fuzzy logic, generated by overlay of maps previously obtained, separating only the final maps of rainfall. Risk maps were generated separately according to the seasonality of rainfall data. Thus was created a fire risk map through the overlay of maps with the map of annual rainfall, a second map removing the map of annual rainfall and adding map of rainfall in the rainy season, a third map withdrawing map and adding the previous rainfall map of rainfall for the first period of drought and finally a fourth and final map removing the rainfall of the first dry period and adding the rainfall map for the second period of drought. The overlay results can be checked in the following figures.

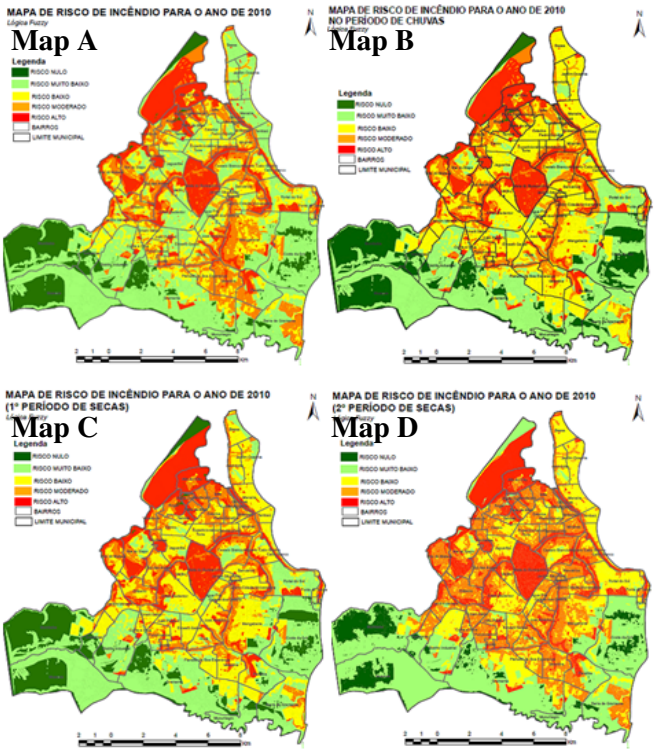


Fig. 14. Fire Risk Maps

Looking at the map of fire risk (Map A) involving the variables proposed in this paper and adding the map of annual rainfall in 2010, it was observed that 40% of the study area showed very low risk, followed by 19% of the municipal area that had low risk, 16% moderate risk, 14% had high risk and 11% of the study area obtained a risk factor null.

According to the map B, it was found that 31% of the land area had very low risk and 31% of the municipal area had low risk, 11% moderate risk, 13% had high risk and 14% of the study area obtained a risk factor null.

According to the map C, it was observed that 32% of the land area got very low risk and 29% of the municipal area had low risk, 13% moderate risk, 14% had high risk and 13% of the study area resulted in a risk factor null.

By analyzing past three maps, it was found that there was no significant difference between the ratings of areas at risk of fire and it was found that risk factors for low and very low showed a larger area of coverage due to the fact that, according to the data rainfall of the Local Executive Agency for Water Management, even though the first period of drought has submitted a rainfall rate less than the rainy season and the annual period, were still registered high levels of precipitation compared to the second dry period in the city, thus presenting, ideal climatic conditions not conducive to fire occurrence.

According to the map D, it was observed that 29% of the study showed very low risk, followed by 26% of the municipal area with moderate risk, 29% with very low risk, high risk showed 16% and 8% of the study obtained a risk factor null.

The maps generated with the use of fuzzy logic it was observed that the second dry season, between the months of September, October, November and December, represented the most critical time in relation to the risk of fire compared to the previous models seasonal due to the fact the above period occur a few rainfall compared to other times of the year. This result is also justified by the fact that in September the rains cease, so that the vegetation begins to show water deficiency. Only the fire risk map for the second dry season the moderate risk factor appeared in the highest percentage in relation to other classes of maps where the risk of larger scope were between the low and very low risk. It is also important to note that the road network and high population density contribute to the occurrence of fire, because although not determine their behavior, are likely start areas [6].

4 Conclusions

According to the models presented in this research, it was possible to evaluate the propensity of fires using a few variables and tools that are able to contribute to fighting and controlling fire in the study area. Using spatial analysis was possible to evaluate and diagnose clearly the classification of areas at risk of fires in Joao Pessoa city.

The use of a GIS which is a decisive geospatial decision making in the planning, prevention and policy implementation in public administration, which shows good efficacy in the management of information within the maps of fire risk. The methodologies used for generating predictive models required a high degree of

organization of data in your development process and also a high degree of knowledge of the functionality of a GIS, besides generating the maps have presented a low-cost production.

It is recommended that managers creating an information policy for people with small precautions that can be taken to avoid greater losses. Establish public awareness campaigns associated with disclosure of risks, training of staff to deal with dangerous situations and continuous monitoring of the areas considered most prone to fire. Attitudes are simple, but aimed at the population's safety and prevent damage that may be caused by the occurrence of fires. As a suggestion for future research, it is recommended that the proposed model compared with the areas actually affected by fire, validating the methodology and establishing finally a reliable study. As well as performing new tests involving other variables not used in this methodology, such as wind speed, aspect, elevation, vegetation water stress, among others.

Finally, it is concluded that the proposed model provides valuable information about the areas most liable to be affected by fire.

References

1. Chuvieco, E.A., Yebra, I., Nieto, M., Salas, H., Martin, J., Vilar, M.P., Martínez, L., Martín, J., Ibarra, S., De La Riva, P., Baeza, J., Rodriguez, J., Molina, F., Herrera, J.R., Zamora, M.A.: Development of a framework for fire risk assessment using remote sensing and geographic information system Technologies. Ecological Modelling 221, 46–58 (2010)
2. Burrough, P.A., Mcdonnell, R.A.: Principles of geographical information systems. Oxford University Press, Oxford (2005)
3. IBGE – Brazilian Institute of Geography and Statistics, http://www.ibge.gov.br
4. Ribeiro, L., Koproski, L., Stolle, L., Lingnau, C., Soares, R., Batista, A.: Zoning of Forest Fire Risks for the Canguiri Farm, Pinewoods (PR). FLORESTA, Curitiba, PR 38(3) (July/September 2008)
5. Soares, R.V.: Forest fire: control and use of fire. FUPEF, Curitiba (1985)
6. Prudente, T.D.: Geotechnology applied to mapping forest fire risk in the National Park of Chapada dosVeadeiros and surrounding area. UFU-MG. 114fl.: il (2010)

Application and Study of Three in One Gas Forecast Technology in Non-excavated Area of Gas Tunnel*

Ding Wang[1,2,3,**], Chang-wu Liu[1,2], Jie-bin Zhou[1,2,3], and Yan Zeng[3]

[1] State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University, Chengdu, Sichuan 610065, China
[2] College of Water Resource and Hydropower, Sichuan University, Chengdu, Sichuan 610065, China
[3] Sichuan Institute Coalfield Geological Engineering Exploration and Designing, Sichuan Coalfield Geology Bureau, Chengdu, Sichuan 610070

Abstract. The super colossal gas explosion accident happened in Dong-Jiashan tunnel of Du (Jiangyan) - Wen (Chuan) expressway, 2005. It led 44 people to die and shocked at home and abroad. From then on, increasing numbers public and the related functional sectors have paid more attention to gas safety managements during the gas tunnel excavated. It is well known gas existence in stratum should meet three necessary conditions, gas sources, channels and sealing conditions. The gas forecast technology by TGP geophysical, geological drilling and borehole gas emission content estimate composed can predict the gas sources and gas channels and gas sealing conditions. The borehole gas emission content estimate can predict the gas content in non-excavated area of gas tunnel. It is successfully applied to the non-excavated area of Le-Tunnel, the gas emission content: $0.164m^3/min<0.5m^3/min$. The forecast segment belongs to the low gas work area. Practice has proved that the three in one gas forecast technology can offer a further reference method for the same type of gas tunnel prediction technology.

Keywords: non-excavation area of gas tunnel, tunnel geology prediction system, the advanced geological drilling, gas emission content, three in one gas forecast technology.

The prediction for gas occurrence rule in front of tunnel excavated work face is the problem which didn't solve well, so it cause engineering geology and tunnel engineering concern [1]. As is known to all, the first priority of predicting gas occurrence rule is to determine whether the gas exists. Now, at the exploration stage of tunnel project, we forecast the gas outburst mainly through investigating the existing data of adjacent coal mines and gas fields.

* Fund program： A grant from the major state basic research development program of china (973 program) (No. 2010CB226802) and national natural science foundation of china(No.50574064, 50879049).

** Author's brief introduction： Wang Ding(1979-)， male， PhD students， engineer. The research direction: underground engineering. 39970374@qq.com

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 335–345, 2013.

At the same time, the structure of coal mines and the gas parameter are also important to prediction [2,3]. During construction, in order to ensure safety, we manage the gas on the basis of gas concentration which is obtained by automatic detection or artificial detection [3].

According to the gas forecast of Le-Tunnel engineering of Ya(An)-Lu(Gu) expressway ,we put forward the 'three in one' gas forecast technology. At the exploration stage of this tunnel project, we confirmed Le-Tunnel is high-gas tunnel by sample analysis of 5 boreholes. Because the quantity of drilling is very limited, we cannot accurately grasp gas occurrence conditions of the whole tunnel. During construction, automatic detection or artificial detection system also can not accurately grasp gas occurrence conditions of non-excavated area. Therefore, The main problem we explored is the gas forecast technology of the constraints in front of non-excavated area.

1 Geologic General Situation of the Tunnel

Le-Tunnel is a double hole one way traffic tunnel , the pile number of the left line are from K205 + 540 to K207 + 769. It is belong to the long tunnel which is 2229 m long.

Geological structure of predicting area: predicting areas' stratum is Baiguowan formation of the Triassic-Jurassic. Weakly weathered mudstone、silt-sandy mudstone 、quartz sandstone are the main stones, which is medium-thick layered, fracture development, rock breaking, block structure, belong to the soft rock or half-hard rock. The buried depth of predicting area is from 134 m to 195 m, no geological structure connected with surface such as fault in the area, so it had better air tightness.

Drilling hole explore suggests that the rock stratum of carbonaceous mudstone located in the pile number of K205+990. This rock stratum have 20° angle with tunnel excavation direction, about 22 m thickness.

2 Principles, Methods and Equipment

2.1 Principles

We conventionally consider that gas is called by a general term for all kinds of noxious gas released from rock mass or oil-gas. After years of research, Zhou Shining, Wang Xingshen, Yu Qixiang who are the professor of China university of mining and technology found that gas can move through the joint and fissure, fault and other geological structures driven by the pressure and concentration difference. It will be gathered to form an airbags, because of good trap and storage conditions of anticline, fold and so on. How much gas the rock stratum saved mainly depends on the closed condition: such as the buried depth of coal seam, the permeability of coal seam and surrounding rock, geological structure and storage conditions (such as the adsorption of coal properties, porosity of coal properties, etc.). [4-9] Thus it can be seen that if we want to grasp gas occurrence conditions of the whole tunnel, we should clearly know the gas source, gas migration pathway, and gas sealing condition of the predicting area, any of them is indispensable.

Combined the TGP geophysical prospecting and geologic advance prediction can effectively detect the lithological character of surrounding rock, broken degree, geological structures, etc. At the same time, we can grasp the gas content by prediction of gas emission quantity. Therefore, integrated the TGP geophysical prospecting、geologic advance prediction and prediction of gas emission quantity, we put forward the 'three in one' gas forecast technology, which can more accurately determine the gas source, gas migration pathway, gas sealing condition, so as to effectively predict the non- excavated area of gas occurrence conditions.

2.2 Methods

2.2.1 TGP12 Tunnel Geological Advanced Prediction System

TGP12 tunnel geological advanced prediction system is specially designed for tunnel and underground engineering, which is the advanced geological forecast research and development of geological prospecting technology achievements. In 2005, it is considered to reach the international advanced level by appraisal of the panel which make up with domestic famous expert in tunnel.

Working principle of the TGP：TGP12 tunnel geological advanced prediction system can predict tunnel geological conditions ahead by using the elastic wave which will form phenomenon of elastic wave reflection when encounter acoustic impedance interface, such as geological interface of lithological character changed, structure fracture zone, karstic zones.

Three-component geophone of TGP12 system placed in the receiving aperture, coupled with butter. Blasting adopts open circuit triggering method; blast hole charge 50 ~ 80 g. Under the condition of filling water, initiation and data acquisition is in turn. Survey line and borehole layout can be seen in figure 1.

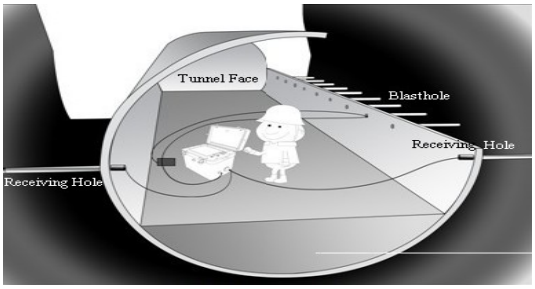


Fig. 1. Survey line and drilling hole layout diagram for TGP

2.2.2 Advanced Geological Drilling

Using permissible tunnel drilling rig, according to the size of hole bitmap to arrangement borehole (as shown in figure 2). Along the advancing direction of tunnel drilling, geological situation revealed by advanced geological drilling is on-site record by geological professional technical person. At the same time, determined the construction geological situation in front of us through the relevant test. Specific methods are as follows:

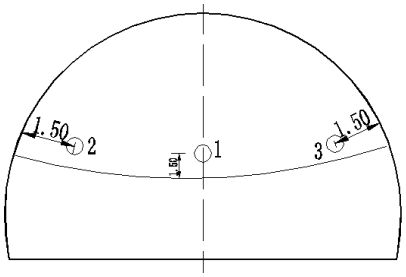


Fig. 2. The layout diagram of drilling hole

1.Determine by the drilling speed

As far as we know, rig drilling speed is uniform at the same rock. Therefore, in combination with strata of tunnel excavated and the changes in the rate of drilling in the drilling process or whether there is sticking phenomenon, etc. We can judge whether the rock in front of us is the bad geological body.

2. Determine by the washing fluid

In the drilling process, we can determine the change of the rock drilling ahead through color changes of the washing fluid, Therefore, we can gasp the lithological character of the advanced rock stratum by identifying rock powder composition of the flushing fluid. At the same time, we can determine whether there is abnormal body ahead and abnormal body property, development depth and scale of abnormal body through impurities composition in the flushing fluid.

2.2.3 Determination of Gas Emission Quantity [10]

Using optical interference methane detection alarm to detect the gas concentration of this area, which located in the tunnel face and about 40 m behind the tunnel face. Generally, arrange 8 ~ 12 measuring points at the tunnel face, arrange 1 ~ 2 measuring points Close to the drill hole. Arrange 5 ~ 10 measuring points at 40 m behind the tunnel face and detection frequency is determined by the borehole gas emission situation. Measuring point layout can be seen in figure 3.

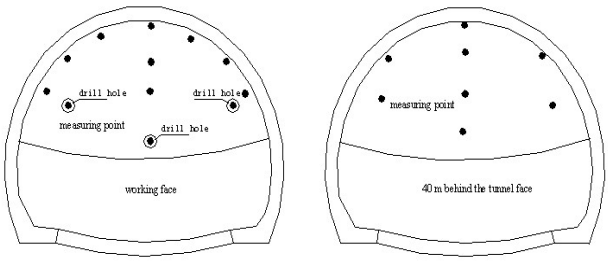


Fig. 3. The distribution diagram of the monitoring points of gas concentration

We can achieve air volume from chimney fan. According to the standard method for determining wind speed with high-speed wind watch. Product of wind speed and cross section is air volume.

Using a thermometer, portable air pressure meter direct read the temperature and atmospheric pressure.

Formula for gas emission quantity:

$$q = 3.526 \times (273 + T) \times Q \times \frac{a}{P} \tag{1}$$

In the Formula : q is the absolute gas emission quantity,m^3/min; T is the measuring point temperature, ℃ ; Q is the air volume of measuring point, m^3/min ; a is the gas density of measuring points, % ; P is atmospheric pressure, hPa。

The return current gas of the tunnel face can be approximate as borehole gas emission.

3 The Existence of Gas Prediction

3.1 Detection Results

(1) In different drilling depth, Gas concentration, temperature, atmospheric pressure and air volume gas parameters of the typical parts such as vault of working face, orifice, the return air of working face and so on can be determined. the determination results are shown in table 1.

Table 1. The detection results of gas parameters

Mileage	Air volume (m3/min)	temperature (℃)	Air pressure (hPa)	location	concentration (%)
K205+847	188	28	792	vault	0.02
				In the return air	0.01
K205+855	190	28	792	vault	0.02
				orifice	0.03
				In the return air	0.01
K205+875	195	27	792	vault	0.03
				orifice	0.04
				In the return air	0.02
K205+892	200	27	792	vault	0.05
				orifice	0.08
				In the return air	0.03
K205+910	202	27	792	vault	0.06
				orifice	0.10
				In the return air	0.04
K205+950 (the end hole)	199	27	792	vault	0.08
				orifice	0.12
				In the return air	0.05
K205+950 (stable)	205	27	792	vault	0.10
				orifice	0.18
				In the return air	0.06

3.2 Data Analysis

3.2.1 The TGP Geophysical Exploration Analysis

(1) Main technical parameter and drawing of the TGP system

1. Parameters of TGP: Each sampling rate is 4096 times; sampling interval is 0.100 ms.

2. forecast range: There are 150 m in front of the import of the left line, just from K205 + 847 to K205 + 997.

3. geotechnical elastic parameter：The average longitudinal wave velocity (Vp) is 1870 m/s; The average shear wave (vs) at the rate of 1050 m/s

4. The result image of migration homing of p-wave in same side and the result image of migration homing of p-wave in opposite side are shown in fig.4 and fig.5.The result image of migration homing of three components in same side and the result image of migration homing of three components in opposite side are shown in fig.6 and fig.7.

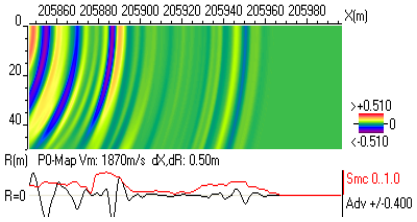


Fig. 4. The result image of migration homing of p-wave in same side

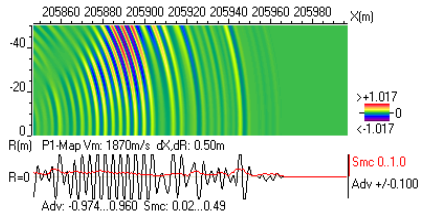


Fig. 5. The result image of migration homing of p-wave in opposite side

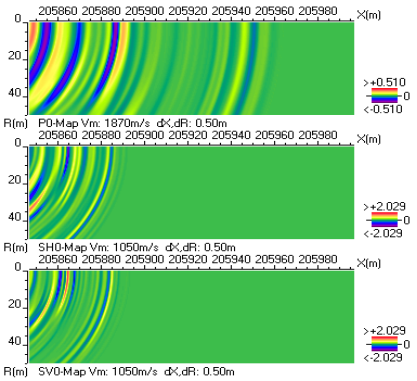


Fig. 6. The result image of migration homing of three components in same side

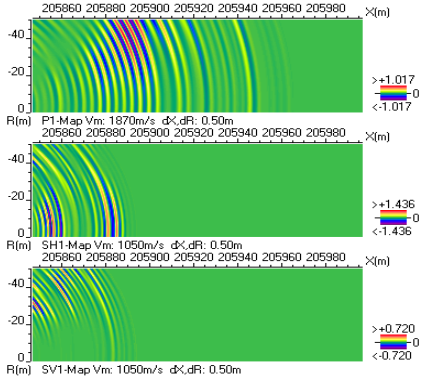


Fig. 7. The result image of migration homing of three components in opposite side

(2) results of the TGP prediction

Through the analysis of the fig.4, fig.5, fig.6, fig.7, it can be seen that：

1. K205+847~K205+900(53m)

P-wave in same side and in opposite side have Strong reaction in this area, shear wave reflection is also strong. P-wave in same side and in opposite side longitudinal wave and shear wave decay curve characteristics of interface migration image is negative. According to the existing geological conditions and possible adverse geological, this paragraph of surrounding rock is broken. At the same time, joint, crack development, medium permeable, possible influence on the stability of surrounding rock is not allow to ignore ,these parts of the possibility of large deformation is bigger. Especially the surrounding rock of K205 + 880 ~ K205 + 890 section is very broken, water-bearing fracture is strongly.

2. K205+900~K205+915(15m)

P-wave in same side and in opposite side in this period had obvious reflection, shear wave reflection is not obvious. Longitudinal wave and shear wave decay curve characteristics of interface migration image of P-wave in same side and in opposite side is negative. According to the existing geological conditions and possible adverse geological, surrounding rock of this section is relatively complete, there may be little water zone.

3. K205+915~K205+960(35m)

P-wave in same side and in opposite side have Strong reaction in this area, shear wave reflection is not obvious. P-wave in same side and in opposite side longitudinal wave and shear wave decay curve characteristics of interface migration image is negative. According to the existing geological conditions and possible adverse geological, this paragraph of surrounding rock is broken, water-bearing fracture is strongly.

4. K205+960~K205+997(37m)

In this section, reflection of P-wave in same side and in opposite side is very weak, shear wave reflection is not obvious. There is not clear reflection of the rest part.

3.2.2 Analysis of Advance Geological Drilling

(1) In the section of K205+847~K205+895, drilling is with occasional sticking, jumping drill, but it is smooth in general. So we speculate that surrounding rock of this section is relatively broken, locality relatively complete. Through the identification of the rock powder composition in the hole, the main rocks are powder sandstone, mudstone and quartz sandstone.

(2)In the section of K205+895～K205+910, drilling is with occasional sticking, but it is smooth in general. So we speculate that surrounding rock of this section is relatively complete, locality relatively broken and Joints and fissures is bad development. Through the identification of the rock powder composition in the hole, the main rocks are powder sandstone, mudstone and quartz sandstone.

(3)In the section of K205 + 910 ~ K205 + 950, drilling is very difficult, drilling speed under 0.2 m/min. At the same time, phenomenon of sticking, jumping drill is seriously, so we speculate that surrounding rock of this section is relatively broken, locality relatively complete. Through the identification of the rock powder composition in the hole, the main rocks are powder sandstone, mudstone and quartz sandstone.

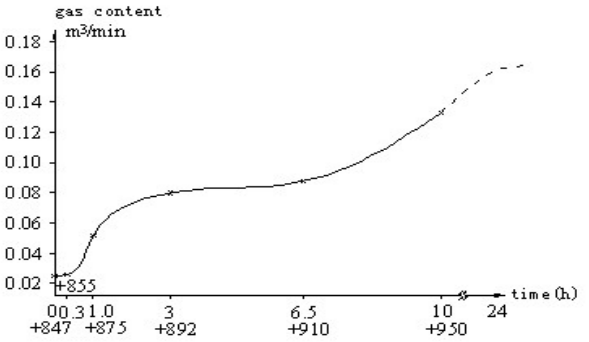
Determine the surrounding rock is broken, joint fissure development through the color changes and impurity elements of flushing fluid.

3.2.3 Calculation Analysis of Gas Emission Quantity

According to the formula (1),we calculate gas emission in different drill in drilling deep by the gas concentration, air volume, temperature, atmospheric pressure value in the return current of working face (shown in table 3).Calculation results are shown in table 2, the gas emission changes are shown in figure 8.

Table 2. The variation of gas content in different mileage

Mileage	K205 +847	K205 +855	K205 +875	K205 +892	K205 +910	K205 +950 (the end hole)	K205 +950 (stable)
Gas content	0.025	0.026	0.052	0.080	0.088	0.133	0.164



Note: "+ 847" is the end mileage of different drilling depth to detect gas. The drilling lasted 10 h, it is unlikely to continue drilling reach to the pile number of K205 + 950, borehole gas emission keep basically stable after stop drilling 14 h ago. The dashed part of variation of gas content is the changes of gas Situation in the final hole, gas content finally tends to a stable value.

Fig. 8. The variation diagram of gas content

4 Analysis of Gas Results Prediction

By the discusses of TGP geophysical prospecting, drilling and gas emission estimates "three in one" comprehensive gas prediction technology, combined with the tunnel geological prospecting data, we can know that：

(1) the prediction range: the prediction by TGP12 tunnel geological advanced prediction system is relatively clear in the section of K205+847 ~ K205 + 960,so we adopt geological prediction; Advanced drilling range is from K205 + 847 to K205 + 950. According to the forecast results of geophysical prospecting and drilling, the prediction range is K205 + 847 ~ K205 + 960.

(2) the section of K205+847~K205+890. Through results of the TGP geophysical prospecting and drilling predicted can be seen that surrounding rock of the section is relatively broken and also existed joints and fissures to form the gas migration pathway. At the same time, there have not such faults and fissures connected to the surface. So it forms the condition for gas storage. By the measurement of drilling gas content, it shows that from figure 13 the gas content at about 0.08 m^3 / min when drilling to K205+890. It instructed that section of K205+847~K205+890 exists gas. Through the identification of the rock powder composition in the hole, the main rocks are powder sandstone, mudstone and quartz sandstone. There are no coal or coal measures strata in this section, it instructed that it does not have conditions for gas generating. The other section of a small amount of gas are transported to this section mainly through the joint or crack, and eventually preserved. But water-bearing fracture is strongly in this section, shows that the groundwater take away the large amounts of gas in the long geochronology.

(3) the section of K205+910~K205+950.The surrounding rock of this section is relatively complete. In combination with gas content change trend of figure 13 shows, this gas emission from 0.08 m^3 / min up to 0.088 m^3 / min, with a little change. Therefore, we think that the gas is from K205 + 847 ~ K205 + 890 segment which continues to release. Because this paragraph neither formation does not has gas generating conditions, nor has gas migration channel.

(4) the section of K205+910~K205+950 .Through results of the TGP geophysical prospecting and drilling predicted can be seen that surrounding rock of the section is relatively broken and also existed joints and fissures to form the gas migration pathway. At the same time, there have not such faults and fissures connected to the surface. In combination with gas content change trend of figure 13 shows the gas content is about 0.088 m^3 / min when drilling to K205+910. The gas content is up to 0.133 m^3 / min when drilling to K205+950. Gas emission keep basically stable after stop drilling 14 h ago, and the gas content is up to 0.164 m^3 / min. It instructed that section of K205+847~K205+890 exists gas. Through the identification of the rock powder composition in the hole, the main rocks are powder sandstone, mudstone and quartz sandstone. There are also no coal or coal measures strata in this section, it instructed that it does not have conditions for gas generating. But why the gas content is increased obviously? Mainly because coal-bearing strata (carbonaceous mudstone layer) in the section of K205 + 990 ~ K206 + 054 experienced the period of biological gasification, it has the condition of the gas generated. This gas can be transported to section of K205+910~ K205+950 mainly through the joint or crack, and eventually preserved. Maybe the seal gas will be released when excavate tunnel, and form a gas hazards.

5 Conclusion

Using the TGP12 advance geological forecast system and advanced drilling can effectively detect changes in lithology interface geological, tectonic fracture zone, rock crushing levels, as well as lithology, so as to determine whether there is gas source, gas migration pathways and gas closed condition. By borehole gas emission measurement

can determine what formation to save gas, can also corroborate whether they have gas closed condition. therefore, integrated the TGP geophysical prospecting, drilling, borehole gas emission measuring the formation of the "three in one" gas prediction technology can more accurately predict the tunnel non-excavation area of gas occurrence conditions, which was proved to be feasible at Le-Tunnel.

Through the "three in one" gas prediction technology apply to Le-tunnel, currently in the normal ventilation of the tunnel, the gas concentration is low at tunnel. The gas emission of forecast area is lesser and the maximum value of gas emission is $0.164m^3/min$ $<0.5m^3/min$. According to "gas tunnel railway technical specifications" classification standards of gas work area, K205 +847 ~ K205 +950 section is low gas work area.

By the discusses of TGP geophysical prospecting, drilling and gas emission estimates "three in one" comprehensive gas prediction technology, combined with the tunnel geological prospecting data, Le-tunnel K205 +847 ~ K205 +950 section of exposed stratigraphic is not coal or coal bearing stratum, and is not undergo biochemical or coalification metamorphism gasification period. It is not coal and coal measures strata gas and gas generating conditions is poor. From the drilling gas emission, geophysical prospecting, drilling results, can be seen that the segment strata have a certain gas closed condition, and the gas of surrounding coal-bearing can be through fracture zone, joints, cracks and so on to supply the rock formation of tunnel through area so that form gas migration. If the construction process ventilation, gas safety management may not be implemented, may cause localized accumulation of gas or gas overrun, lead to gas accident.

The "gas tunnel railway technical specifications" into the tunnel gas work area is mainly built on the basis of geological prospecting work, however construction of the actual operation still determination of gas emission in the tunnel face, and thus the gas tunnel work area partition. Whether relying on geological survey data or field measurement data, or a combination of both, cannot accurately grasp the gas deposit situation of non- excavation area. Therefore, during the construction phase we should take advantage of gas advanced prediction technology to divide into gas tunnel work area categories. This article explores "three in one" gas prediction technology that consist of the TGP geophysical, drilling, drilling of gas emission estimates can offer a further reference method for gas forecast of tunnel no excavation area.

References

1. Fu, H., Li, Y.: Briefly Introduce to Forecast Technology of Karsts and Gas During the Period of Tunnel Construction. Mining Technology, 46–48 (2007)
2. The Industry Standards Compilation Group of People′s Republic of China. Technical code for the coal and peat geological exploration (DZ/T0215–2002). China Railway Press, Beijing (2002)
3. The Industry Standards Compilation Group of People′s Republic of China. Technical code for railway tunnel with gas (TB10120–2002/J160–2002). China Railway Press, Beijing (2002)

4. Wang, S., Yu, Q.: Theory and technology of Mine disaster prevention and control. China University of Mining Press, Xu Zhou (1997)
5. Zhang, Z., Zhang, Y.: Gas geology rule and gas prediction. Coal Industry Publishing House, Beijing (2005)
6. Cao, Y., Li, K., Zhang, Y.: The Structure Classification of Tectonic coal. A. Symposium on Gas-geology. Coal Industry Press, Beijing (1995)
7. Cao, Y., Zhang, Y., Li, K.: The Tectonic Coal of Dynamic Metamorphism and Evolution Rules. Technical for Coal Geology and Exploration, 15–17 (1996)
8. Fu, J., Liu, D.: Condition of gas migration, reservoir. Science Press, Beijing (1992)
9. Han, D., Yang, Q.: China coal field geology. Coal Industry Publishing House, Beijing (1979)
10. Wang, D., Liu, C., Wang, D.: Artificially Statistical Monitoring and Preventing the Gas Hazard of Large Cross-sections Underground Engineering. Metal Mine (suppl.), 441–444 (2011)

Scale-Free Model in Software Engineering: A New Design Method

Zhengxu Zhao[1] and Yang Guo[2]

[1] School of Information Science and Technology, Shijiazhuang Tiedao University, Shijiazhuang 050043, Hebei, China
[2] School of Mechanical Engineering, Shandong University, Jinan 250061, Shandong, China
zhaozx@stdu.edu.cn, guoyang1013@126.com

Abstract. Along with software systems become larger and more complex, engineers need better ways to design and maintain these systems. We apply complex network theory into software engineering for this purpose. Recently researchers have found most of complex systems comply with a scale-free network model. This paper presents an initial research into the small-world and scale-free phenomenon within the development of a three-dimensional visualization platform, Total Discovery of Space, which has applied to practical aerospace engineering successfully. It is expected that this model will be practical for software engineering.

Keywords: Information visualization, Function-call graph, Scale-free network, Small world effect, Power-law distributions.

1 Introduction

1.1 Information Visualization

Visualization is any technique for converting data into graphics or images displayed on the screen and interactive processing through the use of computer graphics and image processing technology. It involves many fields such as computer graphics, image processing, computer vision, computer-aided design, and become a integrated technology focusing on a series of problems like data representation, processing and decision analysis. Information visualization concentrates on the use of computer-supported tools to explore large amount of abstract data to reinforce human cognition.

Due to the current rapid development of aerospace industry, three-dimensional visualization technology has been widely applied in aerospace field. There are many difficulties in the aerospace tracking mission. For example, due to the optical equipment on the ground can't trace the flight state of target spacecraft, technical engineers only monitor the running state by the remote sensing data. These data were often processed into various tables and curves for experts to analyze and deduce the flight state of target spacecraft, capturing the changes of positions and attitudes. The serious problem of this simulation method is unable to provide a way for technical engineers in an intuitive manner. The ground control center at first adopted the method of

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 346–353, 2013.

three-dimensional animation to reappear the tracking state of target spacecraft. Although it can simulate the current state, the tracking data from observation stations could be out of use in simulation process. The data were only sent to experts who can understand and analyze them. Gradually the requirement of spacecraft tracking visualizing technique is getting higher and higher. The traditional approach, using Vega prime or Satellite Tool Kit to simulate the state of target spacecraft, has been unable to meet the demand like big data processing and real-time rendering. So there is an urgent need for a three-dimensional visualization system that can support real-time rendering and big data processing. Our visualization system, Total Discovery of Space, has been developed in this context.

Total Discovery of Space, often abbreviated to "TDS", provides three-dimensional, photo-realistic, real-time viewing of the Solar System, the galaxy and the universe. The design intention of TDS is providing a visualization platform for deep space exploration mission. Unlike most other planetarium software systems, TDS doesn't confine users to the surface of Earth. Users can travel throughout the solar system, to any of over 100,000 stars, or even beyond the galaxy. Considering the stability and security of the aerospace software, we have proposed a software engineering methodology for developing systems using complex network theory, and successfully applied to the development of TDS.

1.2 Software Systems as Networks

Complex network is being considered as an important interdisciplinary approach to representing complex systems these years. Small-world phenomenon and power-law distributions have been found in many instances of natural and man-made systems. For understanding the behavior of numerous and large complex systems, networks have become a promising research area in fields such as the Internet, and also the World Wide Web. These networks are considered to show small-world and scale-free behavior. Small-world here means that characteristic path length is very short in a network, regardless of the clustering property. Scale-free behavior implies that the network lacks a "characteristic length scale" and the distribution of their network graphs follows a power-law. Power-law distribution means that small values are greatly common, and greater values are extremely rare [1-2].

The study of software systems as networks is an emerging field. It may contribute to solving problems in software engineering, such as managing the complexity of software growth, or how to create better programs for software engineers. This idea of software systems as networks is not novel; it has been presented by Knuth twenty years ago. Knuth asserted that "a program is best thought of as a web instead of a tree".

2 Preliminaries

2.1 Small-World Phenomenon

One of the most widely discussed of network phenomena is the small world effect, which means the typical network distances between vertices are surprisingly small.

Small-world phenomenon originated in a 1929 short story entitled Láncszemek (Links) by Hungarian author F. Karinthy. In the 1960s, psychologist Stanely Milgram conducted a small -world experiment by mirroring this story in examining the chains of acquaintances among individual networks over the U.S. population. He ran an experiment to measure the average number of intermediate acquaintances needed to deliver a letter addressed to a stockbroker in Pittsburgh. Each person was required to hand the letter over to someone with whom they believed can eventually deliver the letter. This experiment revealed how society was connected by approximately six degrees of separation. Since then the notion of six degrees of separation has thus gained iconic status in today's popular culture. In mathematical terms, small world effect is a hypothesis that the mean distance is small, in a sense that will be defined shortly.

Applied mathematicians and physicists have researched the small-world effect in social, technological, and biological networks. Watts and Strogatz has provided a formal presentation using graph theory concepts. They have studied lots of real networks, finding the shorter average path length and higher clustering coefficient. The network with these features is called a small-world network. The average path length in a small-world network is approximate with that of the random network, where the average path length in random networks is proportional to the logarithm of their size. Clustering coefficient is inversely proportional to the size of network, which is significantly larger in a small-world network than that for the random network.

2.2 Scale-Free Phenomenon

Scale-free here means that the network lacks a "characteristic length scale". The method to certify a scale-free phenomenon is to see if it statistically shows up in the form of a power law distribution. A power law is a type of probability distribution and has attracted particular attention over the years for its mathematical properties.

In power law distributions, the number of occurrences N_k of some event of size k is proportional to k raised to some power. Mathematically, a quantity x obeys a power law if it is drawn from a probability distribution

$$p(x) \propto x^{-\alpha} \tag{1}$$

where α is a constant parameter of the distribution known as the exponent or scaling parameter. But because there're not many empirical phenomena which obey power laws with all values of x, the power-law distribution usually applies only for values larger than some minimum x_{min} and the tail of distribution follows a power law in this case.

Power laws can be also defined as a complementary cumulative definition under some circumstances, where the probability that a random variable takes at least a value is proportional to a negative power of that value, that is:

$$P(X \geq x) \propto cx^{-k'} \tag{2}$$

2.3 Function-Call Graph

Function-call graphs represent an important information space of software systems. The function-call graph, which presents the calling relationships between functions, is a useful representation of programs that can aid understanding for engineers. It is a directed graph in which each node represents a function and each edge (f_1, f_2) indicates that function f_1 calls function f_2. Thus, a cycle in the graph indicates recursive function calls.

3 Related Work

3.1 Existing Planetarium Software Systems

Vega Prime is the most productive COTS visualization tool for real-time 3D development and deployment of simulation applications. It is a product of Multigen-Paradigm Company. Ideal for both high-performance and low cost hardware platforms, Vega Prime's extensible plug-in architecture facilitates the rapid design and prototyping of real time 3D applications by utilizing the most sophisticated technology available, within an easy-to-use toolkit[1].

Satellite Tool Kit, often referred to STK, is described as a physics-based software package from Analytical Graphics Inc. It allows engineers to perform complex analyses of ground, sea, air, and space assets, and share results in one integrated solution. At the core of STK is a geometry engine for determining the time-dynamic position and attitude of assets, and the spatial relationships among the objects under consideration including their relationships or accesses given a number of complex, simultaneous constraining conditions. It is originally created to solve problems involving Earth-orbiting satellites[2].

3.2 Empirical Analysis of Systems Based on Complex Network Theory

There have been substantial works on small-world phenomenon and power-law distributions in natural phenomena or the web over recent years. Until recent years researchers discuss the structure of software systems combined with the complex network theory.

Valverde et al. [3] first used the method of complex network to analyze the topological structure of the Java Development Framework. They had finally concluded that the software architecture graphs, where nodes represented classes and edges means relationship between them, revealed small-world and scale-free behavior. Myers [4] also found collaboration graphs of some software systems written in C/C++ displaying the same properties and proposed some measures of network topological structure to discuss the relationship to software engineering practices. Potanin et al. [5] had examined numerous different systems written in different languages and studied the graphs where objects as nodes and references as edges. Finally they had

[1] http://www.presagis.com/products_services/products/modeling-simulation/visualization/vega_prime/

[2] http://en.wikipedia.org/wiki/Systems_Tool_Kit

drawn a conclusion that the graphs obeyed a power-law distribution. Concas et al. [6] had presented a study which used an implementation of an object-oriented system in order to search for scaling laws. He had studied some system properties, such as the distributions of methods or classes and inheritance hierarchies. They also explained why the out-degree distribution showed log-normal or power-low behavior in class graphs. Louridas et al. [7] provided evidences to show that the distribution with long, fat tails in software systems appearing at different levels of abstraction in diverse languages or systems were much more common than that known before.

4 A New Design Methodology

Complex network theory is a new approach in studying different types of large complex systems in the abstract worlds. In complex networks, it displays a surprising degree of robustness: while one node even the key component regularly malfunctions, local failures rarely lead to the loss of the global network. The stability of complex systems is often attributed to the redundant wiring of the functional network.

But for the software system, while a function fails to implement, the system may throw an exception or even collapse immediately. This phenomenon is absolutely not allowed in some important fields, especially the Aerospace measurement and control area. However, with the continuous increasing scale of software system，the existence of loopholes or the fault could be inevitable. It is still to be solved that how to make the software systems still provide consecutive error-free service when internal errors or external interference occurred. Considering the robustness of the network, if the network composed of functions meets the complex network characteristic, Parameters can be passed on by another path when some function failures so that the system of sudden collapse can be avoided.

The stability of software systems mainly depends on a lot of test latterly at present. In this paper, we will propose a new software design methodology using complex network theory to construct system model for guiding software development. Taking the TDS system as an example, we studied this system at the function level and extracted the calling relationship to construct complex network model.

4.1 Statistical Analysis

We have elaborated a reconstruction algorithm based on Windows systems for obtaining the function-call graph from software systems [8]. Fig. 1 shows the directed function-call graph for TDS, in which the size of nodes is proportional to their node degrees. There are 2123 functions and 10060 calls in the function-call graph and the average degree is 4.739. The average path length 3.771 is proportional to the logarithm of network size and the diameter of graph is 12 referred the calling path from function *readTypeColor* to *Star::getOrbitingStars*. The average clustering coefficient of this graph is about 0.0767, which is significantly larger than that for the same size of random network, where the clustering coefficient is inversely proportional to the network size.

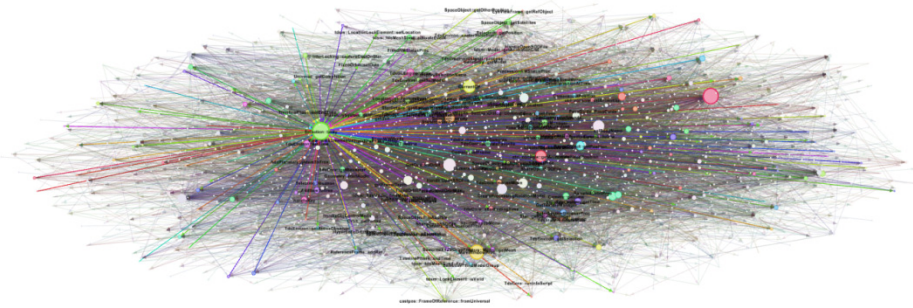


Fig. 1. Directed function-call graph for TDS

4.2 Degree Distributions

The degree distribution is an important attribute for network analysis. In the directed graph, node degree can be divided into in-degree and out-degree, which respectively represents the number of tail endpoints adjacent to a node and the number of head endpoints adjacent to a node. The degree distribution of a network is defined to be the fraction of nodes in the network with degree k, which is also sometimes presented in the form of a cumulative degree distribution, the fraction of nodes with degree greater than or equal to k. Fig.2 shows the degree distribution of function-call graph for TDS. From this figure, most of nodes in the network have low degree, but there is a significant "heavy tail" trend similar to the power law distribution, corresponding to nodes with substantially higher degree. The highest degree of node in the network is 285. Such a well-connected node is often called a hub.

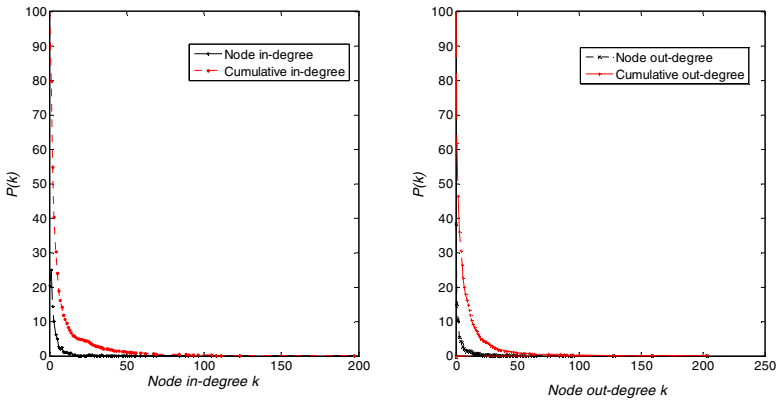


Fig. 2. Degree distribution of function-call graph for TDS

We also have studied the distribution and plotted the node degree and cumulative degree distributions in log-log scale, shown in Fig.3 and Fig.4 respectively. In the two figures longitudinal coordinate indicates the number of nodes in function-call graphs with degree greater than or equal to k and the lines express power-law fits on log-log scales.

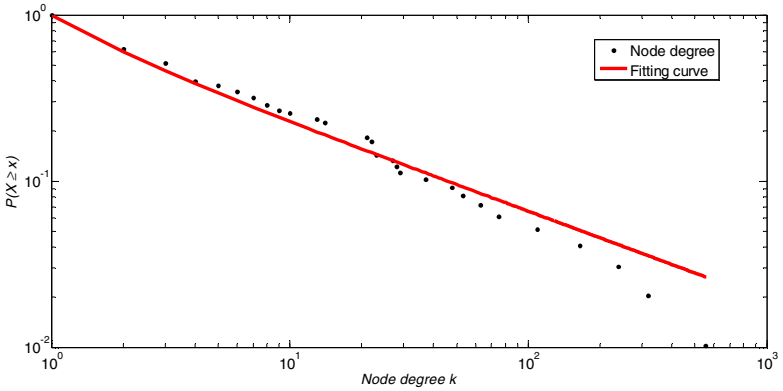


Fig. 3. Node degree distribution for TDS

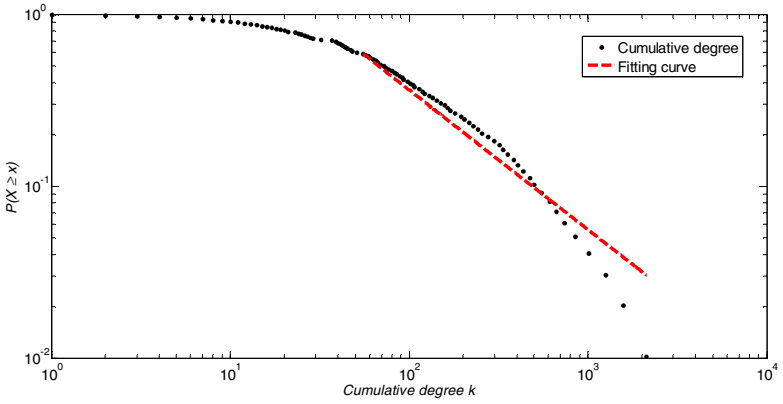


Fig. 4. Cumulative degree distribution for TDS

5 Conclusion

With the recent studies in the research of complex networks, we have proposed a new software design methodology by analyzing the function-call relationship and constructed a scale-free model for system development. In this model, the degree distribution of function-call network displays an obvious heavy-tail distribution like other real networks obeying power laws, and the average clustering coefficient is also larger than that of a random network in the same size, which is similar to the phenomenon of

small world effect. This model has been applied to the development of TDS successfully and will provide a great convenience to engineers for the maintenance or upgrade in the life cycle of software systems. Further work is to extend this model to data provenance model so that software engineers can master the control flow and data flow better for maintaining systems.

References

1. Newman, M.E.J.: The structure and function of complex networks. Siam Review 45(2), 167–256 (2003)
2. Zhao, Z., Zhao, L.: Small-world phenomenon: toward an analytical model for data exchange in Product Lifecycle Management. International Journal of Internet Manufacturing and Services 1(3), 213–230 (2008)
3. Valverde, S., Ferrer-Cancho, R., Sole, R.: Scale-free networks from optimal design. Europhysics Letters 60(4), 512–517 (2002)
4. Myers, C.R.: Software systems as complex networks: structure, function, and evolvability of software collaboration graphs. Physical Review E 68(4), 61–76 (2003)
5. Potanin, A., Noble, J., Frean, M., Biddle, R.: Scale-free geometry in OO programs. Communications of the ACM 48(5), 99–103 (2005)
6. Concas, G., Marchesi, M., Pinna, S., Serra, N.: Power-laws in a large object-oriented software system. IEEE Transactions on Software Engineering 33(10), 687–708 (2007)
7. Louridas, P., Spinellis, D., Vlachos, V.: Power laws in software. ACM Transactions on Software Engineering and Methodology 18(1), 1–26 (2008)
8. Guo, Y., Zhao, Z., Zhou, Y.: Complexity analysis with function-call graph on Windows software. International Review on Computers and Software 7(3), 1149–1153 (2012)

Dynamic Variation of Vegetation Fraction for Ion-Absorbing Type Rare Earths Ore in South China Based on Landsat Data——Case Study of Longnnan Rare Earths Mines

Siwen Liu[1,2,**], Hao Wang[3], Jiling Song[3], Xiaolu Fan[3], and Mingzhong Tian[3,*]

[1] National Research Center for Geoanalysis, 100037 Beijing, China
[2] Key Laboratory of Ecological Geochemistry , Ministry of Land and Resources, 100037 Beijing, China
[3]School of Geosciences and Resource, China University of Geosciences, 100083 Beijing, China

Abstract. Ion-absorbing Type Rare Earths Ore mining caused a series of environmental problems. The thesis focused on the study of the district for heavy rare earth in Longnan County, Jiangxi Province, South China. It examined characteristics of spatial-temporal variation for vegetation fraction (*vf*) in study area by 10 images of Landsat TM and ETM+ during 1988 to 2009. The results supported that: (1) there were spatial-temporal differences of *vf* in each mineral district and the areas for medium to high *vf* dominated in mineral districts, but low *vf* distributed on concentrated mining districts; (2) dynamic variation for *vf* differentiated in each mining area each stage, no change areas dominated in mineral districts, but negative change areas only distributed on concentrated mining districts; (3) the history of production, various technologies and preventive measures were the main factors caused spatial-temporal differences of *vf* in mineral districts.

Keywords: TM, ETM+, vegetation fraction, rare earth mineral districts, Longnan, Jiangxi province.

1 Introduction

Normalized Difference Vegetation Index (NDVI) is one of the most important indicators for land use of mineral districts and geomorphic landscape variations which is also an essential reference frame for soil geochemistry investigation and evaluation of land quality [1, 2]. Nowadays, using NDVI to monitor mineral environment are mainly applied in the coal [3-5] and iron ore mining areas [6], but seldom used in other types of mining areas.

Ganzhou with a long history of rare earth mining is well known as "The Heavy Rare Earths Township" around the world. Meanwhile, mining has caused a series of

* The first corresponding author: Mingzhong Tian, Tianmz56@126.com
** The second corresponding author: Siwen Liu, siwenzliu@126.com
The first two authors contributed equally to this paper.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 354–362, 2013.

problems of vegetation deterioration and bedrock exposion. Longnan County, in the south of Ganzhou, with its rare earth mining history more than 30 years , has made tremendous contribution to local and national economic development and laid foundation for its mining industrial system. Meanwhile, rare earth mining has led to a series of environmental issues such as water and soil loss, water pollution and secondary geological calamity and so forth. Wang et al [7] applied multi-source and multi-temporal remote sensing image classification method on evaluating the surrounding environment of rare earths mining areas in the south of Ganzhou, however, there are seldom systemic researches on dynamic change for *vf* in this area. It chose 10 images of Landsat TM and ETM+ from 1988 to 2009 to analyze the spatial-temporal variation of vegetation fraction in rare earths mining areas, Longnan County, which aimed at providing evidences to soil geochemistry investigation, land use evaluation, and soil remediation.

2 Data Source and Methods

2.1 Study Region

Longnan County is situated in the southern most of Jiangxi Province (show in figure 1). Its geomorphic is mainly tectonic denuded hills with basins like beads in the whole

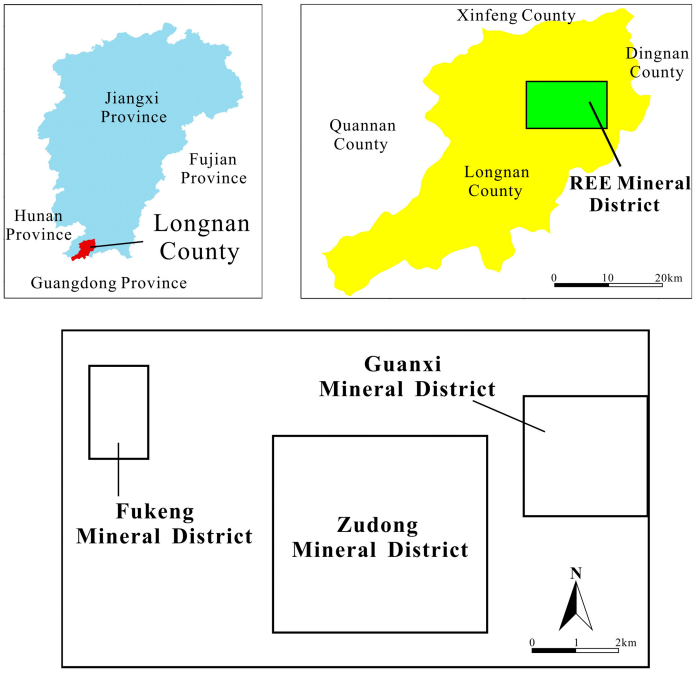


Fig. 1. The Location of Longnan County and 3 REE Mineral Districts

county. The formation lithology for the study region is mostly Yanshannian granites. The weathering crust of granites on the ridge and the top with thickness is nearly 10 to 15 meter which develop iconic rare earth mine. Longnan County is renowned for its heavy rare earths whose reserve accounts for 70% of all over the world. Zudong, Guanxi and Fukeng are the three major mineral districts (show in figure 1). Zudong and Guanxi began mining in 1971, and Fukeng started in 1988. The initial process of producing was pond leaching before 1994, but the in situ leaching technology was widely used after 1994. After 2002, the mines in county basically stopped exploiting because of short of supplies, price decline and government's restriction.

2.2 Data Acquisition and Processing

The data comes from International Scientific Data Service Platform (ISDSP) and Global Land Cover Facility: Earth Science Data Interface (GLCF). The resolution for Band 1 to 5 of TM and ETM+ is 30 meter. All images have gone through systemic radiation correction, geometric correction of ground control points and DEM topographic correction. The specific information of images showed in table 1. Image processing and calculation were under of the ENVI 4.7.

Table 1. the detail information for 10 Images

Image No.	Source	Sensor Type	Data	Path/Row	Mean Cloudiness/%
1	GLCF	TM	16th/Oct/1988	121/43	0%
2	GLCF	TM	9th/Oct/1991	121/43	0%
3	GLCF	TM	5th/Oct/1993	121/43	0%
4	GLCF	TM	18th/Sep/1995	121/43	0%
5	GLCF	ETM+	25th/Dec/1999	121/43	1%
6	ISDSP	ETM+	12th/Oct/2001	121/43	0%
7	ISDSP	ETM+	7th/Nov/2002	122/43	0%
8	ISDSP	ETM+	3rd/Oct/2004	122/43	0%
9	GLCF	ETM+	11th/Nov/2006	121/43	1%
10	ISDSP	TM	2nd/Nov/2009	121/43	0%

2.3 Methods

The value of reflectivity each pixel composes of complete vegetation fraction and non vegetation fraction for combing linearly weighted sum on dimidiate pixel model [8-10], as formula:

$$S=S_{veg}+S_{soil} \tag{1}$$

It defined the contribution rate of complete vegetation fraction is *vf*, and the contribution rate of non vegetation fraction is 1-*vf*. So formula 1 converts

$S=vf \bullet S_{veg}+(1-vf)\cdot S_{soil}$, and $vf=(S- S_{soil})/(S_{veg}- S_{soil})$[11-13]. Similarly, the value for NDVI of each pixel compose of complete vegetation coverage and non vegetation coverage, and it deduces the equation of vegetation fraction (*vf*) and NDVI, that is,

$$vf= (NDVI- NDVI_{soil})/ (NDVI_{veg}- NDVI_{soil}) \quad (2)$$

$NDVI_{soil}$ stands for NDVI value of bare area or non vegetation fraction districts, and $NDVI_{veg}$ stands for NDVI value of complete vegetation fraction districts in the above equation. It used respectively the value of accumulative percent 1% and 99% of NDVI as the value of $NDVI_{soil}$ and $NDVI_{veg}$, and then drew conclusions of vegetation fraction for rare earth mineral districts in Longnan County in different periods.

It used grading standard to vegetation fraction and then defined the thresholds to judge the variations of *vf* [3]. It calculated the average difference between *vf* for the investigated year and *vf* of the previous year, whose results divided by the its standard deviation. The results could be used to vertify whether the variation of vegetation fraction is significant [3, 14]. It defines the value of threshold as 1, when the result is greater than 1 that shows positive change. While it shows negative change if the result is less than -1. And it shows no change when its absolute value is less than or equal to 1. The entire specific calculate processes could be founded in Wei et al [3] and Morawitz et al [14].

3 Results and Analysis

3.1 The Characteristics for Spatial-temporal of Vegetation Fraction

The article calculated the *vf* value in 3 mineral districts according to equation 1 and equation 2, and then graded vegetation fraction in different period. The *vf* value is less than 30% that indicates low vegetation fraction .And the *vf* value is equal or greater than 30% that indicates medium to high vegetation fraction [13]. The result showed as figure 2.

The areas for medium to high *vf* dominated in each mineral district, though the percentage for areas of low *vf* was small nearly 20 years in figure 2. Besides, there were spatial-temporal differences of *vf* in each mineral district. The areas for low value of *vf* with the characteristics of stability and tiny varied range were the concentrated mining districts in Zudong, Guanxi and Fukeng, although the area proportion and spatial position were diverse. The variation tendency for both areas of low *vf* and medium to high *vf* in each mineral district was fluctuant and opposite in 1988 to 2009. In addition, the maximum of the relative areas for medium to high *vf* appeared in 1995.

The district for low *vf* in Zudong was mainly in the north-center of the mining area, which was also the concentrated rare earth mining district for the whole Longnan County. The distribution trend for low *vf* in west and south of Zudong mineral district decentralized in 1995 to 2001, which might be relevant to the increase of the illegal mining actives, however, the extent of variable for *vf* area became small in the period.

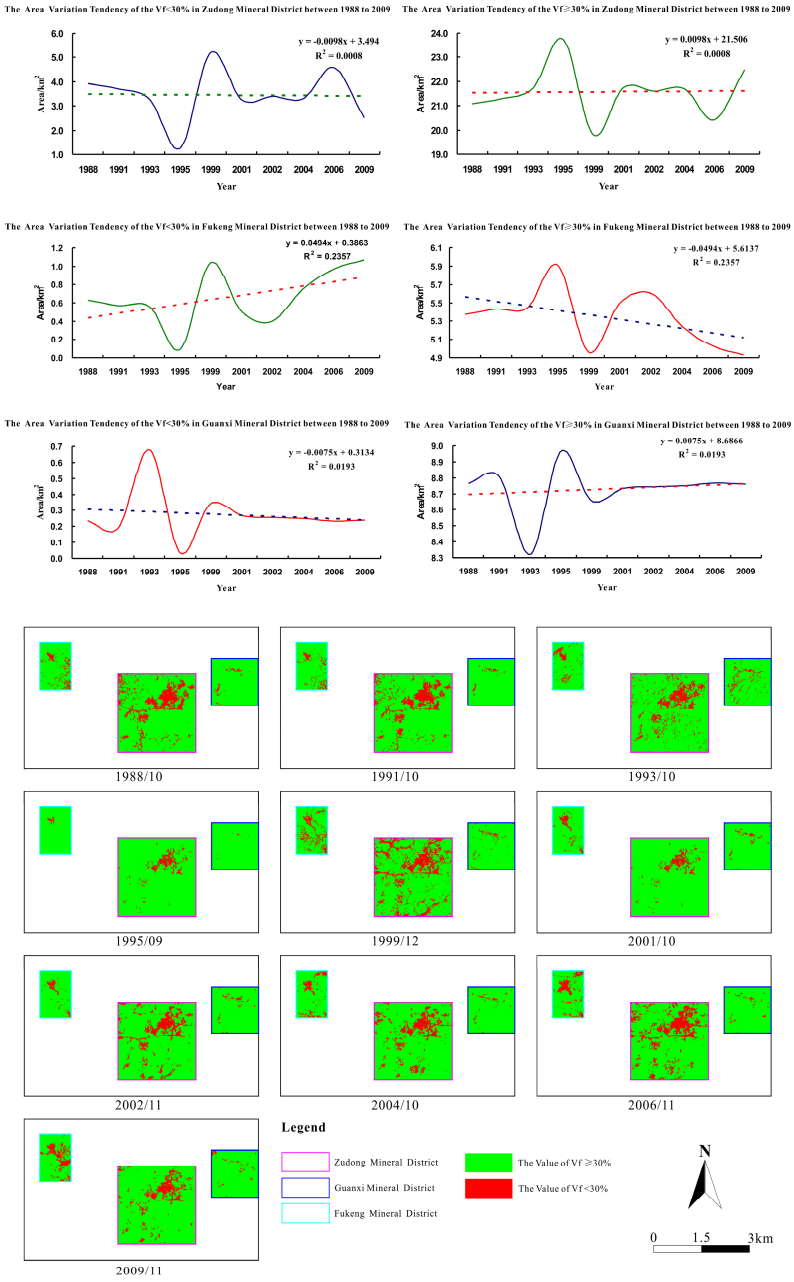


Fig. 2. The Area Variation Tendency of the different Value for *Vf* in Mineral Districts and Spatial-temporal Variation of Vegetation Fraction of 3 Mineral Districts in Longnan Country during 1988 to 2009

In 1995 to 2001, the area for low *vf* varied significantly in Zudong. The area for low *vf* in west and south enlarged obviously in Zudong in 1999, although reduced in 1995 and 2001. The area for low *vf* in west and south of Zudong increased by a slight in 2002 to 2006, however, the tendency was not very significant in 2009. The districts for low *vf* concentrated in the north-center of Guanxi, which distributed dispersedly in southwest at the same time. Not only was the value of low v*f* was most dispersed but its area were largest in 1993 during 1988 to 2009. In general, the *vf* in Guanxi was relatively steady that means its small variable extent and low vegetation deterioration. The low *vf* mainly distributed in the north-center and southeast in Fukeng. In 1988 to 1999, the low *vf* diffused from southeast to west in Fukeng, however, the tendency became not very obvious in 2001. The low *vf* spreaded from southeast to both northeast and south in 2001 to 2009. Generally speaking, the areas for medium to high *vf* in both Zudong and Guanxi increased by a slight nearly 20 years, whose relative area extent of the variable were respectively 15.82% and 7.17%.The areas for medium to high *vf* in Fukeng declined inconspicuously and the extent of the variable for relative areas was 16.33%.

3.2 Analysis to Dynamic Variation for Vegetation Fraction

According to the time limit for the conversion of mining technologies and stop mining, remote sensing images could be classed as 3 periods as follows: the first stage, the period for pond leaching (refer 1988, 1991 and 1993); the second stage, the period for in situ leaching (refer 1995.1999 and 2001); the third stage, the period for stop mining (refer 2002, 2004, 2006 and 2009). It calculated the average of each period as the *vf* for the mineral district in the stage by using the same path and row images. The methods from reference 3 and reference 14 were used to calculate dynamic variation for *vf* in each mineral district for 3 stages. Positive change shows the vegetation growth in good condition and the gross increases. On the contrary, negative change indicates the vegetation growth in poor condition and the gross decreases. In addition, no change reflects the vegetation growth in steady condition. All the results showed as figure 3.

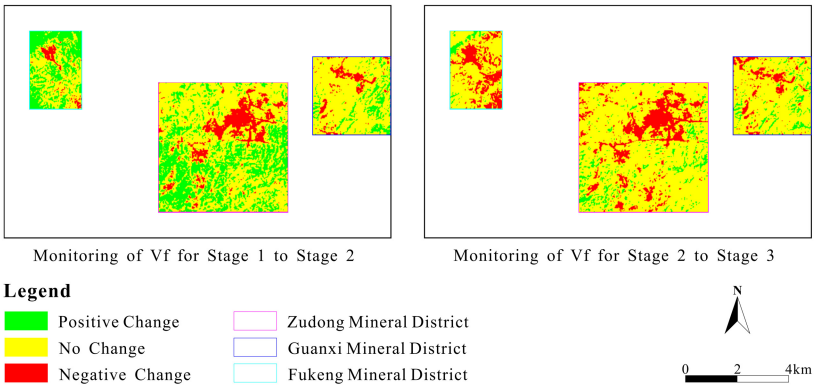


Fig. 3. The Dynamic Change for Vegetation Fraction in Different Stages

Figure 3 showed that the difference for dynamic variation of *vf* in each mining area in 3 stages. From stage 1 to stage 2, the districts for no change mainly dominated in the north and southeast of Zudong, whose areas took up 59.40% of the whole mineral district. The districts for the negative change whose areas took up 11.24% of Zudong only concentrated in core exploitation areas, besides, contributed dispersedly in west and southwest. Positive change areas were mainly in south-center whose areas took up 29.36%. The districts for negative change concentrated in north-center whose areas were equal to the districts for low *vf* and took up 9.05% of Guanxi. Others are positive change districts and no change districts, whose areas took up 11.44% and 79.51% respectively. The areas for no change and positive change of *vf* took up respectively 50.18% and 43.75% of the whole Fukeng. Similarly, districts for negative change concentrated on core exploiting mining, whose areas took up 6.06%.

From stage 2 to stage 3, the districts for negative change in Zudong expanded to south and northwest whose areas took up nearly 16.52%. The districts for no change took up most parts of the mining districts, but the districts for positive change only distributed sporadically in east and southwest. The districts for negative change spreaded to northwest whose area took up 11.35% of Guanxi. The areas for no change and positive change took up respectively 81.98% and 6.16%. The districts for negative change distributed in the core mining and most part for east of Fukeng, whose area took up 27.96%. The districts for positive change distributed mainly in the northwest whose area only took up 3.99% of the whole mineral area.

4 Discussion

The reasons for the spatial-temporal discrepancies of vegetation fraction in each mineral district were concluded in the paper as follows based on field investigations and relative materials.

The production life for each mineral district was different. The longer for mines were opened up, the more serious the environmental problem showed. The relative area for low *vf* in Zudong and Guanxi was larger than Fukeng. The relative area for negative change of *vf* in Zudong was 5.18% larger than Fukeng that reflected the production life effected the dynamic variation for *vf*.

Different mining technologies resulted in various influences on *vf*. In situ leaching technology damaged the vegetation in only limited locality instead of large scale. The vegetation recovered rapidly in non concentrated mining districts in the early period (refer around 1994) due to mining technology change. The area for medium to high *vf* in each mineral district extended and low *vf* merely limited in core exploitation areas from 1993 to 1995. Acid leaching-ore chemicals broke down stress balance of the origin rock-soil which aggravated soil erosion and vegetation degradation in the middle and later periods for using in situ leaching technology. The situ leaching technology caused less disruption for vegetation with stopping mining rare earth gradually during 2000 to 2002. Meanwhile, the relative area for medium to high vegetation fraction has increased slightly in mineral districts in the period.

Relative comprehensive preventive measures influenced *vf*. Comprehensive manage measures have been taken in the mining environments in Longnan County since 1983, however, the recovery for vegetation was not significant. The local has started first stage project on rare earth mines environmental protection since 2007 which played a role in vegetation recover in some degree. The relative areas for low vegetation fraction in Zudong and Guanxi decreased from 2006 to 2009, however, the relative areas for medium to high vegetation fraction decreased in Fukeng due to its fragile environment and the increase of the illegal mining actives.

5 Conclusions

The paper made the following preliminary conclusions based on above-mentioned research for vegetation fraction of rare earth mineral districts in Longnan County.

(1) There were spatial-temporal differences of vegetation fraction in mineral districts. Mining areas usually manifested low vegetation fraction. The total area for low vegetation fraction declined slightly in Zudong and Guanxi nearly 20 years. On the contrary, it increased slightly in Fukeng.

(2) The dynamic variation for vegetation fraction varied from each mining district. The vegetation fraction in mining areas showed negative change that indicates the vegetation deterioration and the gross reduction. But the vegetation fraction in non mining area showed no change reflected that the growth status of vegetation is steady.

(3) The major factors caused spatial-temporal differences of vegetation fraction in mineral districts are the length of production life, different mining technologies and preventive measures. The longer mining history, the more severely the vegetation was destroyed. In situ leaching technology merely played a prominently role in vegetation recovery in the early period. Preventive measures improved the vegetation fraction in mineral districts to a certain degree though not very obviously.

Acknowledgement. This study was supported by Geological Survey Project of China (No. 121201130027).

References

1. Lu, F., Yue, D.P., Guo, X.: A Review on Application of Normal Different Vegetation Index. Forest Inventory and Planning 34(2), 48–52 (2009)
2. Xiaoke, Z.: Research on Temporal and Spatial Variation for Land Cover in Mining Disturbed Zone. Master Degree, Thesis for Taiyuan University of Technology, pp. 1–65 (2010)
3. Wei, S., Sun, Z.P., Li, D.L., et al.: Monitoring of Temporal and Spatial Change of Vegetation in Waste Dump of Haizhou Opencast Coalmine Area Using Multi-temporal Landsat Remote Sensing Images. Acta Ecologica Sinica 29(11), 5860–5868 (2009)
4. Lixin, W., Ma, B.D., Liu, S.J.: Analysis to Vegetation Coverage Change in Shendong Mining Area with SPOT NDVI Data. Journal of China Coal Society 34(9), 1217–1222 (2009)

5. Guangjun, W., Hu, Z.Q., Du, H.Q., et al.: Analysis of Grassland Desertification Due to Coal Mining Based on Remote Sensing——An Example from Huolinhe Open-cast Coal Mine. Journal of Remote Sensing 10(6), 917–925 (2006)
6. Baodong, M., Wu, L.X., Liu, S.J.: Variation Detecting of NDVI for Mine Considering the Seasonal Phase Difference——Taking Qidashan Mine as a Case. The Proceedings for China Association of Remote Sensing Applications in 2010 (2010)
7. Tao, W., Liu, Y.Y., Wang, P., et al.: Application of Multi-source and Multi-temporal RS Image Classification Method in REE Mineral Development and Environment Variation in Ganzhou Mineral District. China Mining Magazine 18(11), 88–91 (2009)
8. Hongyu, S., Wang, C.Y., Niu, Z.: Analysis of the Vegetation Cover Change and the Relationship between NDVI and Environmental Factors by Using NOAA Time Series Data. Journal of Remote Sensing 2(3), 205–210 (1998)
9. Yuecong, Z., Zhao, Z.Q., Li, S.C.: Indicating Variation of Surface Vegetation Cover Using SPOT NDVI in Northern Part of North China. Geographical Research 27(4), 745–754 (2008)
10. Purevdorj, T.S., Tateishi, R., Ishiyam, T., et al.: Relationships between percent vegetation cover and vegetation indices. Remote Sense 19, 3519–3535 (1998)
11. Tao, C., Li, P.X., Zhang, L.P.: Dynamic Analysis of Vegetation Fraction Change in Wuhan Region from 1988 to 2002. Remote Sensing Technology and Application 23(5), 511–516 (2008)
12. Changguang, W., Zhou, Z.X., Xiao, W.F., et al.: Dynamic Monitoring of Vegetation Coverage in Three Gorges Reservoir Area Based on MODIS NDVI. Scientia Silvae Sinicae 48(1), 22–28 (2012)
13. Miaomiao, L.: The Method of Vegetation Fraction Estimation by Remote Sensing. Master Degree Thesis for Chinese Academy of Sciences, pp. 1–110 (2003)
14. Morawitz, D.F., Blewett, T.M., Cohen, A., Alberti, M.: Using NDVI to Assess Vegetative Land Cover Change in Central Puget Sound. Environmental Monitoring and Assessment 114, 85–106 (2006)

Study on Key Links of Data Release in Chinese TMC

Yuan Liu[1,2,*], Liang Wang[1], Rong Zhao[1], and Hai Tan[1]

[1] Research Center of Government GIS, Chinese Academy of Surveying and Mapping, No.28 Lianhuachi West Road, Haidian District, Beijing, China
liuyuan322@gmail.com
[2] Faculty of Geomatics and Geographic Information, Lanzhou Jiaotong University, Lanzhou 730070, China

Abstract. The government and enterprises, in China, set out to cope with the congestion problem using effective solutions about intelligent transportation systems. Not only do they slow down transportation pressure but also make response to green travel, energy conservation and emission reduction. This is a new digital ecological engineering. Traffic Message Channel is regarded as a complete set of Intelligent Transportation System in releasing the dynamic information in real-time manner. It is currently the most popular solutions about transportation congestion. It makes great contribution for digital ecological engineering. This paper will analyze the process of data release. It is crucial to think about new contents of key links of data release in a complex environment of Chinese special transportation infrastructure construction and the uncertain collecting way. There are three main aspects such as data fusion, data matching, data releasing. We put forward reasonable approaches to deal with problems that we encountered by in-depth research on three main aspects. The experiments verified the feasibility of schemes which can improve the accuracy and efficiency about TMC traffic information releasing and provide the technical reference.

Keywords: ITS, TMC, data fusion, data matching, data releasing.

1 Introduction

The government and enterprises, in China, set out to cope with the congestion problem using effective solutions about ITS. Not only do they slow down transportation pressure but also make response to green travel, energy conservation and emission reduction. Real-time traffic dynamic information system is critical system in ITS. Therefore, people carry out a series of high-tech electronic technology such as VICS [1], TMC [2]. The TMC enriches connotation of GPS navigation system and GIS maps in terms of functions and applications.

The TMC technology originated in Europe, nowadays, becomes one of the major European traffic information service modes. It makes a great contribution to solving the traffic jams and meeting the higher demands of travel demand [3].At the same time, TMC FORUM of its attached product promotes the TMC standardization and industrialization. So far, it is the only certified international standard in traffic

[*] Yuan Liu (1986 -), she is from the city of Hengyang in Hunan. The major of research is GIS.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 363–369, 2013.

information standardization industry [4].In 2005, ERTICO brought RDS-TMC in China. With the development of it, it has ensured that a series of large-scale activities such as Beijing Olympic Games in 2008, Shanghai World Expo in 2010, and so on. However, TMC in China is immature. For adapting to complexity in China, people have to put forward new solutions and relevant measures to deal with unexpected situations. Therefore, we need to conduct in-depth research about the key links of data release, and put forward new problems and solutions.

2 A Basic Processing of Data Releasing in TMC

According to the principle of RDS-TMC [5, 6], data releasing follows several steps (Figure 1).

(1) Gaining data sources. All kinds of traffic information will be unified to a processing platform for operating, analyzing and processing. The data accessed can be divided into two kinds of data sources (internal data about point source and external data about line source).Firstly, point sources are FCD data from vehicles equipped with GPS. Secondly, we inert partners' traffic information sources of relevant cities. Thirdly, we collect real-time traffic information using personal acquisition system.

(2) Setting up Location Table of traffic information in the range of a region or city by ourselves. It is the regional or urban road information. A traffic unit consists of code of a location point and its offset direction [7]. These data take or receive relevant information and are independent of map version.

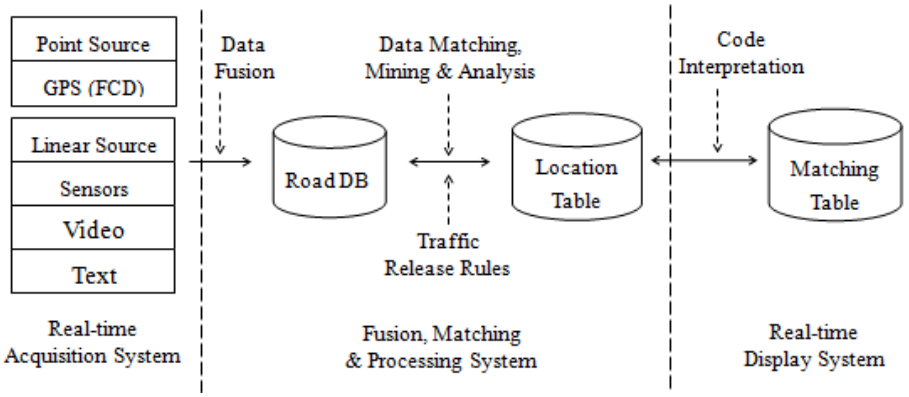


Fig. 1. Transmission data process in TMC

(3) Loading Matching Table of a certain version of the TMC traffic in a city or an area. For realizing spatial display, spatial analysis and processing in the processing platform of traffic information, we need to match them to the underlying road database. So the process needs to build a matching relationship between location points and roads in the basic electronic map.

(4) Coding interpretation and feedback. When processing platform of traffic information has completed processing data, it could send TMC message to one or more of clients. Standard TMC user messages provide five basic items of broadcast

information [8]. When clients receive TMC message, they employ MT to reconstruct the original message. The forms could be video images or text feedback.

3 Discuss Key Links of Releasing Traffic Information in TMC

3.1 Integrating a Variety of Traffic Information Data Sources

According to different data formats and standards provided by the traffic data suppliers, there are two methods to implement data source integration:

(1) Integrating internal data. At present, we directly use of current popular FCD detection technology to realize internal information fusion of electronic map.

(2) Integrating external data. External data fusion mainly uses custom mapping table technology. Because of uncertainty of storage format of the data, the difficulty of the data fusion will be more and more important. The core idea is to create the mapping relationship between customers' data and their own data of electronic map. There are two ways to solve problems.

a) Rely on customers' traffic flow information.
All kinds of sensors (such as "microwave detector", etc.) and "video detector" are common professional equipment of collecting data. It is critical to create a corresponding set of mapping table for each data source provider. We need to match a corresponding set of mapping table according to serious of sensors' message. However, these relationships between both of them are one-to-one, one-to-many, many-to-one, many-to-many, etc. Figure 2 is that an example about customers' freeway traffic data integrated into the corresponding services' system in Beijing venues of 2010 Shanghai World Exp.

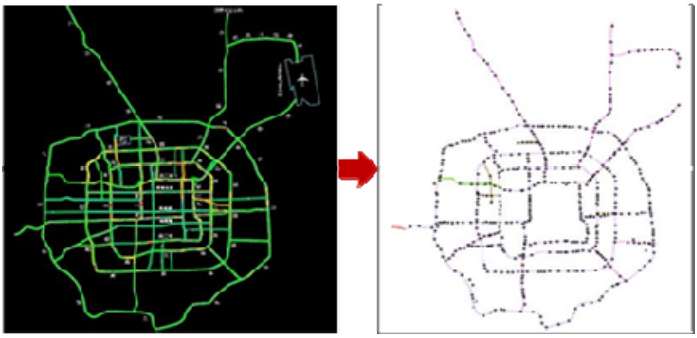


Fig. 2. Customers' freeway traffic data integrated into unified servers' system

b) Rely on listening to the radio, the service hotline about traffic accidents.
Its existence is characterized in the form of text broadcasted, which makes up for blind sides of sensors. It is difficult to fast search corresponding to the intersection between the sections. We need to unify data access specification in order to avoid inputting disordered and wrong data. Therefore, before accessing this kind of data, we need to subdivide roads of the city into details. You can refer to the following instance (see Table 1). It establishes a certain section between two road crossings.

Table 1. An example that one text transformed into a certain links

Road name	Direction	Start Point	End Point	Link ID	ID	Length
A	W-E	B	C	595662003216	1	84
A	W-E	B	C	595662003247	2	17
A	W-E	B	C	595662003250	3	51
A	W-E	B	C	595662003251	4	158

Multi-channel data fusion also leads to overlapping redundancy or different conflicts. Hence, we take effective measures to realize accurate information sharing, which is to be thinking.

3.2 Creating Location Table

Location Table of TMC can determine the range of data release and label. The production rule of LT has been relatively complete, but some principles of LT are not applicable in China. There are classic characteristics that Location Tables layout different content. The TISA distributes 31 location table codes to China for its vast territory.

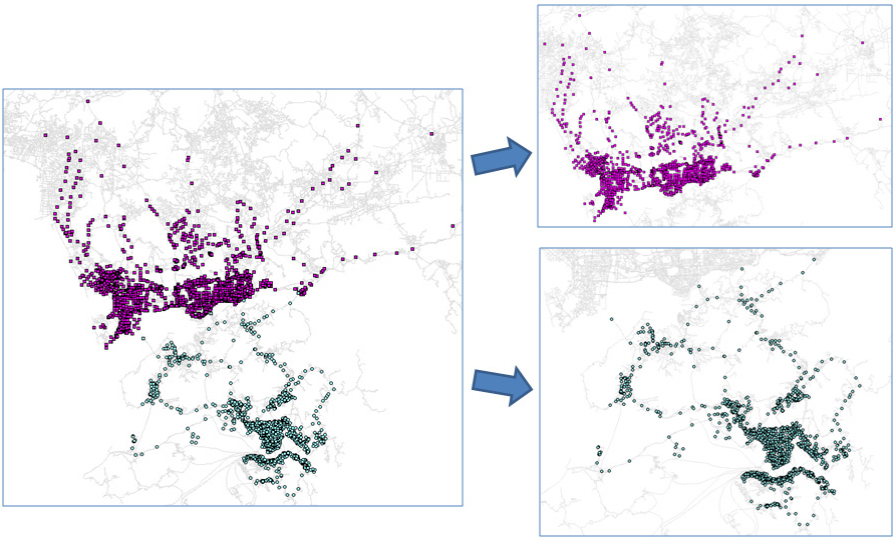


Fig. 3. Separation and Mergence Results

One table may contain a city or a province, or even several provinces. It can control well in various local cities, but increases cost about production, management and version maintenance, and relations of road links among location tables may be missed. Nowadays, we are able to abstract effectively different city from an enormous table using relevant specification, and automatically match intersection of the same position (see Figure 3).Figure 3 is that LT of Shenzhen (pink points) and LT of Hong Kong (green points) are merged (left) or are separated (right).

3.3 Generating Matching Table

After making LT in TMC, we need to load real-time traffic flow data to TMC messages. The key in this process is to establish a relationship between location table and road ink in the electronic navigation map. Therefore, establishing a matching table is establishing a two-way mapping between location table and electronic map. Different electronic map data is banded its unique Matching Table [4]. MT is used to match every traffic unit of location table. The following example shows the process of matching the location point to the electronic map links (see Table 2、Figure 5).

Table 2. The unit of Matching Table related to Location Table

Mesh	Link ID	Table Code	LCD	Direction
595672	595672556789	32	16	+
595672	595672556790	32	16	P

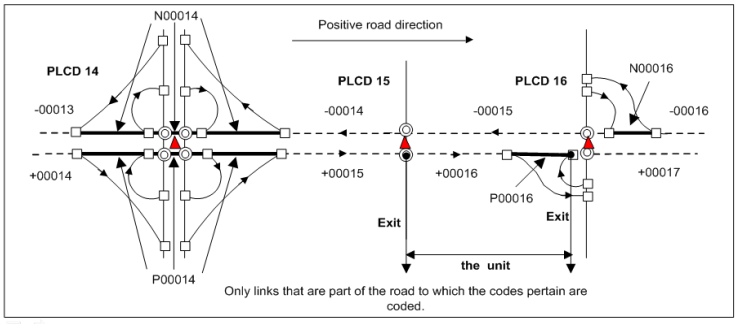


Fig. 4. The process of matching the location point to the electronic map links

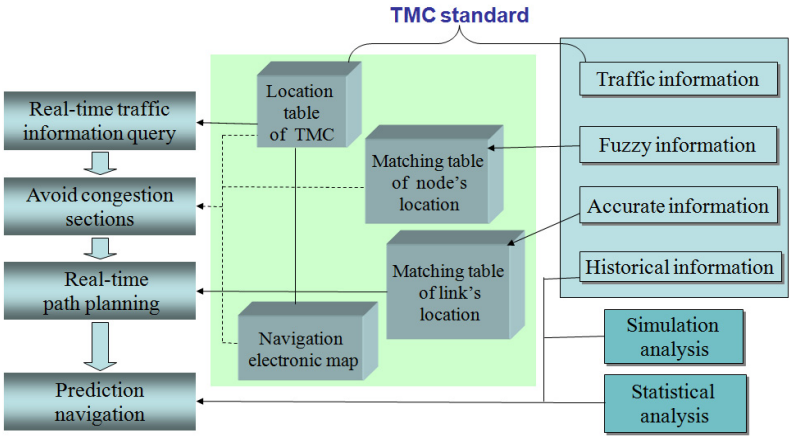


Fig. 5. A diagram of the relationship between LT and MT in application of TMC

Figure 5 is a diagram of the relationship between the TMC coding and navigation electronic map. You can see that matching table which connects the location table and the traffic link guides the process of travel.

3.4 Releasing More Tenses Data

As time goes on, some road will be rectified, newly built, abandoned, or changed road names. We need to change LT and MT to ensure that data is accurate and consistent. It appears the following questions:

(1) Adding location point. The new version releases some codes which customers can't be received by old customers.
(2) Deleting location point. The new version does not release the old position point codes.
(3) Changing location point. The new version of information released and the old version of information released is not consistent.
(4) There are other secondary attributes which do not affect the overall information such as road level, road name, and so on.

The 1), 2), 4) update items are a kind of regular update. But 3) update item often results in the inconsistency between two different versions of maps. When a new version is released, it doesn't get updated at once because of a long and cyclical renewability for an old version. We put forward a way to establish a differential relationship between two different versions in order to realize the accurate interaction between data. Table 3 is an associated form created for them. Generally, updated medium-sized city doubles the efficiency, and data accuracy also will be greatly improved.

Table 3. A difference table about two different versions of LT

ID	Mesh	Version	Level	LCD	Direction	Length	Type	Description
1	595672	8/9	4/4	95/19|20	-/-	90/27|63	4	Divide
2	595672	8/9	3/1	2/13	+/+	45/45	1	Level*
3	595672	8/9	4/4	20/20	-/+	50/50	2	Direction*
4	595672	8/9	2/2	31/201	-/-	78/113	5	Combine
5	595672	8/9	2/2	32/201	-/-	35/113	5	Combine
6	595672	8/9	4/	109/	+/	33/	3	Delete

Note: "*/*"is compared between previous version and next version (previous version/ next version). "Level*"is meaning of changed level; "Direction*"is meaning of changed direction.

4 Conclusions

TMC has been widely used in the Internet, in-car devices and others. By researching the key links of data release in TMC, by and large, we solve a series of problems in

the process of data transmission, and can greatly improve the efficiency and quality of work. Although TMC in China still have an advantage in a great degree, its display effect and intensity, compared with RTIC [9] and VICS, are slightly worse. It is necessary to increase the density of the road network and enhance visual effects. At present, with the development of networking technical and vehicular networking technology, ITS will have a wider application prospect. Finally, it can ensure that the information is obtained in real time, communicated in effective and accurate methods. It is worth further research and verification [10].

Acknowledgements. This paper is funded by National Science & Technology Pillar Program of P.R. China under grand No.2012BAH24B02, the research and demonstration on key technologies of geographic information monitoring in open sea reefs under grant No. 2012BAB16B01.

References

1. Li, H.-H., Liu, D.-M., Wang, J.: Latest Development of VICS System in Japan. Communications Standardization 15, 107–112 (2011)
2. Huang, Z.-Q., Zheng, Y.-Z.: The FM subcarrier and its application in RDS-TMC. Foreign Investment in China 247, 125–126 (2008)
3. Gao, Y., Wen, H.-M.: Technique and Standardization Research of Radio Data System-Traffic Message Channel (RDS-TMC). Journal of Transportation Systems Engineering and Information Technology 7, 49–54 (2007)
4. Han, J.-J., Zhang, L.-D., Liu, D.-C., Zhang, F., Zheng, Y.-P.: Research and Design of Real-Time Traffic Information System: GPRS-TMC. Control Engineering of China (suppl.), 117-119 (2009)
5. Xu, J.-P.: Intelligent Transport System oriented Research on Spatial Data Mining. Tongji University, Shanghai (2007)
6. TISA, TMC forum (2005), `http://www.tmcforum.com`
7. GB/T 20612.3-2006, Traffic and Traveler Information (TTI)-TTI messages via traffic message coding Part 3: Location referencing for ALERT-C (2006)
8. Qin, X.-J., Lan, B., Li, J.-H., Wang, G.-L.: Dynamic Vehicle Navigation System Based on RDS-TMC. Application Research of Computers 5, 182–184 (2006)
9. Li, J.-H.: Real-time traffic information system based on RDS. Digital Communication World 4, 60–63 (2011)
10. Wu, Z.-Q., Liang, Z.-X., Zeng, X.-M.: A research on the C/S intelligent transportation real-time dynamic guidance system based on wireless network construction. Journal of Foshan University (Natural Science Edition) 29(4), 55–60 (2011)

Grassland NDVI Response to Climate Factors in Different Vegetation Regionalizations in China

Shaohua Liu[1], DengHua Yan[1], XiaoLiang Shi[2], Gang Wang[1], Zhe Yuan[1], and Jun Yin[1]

[1] State Key Laboratory of Water Cycle Simulation and Regulation in River Basin,
China Institute of Water Resources and Hydropower Research,
Beijing 100038, China
[2] Information Center,
Yellow River Conservancy Commission of the Ministry of Water Resources,
Zhengzhou 450004, China

Abstract. On the basis of the daily precipitation and mean temperature data of 583 meteorological stations, and AVHRR GIMMS NDVI materials from 1982 to 2006, the NDVI (annual maximum NDVI), annual precipitation and annual accumulated temperature above 10℃ (AT10) of grassland in vegetation regionalizations were obtained in China. Then the Mann-Kendall test and linear trend analysis were applied into the trend analysis of NDVI, precipitation and AT10 on per-pixel and vegetation regionalization scales respectively. And the correlation analysis and binary linear regression analysis were used to study the relation between NDVI and precipitation and AT10. The result showed that the NDVI of grassland was increasing significantly in the subtropical broadleaf evergreen forest regionalization, while the precipitation and AT10 were decreasing. And the NDVI of grassland are positive related to the precipitation and AT10 in temperate steppe and desert regionalization. The NDVI, precipitation and AT10 of grassland were increasing in Qinghai-Tibet plateau alpine vegetation regionalization. However, the NDVI-precipitation correlation and NDVI-AT10 correlation were negative and positive respectively.

Keywords: NDVI (Normalized Difference Vegetation Index), the accumulated temperature above 10℃ (AT10), precipitation, spatial and temporal variation, vegetation regionalizations, grassland in China.

1 Introduction

The relation between climate change and vegetation dynamics are complex in global and regional scales [1,2,3,4]. The spatial and temporal variation of vegetation is research hotspots under the global warming condition. And Normalized Difference Vegetation Index (NDVI) has been effectively used in vegetation dynamics monitoring and the study of vegetation responses to climatic changes at different scales during the past few years [5,6,7,8]. Precipitation and temperature are two key climate factors during the climate changes and most important climate factors for the growth of vegetation. Climate warming has an huge effect on the distribution of precipitation and

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 370–380, 2013.

temperature in China [9,10,11,12,13], and NDVI would be impacted immediately. So studying the relation between climate change and vegetation dynamics is helpful to predict the change of vegetation and protect the ecology.

Many works have been done on the relation between NDVI and climate factors (precipitation and mean temperature) [14,15,16]. However, few works focus on the accumulated temperature, which can represent the thermal condition better than mean temperature during the growth of vegetation. 10℃ is the most suitable initial temperature for vegetation growth in China [17], thus the accumulated temperature above 10℃ (AT10) is an important index of the heat condition, which affects the temporal and spatial distribution of vegetation directly [18,19]. Additionally, the respond of NDVI in regions with different vegetation type and underlying surface to climate change is very different [20,21]. Therefore, it is necessary to analyze the relation between NDVI and climate factors on the region with approximate underlying surface and climate.

The grassland resource is rich in China, and most grassland is located in the northwest of China, where the density of population is low and industrial is underdeveloped. It can reduce the human activity effect on the respond of NDVI to climate change. In this study, the grassland was extracted according to the land-use data of China. And in order to eliminate the effect of underlying surface on the respond of NDVI to climate change, the grassland is divided into different study areas based on the vegetation regionalizations. The purposes of the study are: (1) to detect the trend of NDVI, precipitation and AT10 from 1982 to 2006 for grassland in China and the spatial distribution of the trend; (2) to study the correlation coefficient between NDVI and climate factors (precipitation and AT10) of grassland in different vegetation regionalizations and the spatial distribution of the correlation coefficients; (3) to detect the different respond of NDVI to the change of precipitation and AT10 in different vegetation regionalizations; (4) to construct the binary linear regression between NDVI and climate factors of vegetation regionalizations.

2 Materials and Methods

2.1 Materials

2.1.1 NDVI Dataset

The NDVI dataset in this study are GIMMS (Global Inventory Monitoring and Modeling Studies) NDVI (Normalized Difference Vegetation Index) dataset, which have been corrected for calibration [22]. And it was derived from the AVHRR (Advanced Very High Resolution Radiometer) land dataset at a spatial resolution of 8 km×8 km and 15-d interval from the Global Inventory Monitoring and Modeling Systems group. The annual NDVI dataset from 1982 to 2006 was obtained using the Maximum Value Composite (MVC) method, which selects the highest observation for each pixel from a predefined compositing period to represent the current period. This approach was based on the logic that low-value observations are either erroneous or have less vegetation vigor for the period under consideration [23].

2.1.2 Meteorological Materials

The meteorological materials included daily mean temperature and precipitation from 1982 to 2006 of 583 meteorological stations from the China meteorological data sharing service system (http://cdc.cma.gov.cn/). Annual accumulated temperature above 10℃ (AT10) and precipitation were calculated by the daily meteorological materials, and the overall country grid data of annual precipitation and AT10 with the resolution of 8 km×8 km was obtained from the 583 meteorological stations by the original kriging interpolation with ArcGIS 9.3 software. The location of the 583 meteorological stations are shown in Fig.1.

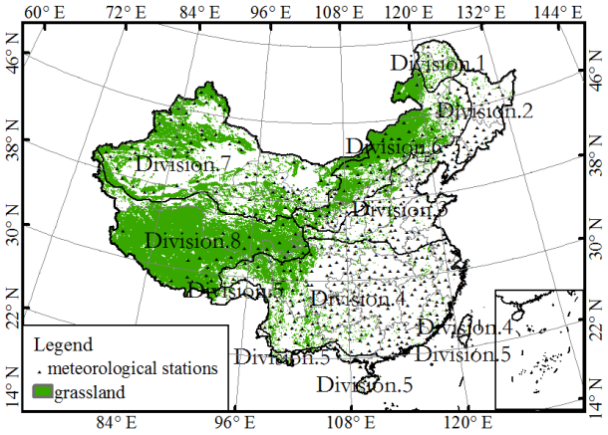


Fig. 1. The distribution of grassland and vegetation regionalizations and location of 583 meteorological stations

2.1.3 Landuse Materials and Vegetation Regionalizations

The 2000 land use map with the resolution of 1 km×1 km was used in the study, which was obtained from Data Center for Resources and Environmental Sciences Chinese Academy of Sciences (RESDC) (http://www.resdc.cn). Landuse were divided into two parts (grassland and non-grassland), and the two parts were resampled to a spatial resolution of 8 km×8 km respectively. In order to eliminate the mixed pixel and improve the purity of grassland, overlaps grids produced by resampling were eliminated from the grassland grids. Finally the distribution of grassland is shown in Fig.1. The data on vegetation regionalizations were digitized from the 1:1,000,000 vegetation regionalizations map of China mapped by the editorial committee of vegetation map of china, chinese academy of sciences. According to the vegetation regionalizations, China can be divided into 8 regions, which are shown in Fig.1. The names of vegetation regionalizations and area of grassland are shown in Table 1. The grassland was mainly distributed in the Division.4 and Division.6-8. Therefore, this study focused on the grassland NDVI in Division.4 and Division.6-8. The NDVI, precipitation and AT10 of Divisions are the average value of every pixel in the grassland Divisions.

Table 1. Areas of grassland in vegetation regionalizations

Name of vegetation regionalizations	Area/10,000 km^2	Division	Choose
Cool temperate and temperate needleleaf forest regionalization	4.81	Division.1	×
Temperate mixed needleleaf and broadleaf deciduous forest regionalization	1.35	Division.2	×
Warm temperate broadleaf deciduous forest regionalization	9.02	Division.3	×
Subtropical broadleaf evergreen forest regionalization	42.31	Division.4	•
Tropical (monsoon) rain forest regionalization	3.67	Division.5	×
Temperate steppe regionalization	67.19	Division.6	•
Temperate desert regionalization	97.36	Division.7	•
Qinghai-Tibet plateau alpine vegetation regionalization	141.66	Division.8	•

2.2 Methods

2.2.1 Mann-Kendall Test

As a non-parametric test method recommended by the World Meteorological Organization, Man-Kendall test method has been widely used in the analysis on the time-series changes of temperature, precipitation, runoff [24,25,26] and remote sensing image [27]. In this study it was applied for the trend tests of NDVI, precipitation and AT10 for grassland Divisions. With the Mann-Kendall test, suppose that an observed sample data set (x_i, i=1,2,…,n) is available, Z follows the standard normal distribution based on the equation as follows:

$$z = \begin{cases} \frac{S-1}{\sqrt{Var(S)}} & S > 0 \\ 0 & S = 0 \\ \frac{S-1}{\sqrt{Var(S)}} & S < 0 \end{cases} \quad (1)$$

Where $S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sgn(x_j - x_i)$, $sgn(x_j - x_i) = \begin{cases} 1 & x_j - x_i > 0 \\ 0 & x_j - x_i = 0 \\ -1 & x_j - x_i < 0 \end{cases}$,

$Var(S) = \frac{n(n-1)(2n+5)}{18}$. The null hypothesis that there is no trend is rejected when the computed Z value is greater than Za/2 in absolute value. And when Z > 0, it means the trend of the data set has an increasing trend. Otherwise, it will be a decreasing trend.

What's more, the significance of β is assessed based on the null distribution of slope, it can be computed by the equation (3) to represent the magnitude of trend.

$$\beta = median\left(\frac{x_j - x_i}{j-i}\right) \quad (1 \le i < j \le n) \tag{2}$$

2.2.2 Linear Trend Analysis

The linear trend was estimated by regressing it as a function of time over the study period. The linear trend of yearly maximum NDVI, annual precipitation and AT10 from 1982 to 2006 on a per-pixel basis (8km×8 km) were examined. Expression is shown in (3) [8].

$$\theta_{slope} = \frac{n \times \sum_{j=1}^{n}(j \times x_j) - \left(\sum_{j=1}^{n}(j)\right)\left(\sum_{j=1}^{n}(x_j)\right)}{n \times \sum_{j=1}^{n} j^2 - \left(\sum_{j=1}^{n}(j)\right)^2} \tag{3}$$

Where x_j means the available in j year; n is the length of the variable, and n = 25 in this study; θ is the linear trend slope. Then to analyze the effects of climate factors on NDVI changes, NDVI-precipitation and NDVI-AT10 correlation were determined on a per-pixel basis from 1982 to 2006. Expression is shown in (4).

$$r_{xy} = \frac{\sum_{j=1}^{n}(x_j - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_{j=1}^{n}(x_j - \bar{x})^2}\sqrt{\sum_{j=1}^{n}(y_j - \bar{y})^2}} \tag{4}$$

where x_j and y_j mean the two related variable in j year. $\bar{x}$ and $\bar{y}$ represent the mean value of two variable in n years. r_{xy} is the correlation coefficient between the x and y; $r_{xy} > 0$ means x is positive related to y; $r_{xy} < 0$ represent x is negative related to y. And the absolute value of r_{xy} is bigger, the correlation between the x and y is better.

3 Results and Discussions

3.1 Trends of NDVI, Precipitation and AT10

The trend of NDVI, precipitation and AT10 are different in vegetation regionalizations with different climate and underlying surface. The grassland NDVI of Division.4 had a significant decreasing trend (5% level), but the grassland NDVI of Division.6-8 showed an increasing trend (table.2). The linear slope of grassland NDVI in Division.6 was the biggest. The result was consistent with the reality that the desertification in some parts of north China has been reversed since the 1980s [28,29,30]. The precipitation was decreasing in Division.6, but the precipitation of Division.4, 7 and 8 were increasing. And the precipitation variations of Division.8 in 1981 and 1996 were abnormal high. The result may owed to the shortage of meteorological stations in Division.8. The AT10 of all the Divisions had a significant increasing trend and the increasing trend was absolutely significant in the 21st century. The trends of precipitation and AT10 in grassland were in agreement with the previous study [11,17].

Table 2. Trend of NDVI, precipitation and AT10 in grassland Divisions

Divisions	NDVI		precipitation		AT10	
	β	Z	β	Z	β	Z
Division.4	-0.0013	-3.0128*	0.6123	0.4437	9.4113	2.4523*
Division.6	0.0020	2.3589*	-2.4034	-1.8450	14.7402	3.6667*
Division.7	0.0006	1.3779	0.8284	1.9385	15.9633	3.8069*
Division.8	0.0007	2.3589*	0.1030	0.0701	9.7521	4.6476*

* represent the test result satisfied significant level at α = 0.05.

3.2 Spatial Distribution Changes of NDVI, Precipitation and AT10

Due to the different terrain and vegetation types, the distribution changes of NDVI, precipitation and AT10 were very different. The precipitation was increasing in the northwest of grassland except for the region of Himalayan Mountain (Fig.2b). And the AT10 was generally increasing overall grassland besides parts of Division.4 and 7. The possible reason is that the part of Division.4 is located in the Hengduan Mountains region with complex terrain and diversity variety of vegetation. The part of Division.7 is located around the Tarim basin with the Tarim desert, the biggest desert of China. The distribution of region with decreasing AT10 is consistent with those obtained by

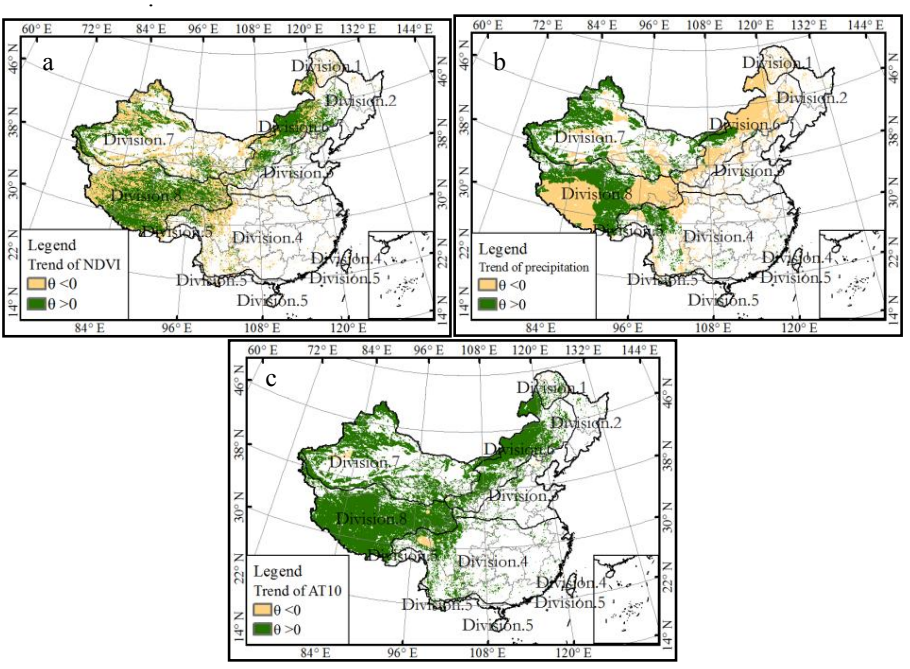


Fig. 2. Spatial distribution trend of NDVI, precipitation and AT10 from 1982 to 2006, (a) distribution of annual NDVI trend, (b) distribution of annual precipitation trend, (c) distribution of annual AT10 trend

Dong et al. (2009). Although both of precipitation and AT10 were increasing in Division.4, the NDVI was overall decreasing. The reason is that the NDVI in Division.4 responds differently to the climate factors because of the complex terrain and vegetation diversity there. However, the NDVI had an increasing trend in most of Division.6-7 (Fig.2a). It is because that the increase of precipitation and AT10 provided much more moisture and thermal conditions for growth of vegetation in this regions.

3.3 Correlation between NDVI, Precipitation and AT10

Different Divisions have different climate and vegetation types, so the responses of grassland NDVI to the changes of precipitation and AT10 were different (Fig.3). The grassland NDVI was negative related to the precipitation and AT10 in Division.4. It is because that the complex topographic and vegetation diversity there made the relation between NDVI and climate factors abnormal. The grassland NDVI was positive related to precipitation and AT10 in Division.6 and 7, and the NDVI-precipitation correlation coefficients were higher than NDVI-AT10 correlation coefficients. The reason is that Division.6 and 7 are located in the arid and semi-arid region, the precipitation is the restriction climate factor for the growth of vegetation there. The NDVI was negative related to the precipitation and positive related to the AT10 in Division.8, especially in

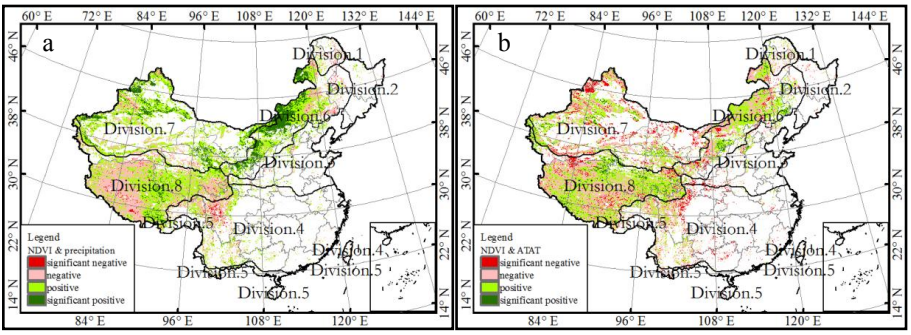


Fig. 3. Spatial distribution of correlation coefficient between NDVI, precipitation and AT10, (a) distribution of correlation coefficient between NDVI and precipitation, (b) distribution of correlation coefficient between NDVI and AT10

Table 3. Correlation coefficient between NDVI and climate factors

Divisions	Vegetation regionalization	NDVI & precipitation	NDVI & AT10
Division.4	Subtropical broadleaf evergreen forest regionalization	-0.1618	-0.5094*
Division.6	Temperate steppe regionalization	0.3611	0.1967
Division.7	Temperate desert regionalization	0.6808*	0.0722
Division.8	Qinghai-Tibet plateau alpine vegetation regionalization	-0.1035	0.2086

* Correlation coefficient at 0.05 significant level.

the Himalayan Mountain region (Fig.3). There are many glaciers and snow covers on Qinghai-Tibet plateau, and the AT10 was increasing there. So the increasing of AT10 not only provided much more heat, but also more moisture for the growth of grass, since the glaciers and snow covers melt by the increasing of AT10.

3.4 Binary Linear Regression Analysis of NDVI, Precipitation and AT10

Binary regression analysis was applied to quantify the relation between NDVI, precipitation and AT10 in different Divisions. The results are shown in table 4 (N, P and T represent NDVI, precipitation and AT10 of the division respectively.). The multiple regression coefficients are absolute significant in Division.4, 6 and 7. It means that the grassland NDVI are closely related to precipitation and temperature, so we can estimate the variation of grass NDVI based on the information of precipitation and AT10 there. However, the multiple regression coefficient was very low in Division.8, so the correlation between NDVI, precipitation and AT10 is weak. The possible reason is that part of the heat was used to melt the glaciers and snow covers in the Qinghai-Tibet plateau in Division.8. So the transformation between moisture and heat impact the real relation between NDVI and precipitation and AT10.

Table 4. Regression equations of NDVI and precipitation and AT10

Divisions	Regression equations	R^2
Divisions.4	N = -0.0370P - 0.1852T + 0.7898	0.2723*
Divisions.6	N = 0.1358P + 0.2244T + 0.1576	0.2699*
Divisions.7	N = 0.0741P + 0.0343T + 0.1694	0.4808*
Divisions.8	N = -0.0046P + 0.0326T + 0.2433	0.0480

* Correlation coefficient at 0.05 significant level.

3.5 Effects of Latitude and Elevation

Previous studies reported that underlying surface factors (as elevation and latitude) have effects on the distribution and growth of vegetation [21,31]. So the relation between the correlation coefficient (NDVI-precipitation and NDVI-AT10) and underlying surface factors (elevation and latitude) of the meteorological stations on the grassland were studied (Fig.4). The result showed that the correlation coefficient (NDVI-precipitation and NDVI-AT10) was higher on the stations in higher latitude, and the correlation coefficient (NDVI-precipitation and NDVI-AT10) was lower on the stations with higher elevation. What's more, the NDVI-precipitation correlation coefficient was more sensitive than NDVI-AT10 correlation coefficient to the underlying surface factors, and the correlation coefficient between the NDVI-precipitation correlation coefficients and underlying surface factors was significant higher than that of NDVI-AT10 correlation coefficient. Generally speaking, both of the correlation coefficient between the correlation coefficient (NDVI-precipitation and NDVI-AT10) and underlying surface factors were low and insignificant, it means that the effects of latitude and elevation on the correlation of NDVI-precipitation and NDVI-AT10 were weak and the NDVI-precipitation and NDVI-AT10 correlations obtained before were effective.

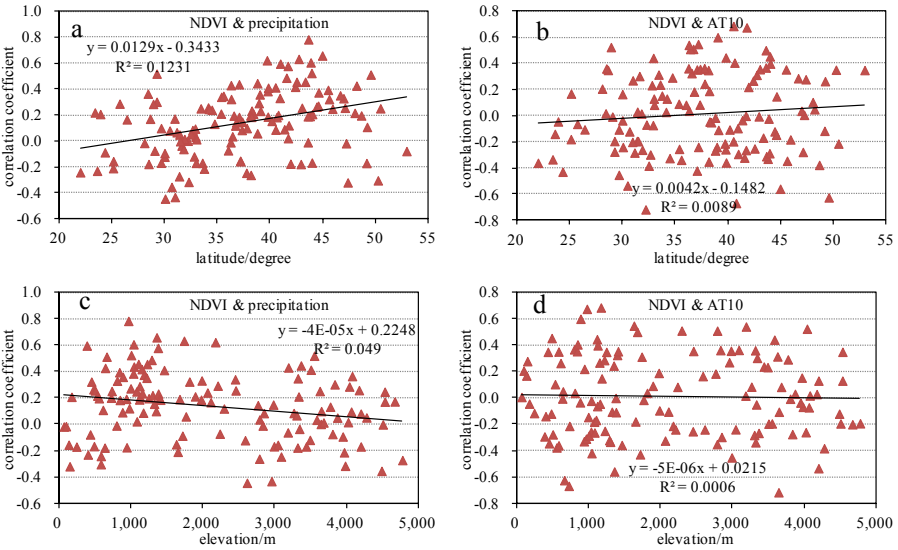


Fig. 4. Relation of correlation coefficient and latitude and elevation in 130 meteorological stations (the station in the grassland), (a) relation between NDVI-precipitation correlation coefficient and latitude, (b) relation between NDVI-AT10 correlation coefficient and latitude, (a) relation between NDVI-precipitation correlation coefficient and elevation, (a) relation between NDVI-AT10 correlation coefficient and elevation

4 Conclusions

In this study, spatial and temporal changes of NDVI, precipitation and AT10 of the grassland were studied from 1982 to 2006 in China. And the relation between the NDVI and climate factors in different vegetation regionalization were analyzed and quantified. The result showed that: (a) the NDVI of grassland was increasing significantly owing to the complex topographic and vegetation diversity in the subtropical broadleaf evergreen forest regionalization (Division.4), while the precipitation and AT10 were increasing. (b) The grassland NDVI were positive related to the precipitation and AT10 in temperate steppe and desert regionalizations (Division.6 &7), especially to precipitation. It means the precipitation is the restrictive climate factor for the growth of grass there. (c) The NDVI, precipitation and AT10 of grassland were increasing in Qinghai-Tibet plateau alpine vegetation regionalization (Division.8). However, the correlation coefficient between NDVI and precipitation was negative and correlation coefficient between NDVI and AT10 was positive. The possible reason is that the transformation between moisture and heat impact the real relation between NDVI and precipitation and AT10, and this transformation makes the AT10 a more important climate factor for the development of the grass in Division.8.

Acknowledgments. This research is supported by the State Key Development Program for Basic Research of China (Grant No. 2010CB951102), the National Natural Science

Foundation Project of foundation of innovation group (Grant No.51021066) and the National Natural Science Founds (Grant No.51279207).

References

1. Knapp, A.K., Smith, M.D.: Variation among biomes in temporal dynamics of aboveground primary production. Science 291, 481–484 (2001)
2. Nemani, R.R., Keeling, C.D., Hashimoto, H.: Climate-driven increases in global terrestrial net primary production from 1982 to 1999. Science 300, 1560–1563 (2003)
3. Weltzin, J.F., Loik, M.E., Schwinning, S.: Assessing the response of terrestrial ecosystems to potential changes in precipitation. Bioscience 53, 941–952 (2003)
4. Wang, J., Meng, J.J., Cai, Y.J.: Assessing vegetation dynamics impacted by climate change in the southwestern karst region of China with AVHRR NDVI and AVHRR NPP time-series. Environment Geology 54, 1185–1195 (2008)
5. Zhou, L., Tucker, C.J., Kaufmann, R.K., Slayback, D.A., Shabanov, N.V., Myneni, R.B.: Variations in northern vegetation activity inferred from satellite data of vegetation index during 1981 to 1999. Journal of Geophysical Research 106(17), 20069–20083 (2001)
6. Wang, J., Rich, P.M., Price, K.P.: Temporal responses of NDVI to precipitation and temperature in the central Great Plains, USA. International Journal of Remote Sensing 11(24), 2345–2364 (2003)
7. Beck, P.S.A., Atzberger, C., Høgda, K.A.: Improved monitoring of vegetation dynamics at very high latitudes, A new method using MODIS NDVI. Remote Sensing of Environment 100, 321–336 (2006)
8. Mao, D.H., Wang, Z.M., Luo, L., Ren, C.Y.: Integrating AVHRR and MODIS data to monitor NDVI changes and their relationships with climatic parameters in Northeast China. International Journal of Applied Earth Observation and Geoinformation 18, 528–536 (2011)
9. Qian, W.H., Lin, X.: Regional trends in recent temperature indices in China. Climate Research 27, 119–134 (2004)
10. Qian, W.H., Lin, X.: Regional trends in recent precipitation indices in China. Meteorology and Atmospheric Physics 90, 193–207 (2005)
11. Ren, G.Y., Guo, J., Xu, M., Chu, Z., Zhang, L., Zou, X., Li, Q., Liu, X.: Climate changes of China's mainland over the past half century. Acta Meteorological Sinica 63, 942–956 (2005)
12. Piao, S.L., Ciais, P., Huang, Y., Shen, Z., Peng, S., Li, J., Zhou, L., et al.: The impacts of climate change on water resources and agriculture in China. Nature 467, 43–51 (2010)
13. Xu, X., Du, Y., Tang, J.P., Wang, Y.: Variations of temperature and precipitation extremes in recent two decades over China. Atmospheric Research 101, 143–154 (2011)
14. Schultz, P.A., Halpert, M.S.: Global analysis of the relationship among a vegetation index, precipitation and land-surface temperature. International Journal of Remote Sensing 16(15), 2755–2777 (1995)
15. Ji, L., Peters, A.J.: A spatial regression procedure for evaluating the relationship between AVHRR-NDVI and climate in the northern Great Plains. International Journal of Remote Sensing 25(2), 297–311 (2004)
16. Xiao, J., Moody, A.: Photosynthetic activity of US biomes: responses to the spatial variability and seasonality of precipitation and temperature. Global Change Biology 10, 437–451 (2004)

17. Dong, J.W., Liu, J.Y., Tao, F.L., Xu, X.L., Wang, J.B.: Spatio-temporal changes in annual accumulated temperature in China and the effects on cropping systems 1980s to 2000. Climate Research 40, 37–48 (2009)
18. Phipps, R.H., Rosemary, J.F., Crofts, F.C.: Relationships between the productions of forage maize and accumulated temperature, ontario heat units and solar radiation. Agricultural Meteorology 14(1-2), 385–397 (1974)
19. Bartholomew, P.W., Williams, R.D.: Cool-season grass development response to accumulated temperature under a range of temperature regimes. Crop Science 45(2), 529–534 (2005)
20. Walsh, S.J., Crawford, T.W., Welsh, W.F., Crews-Meyer, K.A.: A multi-scale analysis of LULC and NDVI variation in Nang Rong district, northeast Thailand. Agriculture, Ecosystems & Environment 85, 47–64 (2001)
21. Zhan, Z.Z., Liu, H.B., Li, H.M., Wu, W., Zhong, B.: The Relationship between NDVI and Terrain Factors –A Case Study of Chongqing. Procedia Environmental Sciences 12, 765–771 (2012)
22. Tucker, C.J., Pinzon, J.E., Brown, M.E., Slayback, D.A., Pak, E.W., Mahoney, R., Vermote, E.F., Saleous, N.E.: An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. International Journal of Remote Sensing 26(20), 4485–4498 (2005)
23. Holben, B.N.: Characteristics of maximum-value composite images from temporal AVHRR data. International Journal of Remote Sensing 7(11), 1417–1434 (1986)
24. Kahya, E., Kalayci, S.: Trend analysis of stream-flow in Turkey. Journal of Hydrology 289(1-4), 128–144 (2004)
25. Ludwig, W., Serrat, P., Cesmat, L., Garcia-Esteves, J.: Evaluating the impact of the recent temperature increaseon the hydrology of the Tet River (Southern France). Journal of Hydrology 289, 204–221 (2004)
26. Xu, Z.X., Gong, T.L., Li, J.Y.: Decadal trend of climate in the Tibetan Plateau-regional temperature and precipitation. Hydrological Processes 22, 3056–3065 (2008)
27. Pouliot, D., Latifovic, R., Olthof, I.: Trends in vegetation NDVI from 1 km AVHRR data over Canada for the period 1985–2006. International Journal of Remote Sensing 30(1), 149–168 (2009)
28. Runnstrom, M.C.: Is northern China winning the battle against desertification? Satellite remote sensing as a tool to study biomass trends on the Ordos plateau in semiarid China. Ambio 29, 468–476 (2000)
29. Zhong, D.C., Qu, J.J.: Recent developmental trend and prediction of sand deserts in China. Journal of Arid Environments 53, 317–329 (2003)
30. Fang, J.Y., Piao, S.L., He, J.S., Ma, W.H.: Increasing terrestrial vegetation activity in China, 1982-1999. Science in China (Ser. C) 47, 229–240 (2004)
31. Jia, G.S., Epstein, H.E., Walker, D.A.: Spatial characteristics of AVHRR-NDVI along latitudinal transects in northern Alaska. Journal of Vegetation Science 13, 315–326 (2002)

Research on Assessment Method of Winter Wheat Water Use Efficiency Based on ET and Biomass with Remote Sensing

Jun E. Fu*, Zhiguo Pang, and Jingxuan Lu

China Institute of Water Resources and Hydropower Research,
A-1 Fuxing Road, Haidian District, 100038 Beijing, 100048 Beijing, China
fuje@iwhr.com

Abstract. The purpose of this study is to make an attempt on establishing a method to assess water management for larger irrigated area using remote sensed information. Firstly, we analyzed the crop water consumption rule during the whole growing stage taking winter wheat as an example. Then the remote sensed assessment method of water use efficiency was established, which took remote sensed ET, biomass and observed precipitation as input data. Finally, the method was applied for the assessment of irrigation water use efficiency in Beijing Tongzhou district over the three growing seasons of 2003-2004, 2004-2005, 2005-2006 of winter wheat. In the end, the feasibility of this method is validated by comparison the results with year-to-year variation of statistical yields. The assessing results can be used for water managers and policy makers in improving irrigation water use efficiency.

Keywords: evapotranspiration, water consumption, biomass, water use efficiency, water productivity.

1 Introduction

Water shortages are already a major global issue in recent years that needs to be addressed urgently, and the problem in the Northwest and the North China is getting particularly serious, which has become a key factor affecting agricultural development. In order to solving the embarrassment of water scarcity, the government increased investment in strengthening construction of non-project measures saving water. Developing water-saving and efficient agriculture is regarded as a basic national policy.

So if water is managed more efficiently, so that crop yield per unit of water consumption increases. For effective monitoring of irrigation and efficient irrigation management, accurate estimation of real irrigation applications is an essential task. Food production and water use are two closely linked processes. As the competition for water intensifies worldwide, water in food production must be used more

* Corresponding author.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 381–388, 2013.

efficiently[1]. Crop yield per unit of water consumed, or "water productivity", is a key element in successful water resource management [2].

Water productivity(WP), is an accurate indicator of agricultural productivity in relationship to the crop's consumptive use of water, which is defined as the ratio of the net benefits from crop to the amount of water required to produce those benefits[2]. Put simply, it means growing more food or gaining more benefits with less water. In the literature water productivity is mostly expressed as the ratio of biomass to evapotranspired water, because separation of E from T was not possible in these cases. The temporal and spatial variation of ET and biomass estimated by remote sensing(RS) can provide the regional difference of agricultural water consumption, so it can be used to assess agricultural water use efficiency objectively. So above equation is changed for: WP_b=B/ET, where WP_b is biomass water productivity[1].

Improving water productivity is one important strategy for addressing future water scarcity. The key objectives of this paper therefore are: (a) to make an attempt on establishing one approach to water management for larger irrigated areas based on remotely sensed free information; and (b) to assess the water use efficiency of winter wheat-growing area using the approach and then to demonstrate its feasibility.

2 Study Area, Data and Methods

2.1 Study Area and Data

Beijing is located in Haihe River Basin, where is extremely short of water resources. Agriculture is the major water consumption industry of Beijing, the irrigation area of Beijing is about 4 600 000 mu, irrigation water accounted for 34% of the city's total water resources amount (Beijing Municipal Water Affairs Bureau, 2008). Among the main crops in Beijing, winter wheat is with the worst water deficit because fewer precipitation in the whole growing season, which is mainly distributing in the eastern and southern plains of Beijing.

The data of remote sensed ET and biomass were collected from the remote-sensed ET and biomass monitoring system in Beijing by using MODIS and TM, which could provides ET and biomass data with 1km resolution from 2003 to 2006 as half a month integrated values, verified by observation data.

We selected 8 stations of planting winter wheat as sample points, which are all below 40m with light loam soil and same climate conditions. As the winter wheat growing season lasts from early last October to early June of this year, we accumulated ET data of each sample point during this period, so that we obtained the wheat water consumption of three growth stages in 2003-2004, 2004-2005 and 2005-2006. The rainfall data from 2003 to 2006 of 6 stations are collected from the China Meteorological Administration Web site, and these data are spatially interpolated to attain the spatial resolution matching remotely sensed data. The data of winter wheat yield per acre are collected from Tongzhou statistical yearbook[3].

2.2 Biomass Water Productivity Calculations

Water productivity is defined as the marketable crop yield per unit of actual total evaporation (ET)[4]:

$$WP = Y/ET. \quad (1)$$

Where, Y is the crop yield, ET is the actual total evaporation. The AquaCrop model, evolved from the FAO I&D Paper No. 33 approach, lead to the equation[5]:

$$B=WP\bullet\Sigma Tr. \quad (2)$$

where, B is the biomass produced cumulatively (kg per m^2), Tr is the crop transpiration(mm). The water consumed and not recoverable in crop production process is normally assessed in terms of evapotranspiration (ET), the sum of productive transpiration by the crop (T) and non-productive evaporation from the soil (E). The spatial estimates of evaporation and biomass were used to calculate biomass water productivity of crop.

2.3 Analysis of Water Consumption Rules of Winter Wheat

Other studies[5] have shown that crop water consumption during crop whole growth period is closely related to the yields and water productivity. In this section, we put the biomass and ET into the above equation for calculating WP_b, and then analyze the relationship among water consumption, biomass and biomass water productivity. The relationships between seasonal ET, WP_b and biomass are given by Fig. 1.

It is obvious from Fig. 1 that WP_b and biomass were increased as ET increased, then dropped as ET continue to increase. It should be noted that ET higher than 300 mm(the blue solid line) gave lower WP_b and higher biomass, while ET higher than 408 mm (the red dotted line) gave lower biomass and WP_b, which shows that much water does not always obtain higher biomass[7]. Thus we get two critical values: $ET(WP_{b\ max})$(300mm) as the blue solid line and $ET(B_{max})$ (408mm) as the red dotted line, which divide water consumption into 3 regions.

Based on above analysis, we can draw conclusions as: ET(Bmax) can be considered as the water consumption on full irrigation; ET(WPb max) can be considered as the base water consumption with ensuring high irrigation efficiency; In a region when the crop water consumption exceeds ET(Bmax), if excluding the influence of rainfall, it can be regarded as overirrigation region; on the contrary, a region when the water consumption less than ET(WPb max) can be considered as the irrigation water deficit areas; The region of water consumption between in ET(WPb max) and ET(Bmax) will get both optimal water use efficiency and crop biomass. Thus, the interval [ETa(WPc'max), ETa(Bmax)] can be determined as the suitable water consumption interval for deficit irrigation in water shortage area. Furthermore, we can calculate that the suitable water consumption interval of winter wheat in Beijing is 300mm~408mm.

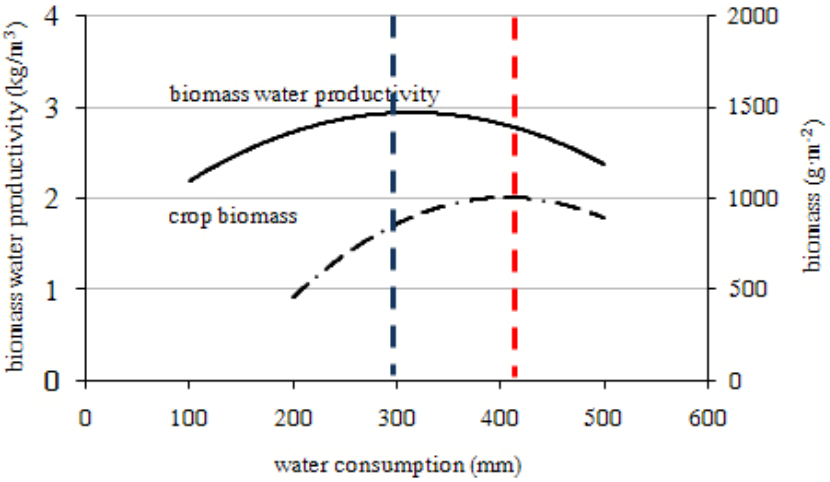


Fig. 1. Relationship of ET and WP_b, water consumption and biomass of winter wheat

2.4 Method and Steps for Irrigation Water Use Efficiency

Based on the analyses, we will establish an approach of assess irrigation performance using remote sensed information. In order to ensure the accuracy of the assessing result, some factors in the following that need to consider: a) The analyses of water consumption rules indicate that in the case of same biomass water productivity, the biomass is different because of the different crop consumption, which means that low water consumption is corresponded to low biomass and high water consumption corresponded to high biomass, but the ratio of both is same. Thus, water productivity is not the only index for assessing irrigation water use efficiency. For this reason, first this paper makes an attempt on dividing the cropped area into three regions by water consumption interval, that can guarantee the existence of a unique biomass water productivity for each region; b) Furthermore, crop water consumption is composed of three components: ET, rainfall and groundwater recharge. The level of groundwater ranges from 10m to 50m in Beijing plain, so its supplement for surface soil water can be ignored. Furthermore, in the steps of assessing the irrigation water use efficiency, the influence of precipitation need to be considered; c） In order to address the difference of irrigating water use efficiency for larger irrigated areas, remotely sensed free information is used.

In general, the main problems in the irrigation water management are irrigation water shortage, over irrigation, unreasonable irrigation mode and etc., all of these could reduce the efficiency of water management and waste a mass of water resource. In view of this, the approach classifies the results into 4 types: water shortage area(C1), reasonable irrigation area(C2), unreasonable irrigation area(C3) and over irrigation area(C4), and the specific meaning is as follows:

i. Water shortage area is defined as the cropped region where the irrigation water cannot meet crop water requirement during crop growing period. In this area with increased irrigation both yield and water productivity increased.

ii. Reasonable irrigation area is defined as the cropped region where the proper timing and quantity of water to be applied to the crops during the crop growing period. It is a high water use efficiency area, where most of the irrigation water use is consumed for the production of biomass in the crop production process. Deficit irrigation is a strategy in this region that allows a crop to sustain some degree of water deficit in order to reduce costs and potentially increase income. It can lead to increased net income where water costs are high or where water supplies are limited.

iii. Unreasonable irrigation area is defined as the cropped region with unreasonable timing and quantity of irrigation water, under this condition yield of crop is adversely affected with excess or inadequate water supply. Excess water from irrigation sinks down into the water table or evaporates from the soil, which reduces the efficiency of water use.

iv. Over irrigation area is defined as the cropped area with over irrigation. The amount of irrigating water exceeds the crop water requirement of whole growing stage, which causes waste agricultural water.

Furthermore, in order to make an attempt on assessing irrigation water use efficiency for larger irrigated areas, remotely sensed ET and biomass are used in the approach. Five steps should be followed to guide use of the approach in assessing irrigation water use efficiency taking winter wheat as an example:

Step 1: The land cover map, used in the approach, is extracted by a multitemporal classification methodology based on Landsat TM images, then we get the area of winter wheat.

Step 2: Taking the remotely sensed ET and biomass of half a month integrated values as input data, and then analyzing crop water consumption rules to deduce the crop water consumption interval[$ET(WP_{b\ min})$, $ET(B_{max})$], on the basis, the crop area under winter wheat is divided into three regions by the two critical values.

Step 3: In the region of $ET \leq ET(WP_{b\ min})$, all pixels in the area is classified as1;

Step 4: In the region of $ET(WP_{b\ min}) < ET \leq ET(B_{max})$, water supplies are limited. The main goal of agricultural water management is ensuring high irrigation efficiency. According to this, we just select the maximal values of WP_b under the same ET to draw the ET-WP_b scatter diagram and then get the relational curve between ET and WP_b such as Figure 1. The relational curve also means the target water productivity curve as the water consumption changes. By comparison of the water productivity of every pixel and the target value of the curve, if the result is equal or greater than 1, the pixel is classified 2, else it will be 3.

Step 5: In the region of $ET > ET(B_{max})$, the amount of irrigating water exceeds the crop water requirement of whole growing stage. The water in excess comes probably from irrigation or rain. Here we use irrigation water consumption(mm) to distinguish,and the formula is as follows:

$$ID=\Sigma(ET-Pe). \tag{3}$$

Where, ID is the irrigation water consumption(mm), and Pe is effective precipitation which is defined as the part of rainfall that can be used to meet the evapotranspiration of growing crops, which is described by the formula: $Pe = P \bullet a$.

Where P is the observed precipitation, and a is the precipitation effective utilization coefficient relating to time rainfall. The following empirical coefficient in China are used: if time rainfall is less than 50mm, and then a=1.0, if it is equal or greater than 50 but less than 150, and then a=0.80-0.75, otherwise it is equal or greater than 150mm, and a=0.7.

If ID is equal or less than 0, and then The water consumption in excess comes mainly from precipitation and the area will be classified 2, else if ID is greater than 0, and then he excessive water is from irrigation and the corresponding area is classified as 4.

3 Results and Discussion

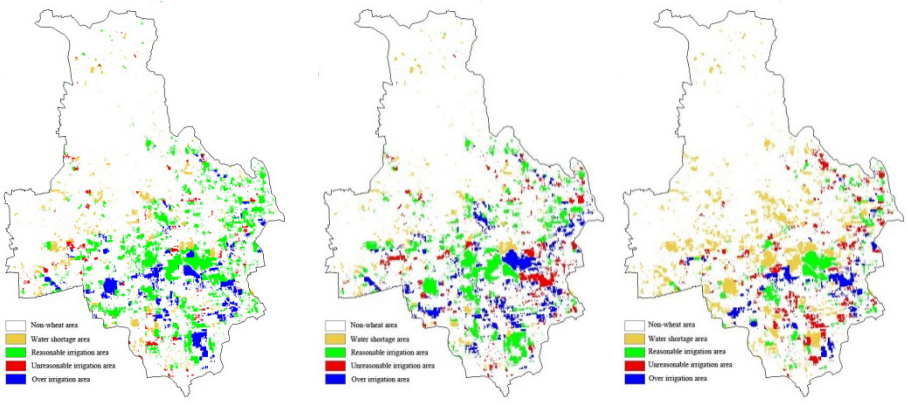
3.1 Application of the Assessing Approach

The method was applied for the assessment of water using efficiency of Beijing Tongzhou district over the three growing seasons of 2003-2004, 2004-2005, 2005-2006 of winter wheat, taking precipitation, ET and biomass by RS as input data. To assess the irrigation performance, WPb was calculated and the relationships among seasonal ET, biomass and WPb of winter wheat were established. The suitable water consumption interval of winter wheat in Tongzhou District of Beijing is 300mm~408mm. Results of assessing water use efficiency showed as follows:

i. In 2003-2004, the order of the area of 4 classes from largest to smallest is: C2(52%)>C4(23%)>C1(17%)>C3(8%). The reasonable irrigation area is largest and the unreasonable irrigation area is smallest;

ii. In 2004-2005, the order of the area of 4 classes is: C2(42%)>C4(26%)>C3(17%)>C1(15%). The reasonable irrigation area is largest and the water shortage area is smallest;

iii. In 2005-2006, the order of the area of 4 classes is: C1(42%)>C2(26%)>C4(17%)>C3(15%). The water shortage area is largest and the unreasonable irrigation area is smallest;



(a)2003-2004 (b)2004-2005 (c)2005-2006

Fig. 2. Assessment maps of water use efficiency for winter wheat in Tongzhou district

3.2 Validation with the Assessing Results

The district yield statistics, to some extent, represent a regional 'ground truth', offer a validation of the assessing results. Table 1 shows the statistical yield and assessing classes for winter wheat. As we know, the greater the sum of water shortage area and unreasonable irrigation area, the lower the district yield, which is consistent with trend like the assessment results as Table 1. The yield of winter wheat in 2004 is highest and it is lowest in 2006. Furthermore we go on the analysis combined with the weather data in 2006, then find that low rainfall and uneven seasonal distribution result in severe water stress, consequently low harvest in the area.

Table 1. The comparison of assessment results of water use efficiency and yield statistics

classes / seasons	C2	C4	C3	C1	C1+C3	Yield (kg/mu)
2003-2004	52%	23%	8%	17%	25%	383.58
2004-2005	42%	26%	17%	15%	32%	364.28
2005-2006	20%	19%	15%	46%	61%	352.81

4 Conclusion

In this paper, we made an attempt on establishing one remote sensed assessment approach to water management for larger irrigated area using remotely sensed free information. The approach was applied for the assessment of water using efficiency of Beijing Tongzhou district over the three growing seasons of 2003-2004, 2004-2005, 2005-2006 of winter wheat, taking precipitation, ET and biomass by RS as input data. In the end, the feasibility of this method was validated by comparison the result with year-to-year variation of statistical yields.

The results of the water consumption rules analysis show that the suitable water consumption interval of winter wheat in Beijing is 300mm~408mm, and much water is not necessary to obtain higher yield. The study also shows that remote sensing information is equity, adequacy and reliability, which can reflect the pixel value differences of water use efficiency of large irrigated area.

The assessing results can be used for water managers and policy makers in improving irrigation water use efficiency.

References

1. Steduto, P., Hsiao, T.C., Fereres, E.: On the conservative behavior of biomass water productivity. Irrig. Sci. 25, 189–207 (2007)
2. Abdul-Ganiyu, S., Amaanatu, M.K., Korese, J.K.: Water use efficiency and productivity for rice (oryza sativa) in the Bontanga irrigation scheme of northern region of Ghana. Agric. Sci. Res. J. 2(7), 362–368 (2012)
3. Tongzhou statistical yearbook, http://www.tz.bjstats.gov.cn/

4. Shah, S., Dalwadi, H.J.: Critical appraisal of an irrigation command and water productivity based on satellite remote sensing. Inter. J. of Water Resour. Environ. Eng. 3(2), 41–45 (2011)
5. AquaCrop: concepts, rationale and operation, http://www.fao.org/docrep/016/i2800e/i2800e04.pdf
6. Xiao, J.F., Liu, Z.D., Duan, A.W., Liu, Z.G.: Studies on Water Production Function of Winter Wheat in Xinxiang District. Henan Agric. Sci. 1, 55–59 (2009)
7. Mo, X., Liu, S., Lin, Z., Xu, Y., Xiang, Y., McVicar, T.R.: Prediction of crop yield, water consumption and water use efficiency with a SVAT-crop growth model using remotely sensed data on the North China Plain. Ecol. Model. 183, 301–322 (2005)

GNSS Investigation in the Early Stage of the Three Gorges Project on the Yangtze River

Zhige Jia*, Gang Liu, Wei Wang, and Yu Zhou

Key Laboratory of Earthquake Geodesy, Institute of Seismology, CEA, Wuhan, China
jia.zg@126.com, jia-zg@sohu.com

Abstract. The GNSS deformation monitoring network of the Three Gorges Reservoir is introduced. The GNSS monitoring strategy from 2011 to 2012 is elaborately analyzed from technology design, field observation, data processing and deformation analysis. The baseline length variations of these GNSS stations from 2011 to 2012 are less than 5 mm, and the vertical variation of each station is about 10 mm. The computing results show that the two-year's variation of the crustal deformation is small, and then the surface structure variation is stable in present around the Three Gorges Reservoir area.

Keywords: The Three Gorges, GNSS, Crustal Deformation, Monitoring Network.

1 Introduction

It is usually suggested that there are two physical mechanisms for reservoir-induced siesmicity: additional loading of water weight to the existing geostress background cause the reservoir-induced seismicity [1]; and decreasing effective normal stress at the fault surface due to increasing pore pressure in subsurface rocks [2]. An important geologic hazard of TGP is reservoir-induced seismicity, known to have caused earthquakes in India (Konya), China (Xinfeng Jiang/Hsingfengkiang), Zimbabwe (Kariba), and Greece (Kremasta) (Talwani, 1997). Like natural earthquakes, before and after reservoir-induced seismicity will produce crustal deformation [3]. For its potential impact on the surface structure around the Yangtze Three Gorges dam site, the crustal deformation monitoring system of the Three Gorges Project is to monitor the Three Gorges Reservoir area, which is from Yichang to Badong [4]. With high precision and real-time monitoring capability, the GNSS (Global Navigation Satellite System) becomes one of the main components to monitor crustal deformation of the Three Gorges Reservoir area, and to provide a basis for monitoring and forecasting earthquakes. This paper first introduces geological background of the Three Gorges Reservoir and then reviews composition of the Three Gorges GNSS deformation monitoring network. The GNSS monitoring strategy is elaborated from technology design, field observation, data processing and deformation analysis. Based on the GNSS data from 2011 to 2012, the level deformation and vertical variation of the

* This work was financially supported by the director foundation (IS200756040) of Institute of Seismology, China Earthquake Administration.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 389–396, 2013.

GNSS station are calculated. The largest baseline length variation between 2011 and 2012 is in the range of 5 mm, and the vertical deformation is about 10mm in the two years. According to usual suggestion, Huanan block is a relatively stable geological block, there is no apparent velocity gradient zone and important differential movement throughout the plate interior [5], [6], vertical motion is less than 1 mm/a, and horizontal motion reaches 6~11mm/a [7]. The level and vertical deformations from 2011 to 2012 indicate that the crustal deformation around the Three Gorges Reservoir is very small in the two-year period. The results show that the surface structure of the Three Gorges Reservoir is relatively stable at present.

2 Geological Background

As the river flows east through the Daba mountains, The Yangtze River encounters a narrow constriction at the Three Gorges of Qutang, Wu, and Xiling (or Sanxia). The Three Gorges area spans two tectonic units, of which the north is Qinling fold belt and the south is Yangtze paraplatform with the boundary of Qingfeng fault. Head area of the Three Gorges Reservoir is located in upper Yangtze platform fold belt of Yangtze paraplatform. The dam of the Three Gorges Reservoir is located on granite bodies of Huangling anticline which is a secondary unit of upper Yangtze platform fold belt. The strength and stability of this anticlinal structure are the principal reasons for the siting of the Three Gorges Dam in its location, near Sandouping town (about 60 km west to Yichang city).

There are three main fault and fracture systems of NE-NNE, NNW-NW and NWW regional faults in the Three Gorges Reservoir area, which are mainly located in surrounding and outlying areas of Huangling anticline, such as Xiannushan fault that is 19km from the dam site, Jiuwanxi fault that is 17km from the dam site, Xinhua fault that is 40km from the dam site, Wuduhe fault that is 35km from the dam site, Tianyangping fault that is 16km from the dam site, Xinhua fault, Gaoqiao fault, Zhoujiashan-Moping fault, Jianshi fault, and so on. Firstly, to the southwest of the Huangling anticline is the NNW oriented Xiannushan fault, which consists of three parallel shear zones [8]. Secondly, the NNE oriented Jiuwanxi fault, which intersects the Xiannushan fault, near Zigui, and disappears just to the south of it. Thirdly, the Zigui basin is crossed by the Gaoqiao fault zone, which is of similar orientation to the Jiuwanxi fault. In addition, the area to the south of Zigui and Badong is characterised by a system of secondary faults, which follow the orientation of the fold system in the area, i.e. of NE [9]. These secondary fault and fracture systems tend to form ‘weak’ zones, which favour slope instability.

There are several common features of these faults as follows: (1) complex structural pattern; (2) the multi-phase activities; (3) frequent small earthquakes. During the early-middle Pleistocene of Quaternary, these fault zones occur strike-slip movement, and of which gave rise to M4.9 ~ M5.1 earthquakes at some time in the past [10], [11], [12], [13]. As a result, future M6.0 earthquakes may most likely occur in the interchange area of Xiannushan fault, Jiuwanxi fault and Tianyangping fault, and the Badong segment of the Three Gorges Reservoir. Moreover, the later become the focus of the GNSS monitoring because lithology and geological conditions of the BADN segment are suitable to induce the occurrence of reservoir earthquakes. The distribution of faults and the tectonic units are shown in Figure 1.

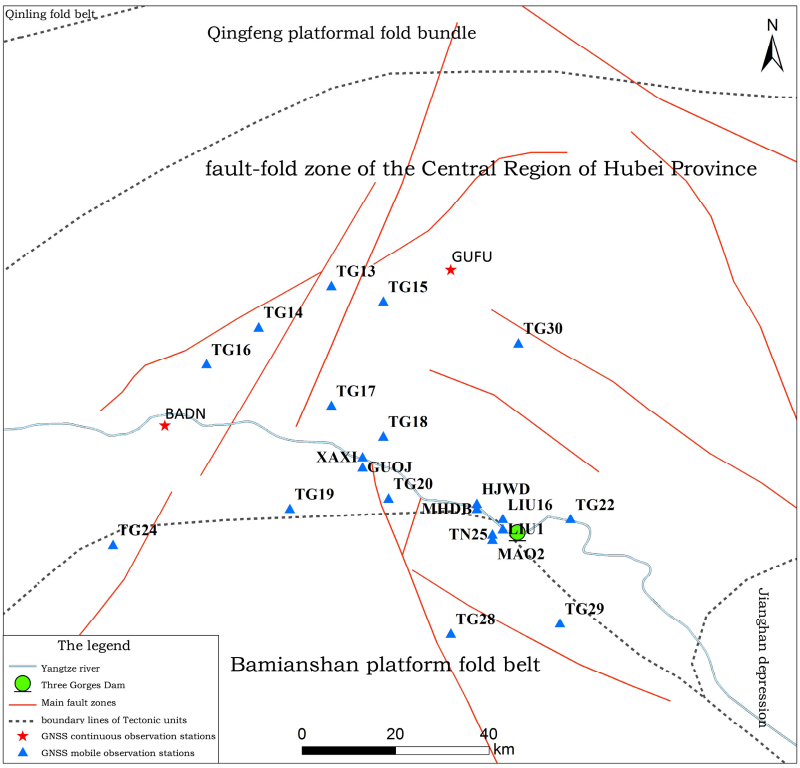


Fig. 1. The GNSS network and geological structure of the study area

3 GNSS Monitoring Network Constitutes

Selection of the GNSS station locations is closely link ed with geological and topographical conditions and tectonic movements. In order to design and construct overall the Three Gorges GNSS monitoring network, we combined geological conditions, historical earthquakes and faults distribution around the Yangtze Three Gorge dam site [14]. The Three Gorges GNSS monitoring network composes of two sections: three GNSS continuous observation stations and 21 GNSS mobile observation stations. All stations are built according to the standards of China Crustal Movement Observation Network, and established in bedrocks. As base stations, three GNSS continuous observation stations were GUFU station, BADN station and Wuhan station. All of 21 GNSS mobile observation stations belong to regional sites which are located around head area of the Three Gorges Reservoir, potential epicenter-source area of Xiannushan fault and Badong area. These regional sites include GUOJ, XAXI, TG13, TG14, TG15, TG17, TG18, TG19, TG20, TG28, TG29, TG30, MAO2, TN25, MAO2, LIU1, LIU16, HJWD, and so on, of which location information are shown in Figure 1.

4 GNSS Observations and Data Processing

4.1 GNSS Observations

(1) Continuous Observation Stations: The Three Gorges GNSS reference stations consist of Wuhan station, BADN station and GUFU station. For the last two station, all data is real-time transported by optical fiber. The two reference stations run well from 2011 to 2012. Running continuous rate and data good rate of the GNSS reference stations in the two years are counted respectively, and the results are shown in table 1 and table 2.

Table 1. The Continuous rates of continuous observation data of the GNSS stations

Station	*Observation time*	*Total days*	*Annual day of no data*	*Total days of no data*	*Continuous rates*
GUFU	2011.1.1-2011.12.31	365	111	1	99.7%
	2012.1.1-2012.12.31	366	356-359, 361	5	98.6%
BADN	2011.1.1-2011.12.31	365	99, 100, 124, 245-249, 300-304	13	96.4%
	2012.1.1-2012.12.31	366	064, 065, 134,162-165,247-264, 274-283,302-324,352-361	68	81.4%

Table 2. The Good rates of continuous observation data of the GNSS stations

Station	*Observation time*	*Total days*	*Invalid data*		*Good rates*
			Annual day	Total days	
GUFU	2011.1.1-2011.12.31	364	none	none	100%
	2012.1.1-2012.12.31	361	none	none	100%
BADN	2011.1.1-2011.12.31	352	none	none	100%
	2012.1.1-2012.12.31	298	none	none	100%

(2) GNSS Mobile Observation: In 2011, the observation time was from November 30 to December 20, with 7 TOPCON NET-G3A GNSS dual-frequency receivers with choke ring antenna. Each observation epoch of sync loops consists of 4 sections and lasts 96 hours, all observation section was exceed 22.5 hours, the data quality was better.

In 2012, the observation time was from September 13 to September 26, using eight TOPCON NET-G3A GNSS dual-frequency receivers with eight TPSCR.G3 choke ring antenna. Each observation epoch of sync loops consists of 3 sections and lasts 72 hours, all observation section was exceed 22.5 hours, the data quality was better.

4.2 Data Processing

GAMIT Software is widely used in satellite orbit determination, ERP solver, crustal movement and velocity field studies and the obtained results is with high accurate [15], [16]. With regards to this, the observation data of the Three Gorges GNSS monitoring network were processed by GAMIT10.4 Software. In order to improve data accuracy, we combine GNSS data around the Three Gorges region with others samples from over ten IGS stations and "China Crustal Movement Observation Network" inside and outside China.

The main GNSS errors consist of satellite clock bias, GPS receiver clock bias, ionospheric refractive error, tropospheric refractive error and phase center variation. In this paper, the satellite broadcast ephemeris parameters is used as clock bias parameters to correct the satellite clock bias model, clock bias parameters obtained by the pseudorange measurements is used to correct the receiver clock bias model, LC observations is used to eliminate the refraction effects by the ionospheric observations, standard atmospheric parameters is used to correct the effects by the tropospheric refraction [17], absolute antenna phase center correction model is used to correct the receiver antenna phase center. Based on the GAMIT software, several correction models are used to improve processing accuracy of the GNSS data, such as solid tide corection model, tide correction model, etc.

Based on GAMIT10.4 Software, the above GNSS stations data is computed as following steps: (1) coordinates and velocities of all IGS stations are transformed to an international terrestrial reference frame (ITRF 2008), (2) seven parameter transformation is implemented, (3) the daily coordinates and velocities of all GNSS stations is computed and adjusted by the above error correction models as weekly periods, (4) the above computing results are transformed to the ITRF 2008. In the above process, the involved IGS stations data consist of SHAO, WUHN, LHAS, URUM, BJFS, KUNM, CHUN inside CHINA, and POL2, KIT3, SELE, KSTU, TSKB, TAEJ, GUAM, IRKT, NTUS, USUD, IISC outside CHINA, and so on.

4.3 Data Analysis

After the calculation process based on the GAMIT software platform, the obtained nominal RMS accuracy of observation stations are 1~3mm in level direction and 3~6mm in vertical direction, the results fully meet the needs of deformation monitoring. Multi-period baseline repetition rate is used to evaluate the GNSS observation data accuracy and Baseline relative positioning situation, which is calculated as formula(1):

$$R = \sqrt{\frac{\frac{n}{n-1}\sum_{i=1}^{n}\frac{\left(c_i - \bar{c}\right)^2}{\sigma_i^2}}{\sum_{i=1}^{n} 1/\sigma_i^2}} \tag{1}$$

In the formula (1), R represents the baseline repetition rate, n represents the number of the baseline period solutions, c_i represents the baseline period solutions, σ_i represents the variance of c_i, $\bar{c}$ represents weighted average of the c_i, which is calculated as the formula (2):

$$\bar{c} = \frac{\sum_{i=1}^{n} c_i / \sigma_i^2}{\sum_{i=1}^{n} 1/\sigma_i^2} \tag{2}$$

According to the formula (1) and formula (2), we calculated the baseline length and the baseline repetition rate, respectivly.

The above computing results consist of baseline length, baseline length variation, baseline repetition rate, and the vertical deformations. These results of the GNSS data from 2011 to 2012 are counted, compared and listed in Table 3 and Figure 2.

(1)The Level Deformation Results: For the interchange area of Xiannushan fault, Jiuwanxi fault and Tianyangping fault, and the Badong segment of the Three Gorges Reservoir, the cross-fault baselines length are mainly calculated and contrasted. The results on variations of the baseline length from 2011 to 2012 are shown in Table 3.

Table 3. The amount of change of the baseline length from 2011 to 2012

Baseline Name	*2011*		*2012*		*Change baseline length (mm)*
	Baseline length	*Baseline repetition rate (mm)*	*Baseline length*	*Baseline repetition rate (mm)*	
BADN-GUFU	49903.8354	0.6	49903.8326	0.8	-2.8
BADN-GUOJ	38743.4771	0.6	38743.4780	0.8	0.9
BADN—XAXI	39071.0374	0.5	39071.0357	0.8	-1.7
BADN-TG13	42195.4471	0.7	42195.4458	0.5	-1.3
TG13_TG14	15589.3858	0.8	15589.3866	0.3	0.8
BADN-TG15	51890.5298	0.7	51890.5307	0.5	0.9
BADN-TG18	41696.7602	0.4	41696.7596	0.9	-0.6
BADN-TG17	31185.2081	1.3	31185.2074	0.5	-0.7
GUFU-TG13	8011.5567	0.4	8011.5567	0.3	0
GUFU-TG14	23498.7192	1.3	23498.7230	0.4	3.8
GUOJ-XAXI	1100.5483	0.4	1100.5479	0.8	-0.4
TG19-TG20	19396.4555	1.2	19396.4573	0.7	1.8
GUOJ-TG29	48459.3976	0.4	48459.3955	0.7	-2.1
TG18_TG28	55186.9655	0.7	55186.9643	0.6	-1.2
TN25-TG19	41023.3491	0.3	41023.3465	0.4	-2.6
MAO2-TG19	38589.2441	0.7	38589.2422	0.2	-1.9
MAO2-LIU1	2889.3766	0.6	2889.3739	0.5	-2.7
HJWD-TG19	35792.9844	0.4	35792.9855	0.5	1.1
TG28-XAXI	39564.6817	0.8	39564.6830	0.6	1.3
BADN-TG28	70006.8607	0.7	70006.8564	0.8	-4.3
BADN-TG29	84865.1713	0.8	84865.1665	1.1	-4.8

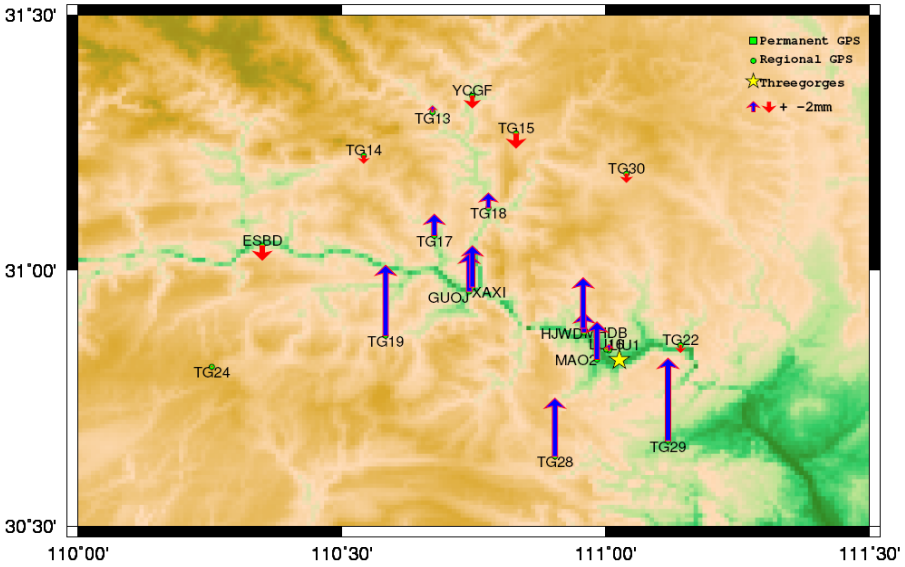


Fig. 2. Vertical variations of the baseline length from 2011 to 2012

The statistical results in Table 3 show that variations of the baseline length from 2011 to 2012 are less than 5 mm; the maximum value of variations is 4.8 mm (BADN-TG29), the minimum value of variations is 0 (GUFU-TG13). Variations of the baseline length are more twice than the weighted error range. The statistical result shows that the level displacement is significant credible and fully meets the requirements of crustal deformation monitoring.

5 Conclusions

The analysis results on the Three Gorges GNSS data basical faithfully reflect the tectonic deformation of the Three Gorges Reservoir area. As a relatively stable geological block, there is no apparent velocity gradient zone and important differential movement throughout inside of the Huanan block, vertical motion is less than 1 mm/a, and horizontal motion reaches 6~11mm/a. The level deformation and vertical deformation are integrated analyzed from 2011 to 2012, the results show that the crustal deformation in the Three Gorges Reservoir area is small from 2011 to 2012, and the surface structure of the Three Gorges Reservoir area is relatively stable at present. The work has also demonstrated that GPS is a very useful tool of for regional geological hazard and crustal deformation monitoring.

Recent construction activities have triggered and reactivated several large landslides. These in turn may affect reservoir capacity and dam safety. We feel that our approach is appropriate for a regional-scale assessment of surface structure stability and geological haza rd and that our results serve to highlight the need for immediate and long-term action plans in this area to ensure that further, inevitable

developments proceed within suitable engineering guidelines. This paper is only given two preliminary analysis results 175 meters after the impoundment water in 2011 and 2012. In the futuer, more multi-source and long-period data would be used for further study.

References

1. Gough, D.I., Gough, W.I.: Time dependence and trigger mechanisms for the Kariba (Rhodesia) earthquakes. Eng. Geol. 10(2-4), 211–218 (1976)
2. Hubbert, M.K., Rubey, W.W.: Mechanics of fluid-filled porous solids and its applications to overthrust faulting. Geol. Soc. Ame. Bull. 70, 115–166 (1959)
3. Yu, T.: Crustal deformation and reservoir-induced earthquakes. Crustal Deformation and Earthquake 11(2), 8–16 (1991)
4. Du, R., Xing, C., Wu, Z., et al.: Crustal deformation of Three Gorges Area. Journal of Geodesy and Geodynamics 24(2), 23–29 (2004)
5. Zhang, P., Wang, Q., Ma, Z.: GPS velocity field and active crustal blocks of contemporary tectonic deformation in continental CHINA. Earth Science Frontiers 9(2), 430–441 (2002)
6. Wang, W., Yang, S., Zhao, B., et al.: Present-day crustal movement velocity field in Chinese mainland. Journal of Geodesy and Geodynamics 6, 29–32 (2012)
7. Li, Q., You, X., Yang, S., et al.: A precise velocity field of tectonic deformation in China as inferred from intensive GPS observations. Sci. China Earth Sci. 55, 695–698 (2012)
8. Chen, S.: Atlas of geo-science analyses of Landsat imagery in China. National Remote Sensing Centre, Chinese Academy of Science. Science Press, Beijing (1986)
9. Wu, S., Hu, D., Chen, Q., Xu, R., Mei, Y.: Assessment of the crustal stability in the Qingjiang river basin of the western Hubei Province and its peripheral area, China. In: Proceedings of the Thirtieth International Geological Congress, Beijing, China, pp. 375–385. VSP International Science Publishers, Zeist (1997)
10. Li, A., Zeng, X.: Engineering Earthquake in the eastern Yangtze Gorges area. Seismological Press, Beijing (1996)
11. Hu, X., Qin, X.: Research on Reservoir-induced Earthquake Monitoring in Three Gorges Area. Engineering Science 5(11), 71–74 (2003)
12. Sun, Y., Tan, C., Wang, R., Hu, D.: An assessment and zonation of regional crustal stability in and around the dam region of the Three Gorges Project on the Yangtze River. Acta Geoscientica Sinica 17(3), 258–268 (1996)
13. Han, X., Rao, Y.: Analysis of genesis of reservoir-induced microquakes in Badong reach of Three Gorges. Journal of Geodesy and Geodynamics 24(2), 74–77 (2004)
14. Xing, C., Gong, K., Du, R.: Crustal deformation monitoring network for three gorges project on Yangtze river. Journal of Geodesy and Geodynamics 23(1), 114–118 (2003)
15. Niu, Z., You, X., Wang, Q., Wang, M.: Velocity fields of crust movement of china: comparison between results with GAMIT and GISPY. Journal of Geodesy and Geodynamics 23(3), 4–8 (2003)
16. Dongchen, E., Zhan, B., Jiang, W., Zhang, S.: High-precision GPS data processing by GAMIT/GLOBK. Chinese Journal of Polar Research 17(3), 173–182 (2005)
17. Lu, C., Wang, J.: Data processing for GPS crust movement monitoring network. Journal of Geodesy and Geodynamics 22(4), 56–60 (2002)

Seismic Spatial Information Grid: Applications of Geo-Informatics in Earthquake Disaster Management

Xiaohong Yang and Qiuwen Zhang*

College of Hydropower and Information Engineering, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan 430074, P.R. China
{yxh_hust,qwzhang_hust}@163.com

Abstract. Earthquake is one of the most serious natural disasters, which has caused huge casualties and economic losses to human society every year. It is very necessary and urgent to construct a platform to prevent and mitigate earthquake disaster. Compared with the traditional seismic data management methods that always only use a single isolated information technology, seismic spatial information grid(SSIG), which is a thematic spatial information grid with applications of geo-informatics in earthquake disaster management, has integrated many kinds of spatial information technology. In this paper, SSIG has been introduced into earthquake disaster management. It describes the construction, data acquisition, storage and application of SSIG. The proposed system was exemplified in Yushu earthquake in Qinghai province of China. The case applications show that the SSIG can effectively manage and analyze massive seismic data. As the system has run very stably, it can be recommended to a national level of China.

Keywords: spatial information grid, earthquake, disaster and risk management, geo-informatics, seismic thematic grid.

1 Introduction

Every year, the natural disasters bring great damage and injury to human society[1,2]. As one of the top ten natural disasters, earthquake has great impacts on the environment and society, and often caused vast economic losses and casualties. So how to mitigate and manage the damages is an important issue faced by mankind.

Earthquake disaster is affected by the geology, environment, society and other factors. Moreover, it is uneven distribution in time and space [3]. It is difficult to manage earthquake disaster. The traditional way is to construct a seismic management database to store the massive earthquake data. Later, the Geographic Information System (GIS) and Remote Sensing (RS) technology were introduced to manage the earthquake disaster. However, these management methods are isolated, disconnect technically. It is difficult to meet the needs of earthquake disaster management.

* Ccorresponding author.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 397–406, 2013.

Spatial information grid (SIG) integrates a variety of spatial information technology. It can organize and dispose spatial data in grid unit under the distributed network environment. Based on SIG, this paper tries to use it to manage earthquake disaster. SSIG has been proposed as an application of geo-informatics in earthquake disaster management. The system collect, store and manage different kinds of seismic thematic data and related data in grid unit. It implemented the acquisition, share, access, processing and analysis of massive seismic spatial information resources.

2 Characteristics and Requirements

2.1 Characteristics of Earthquake Data

Earthquake events occur dynamically both in time and space. The characteristics of seismic data are as follows.

(1) Spatial heterogeneity. Studies have shown that the spatial distribution of seismic activity is complex and random[4,5,6]. The distribution and occurrence of earthquake has a spatial heterogeneity. Different tectonic zone has different earthquake thematic data and various attribute data. Moreover, earthquake disaster risk is different from area to area.

(2) Multi-factor. Earthquake is closely related to many factors no matter in the cause or in the assessment after the disaster. Multiply factors could affect the earthquake disaster, such as tectonic, environment, economics and so on. These impact factors are different between areas. Therefore, the seismic data are massive and complex in both structure and storage.

2.2 Requirements of Earthquake Disaster Management

As the earthquake data are discrete, spatial heterogeneous, complex and multi-factor related, it needs to construct a relevant seismic spatial information management platform. The management system should organize and manage different kinds of information which distributed in different area. Firstly, it should build location-based model and divide the large block into lots of small range grids. Therefore, the seismic data in the same grid can be regarded as uniform. Secondly, every grid could collect and store the earthquake thematic data and various attribute data. The system also could manage and analyze the earthquake data in the grid unit.

3 The Principle of Seismic Spatial Information Grid

3.1 Definition of SSIG

Spatial information grid links all kinds of spatial information resources and realizes their sharing and utilization in wide network [7,8]. Seismic spatial information grid is

a thematic spatial information grid complied with the characteristics of earthquake disaster. It integrates various kinds of spatial information technology to achieve the earthquake disaster early warning, management and service. SSIG is a typical application of geo-informatics in earthquake disaster management. It manages massive earthquake data and other related data.

3.2 Types of SSIG

SSIG concludes both irregular and regular grid. Regular grids consist of a series of polygons with the same size and specifications. On the other hand, irregular grids are some arbitrary polygons with different size and shape. According to the characteristics of earthquake disaster, it needs to construct both irregular and regular ones. They are in different level, and carry different earthquake data as well as the impact factors information.

3.3 Construction of SSIG

Seismic spatial information grid construction is the foundation and key steps of the system. Building an irregular grid needs to create a new layer and then draw grid layer by GIS-based editing. The irregular grid is mainly used for the storage and management of some irregular seismic factor information. The proposed system has chosen the administrative district boundary as the irregular grid. Different levels of administration of the grid store different earthquake information.

Regular seismic spatial information grid is mainly used to assess earthquake losses under the equal area. Drawing regular grid needs to set parameter first. Parameter setting could be manual input or referring to the template layer or element. Then the first grid is drawn through the presetting size and location. The grids are cloned one by one until the interesting region is full of grids with the same size and shape (See Fig.1). Regular SSIG construction process is shown in Fig.2.

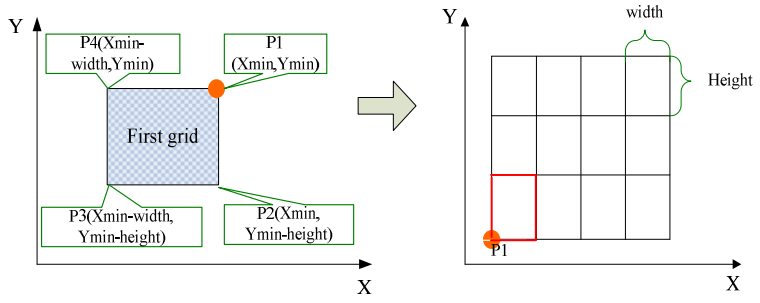


Fig. 1. The construction process of regular grid

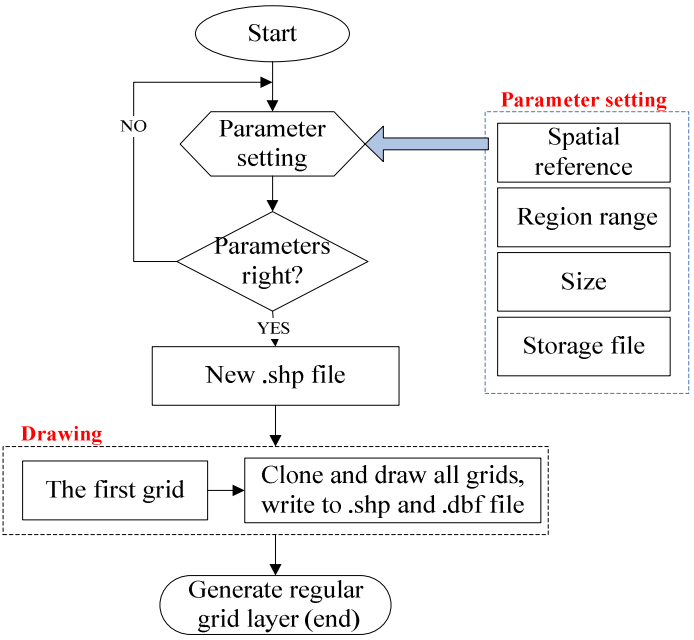


Fig. 2. Flow chart of regular grid drawing

3.4 Data Collection and Acquisition

The grid-based earthquake data collection and acquisition are both by manual input in the layer attribute table and automatic collection. Spatial overlay analysis is adopted in the automatic collection way. Overlaying the blank grid layer and the seismic thematic information layer can generate a gird layer with abundant seismic information. SSIG data acquisition process is shown in Fig.3.

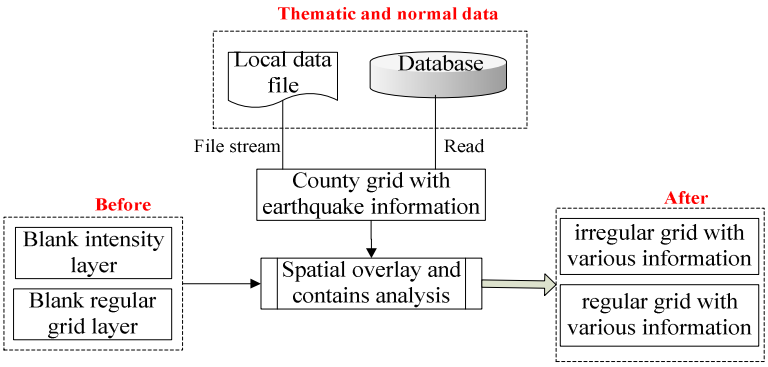


Fig. 3. The flow chart of seismic data acquisition in grid

3.5 Data Organization and Storage

As seismic data are massive, multi-factor related and spatial heterogeneous. It should be stored and managed with appropriate data structure in uniform data standard and spatial reference. This study proposed a mixed data organization model (see Fig. 4). It adopted both ArcSDE and SQL Server to organize and store data in grid unit. Geodatabase model is used to store various data in Relational Database Management System (RDBMS) such as SQL sever, oracle and so on. ArcSDE as a Spatial Data Engine(SDE) is adopted to extend the RDBMS. In this storage schema, all data would distribute into the corresponding grid with different resolution and spatial position.

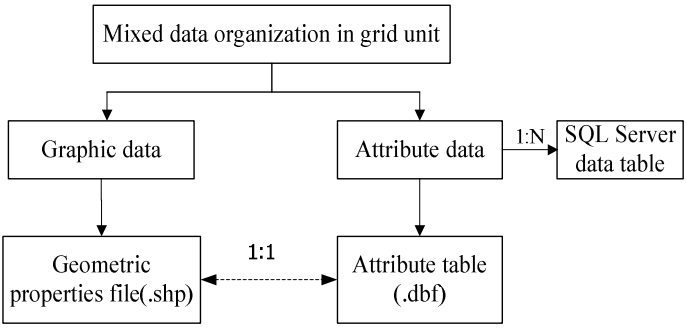


Fig. 4. The model of seismic spatial data organization

3.6 Data Query and Service

The seismic spatial information grid has stored geology, economy, buildings and other related information. It is massive and multi-source heterogeneous. To obtain useful and interesting data from these diverse data, it needs to have a good query function. Data query is an important module of SSIG.

The proposed system has given a bidirectional query way. They are the query from attributes to graph and the query from graph to attributes. The former is some function like identify in ArcGIS. The later refers to the relational database query function and Spatial Data Engine (SDE). From the grid data search, statistics and analysis, it can dig out deeper information from the earthquake data. Some applications such as intensity map drawing and earthquake disaster assessment are provided by SSIG.

4 Application

4.1 Study Area

On April 14, 2010, the Ms7.1 earthquake occurred in Yushu county of Qinghai province in northwest China. The epicenter intensity was nine. The epicenter is mainly located in 96.6 degrees east longitude and 33.1 degrees north latitude.

It belongs to the third region. In this earthquake, 2698 people died, and huge economic losses are caused in the disaster areas. The focal depth of Yushu earthquake is very shallow and the aftershocks are quite frequent.

4.2 Intensity Map Drawing Based on SSIG

Seismic intensity reflects the damage degree of ground and buildings in earthquake disaster [9]. It is related to location, magnitude, epicenter distance and geotechnical properties [10]. In earthquake assessment, isoseismal lines and seismic intensity attenuation map are always used in statistics analysis of the casualties, building damages and other losses in the same seismic intensity.

In this study, ellipse intensity attenuation model is applied in the fast drawing of an intensity map. Before drawing an isoseismal line, some model parameters and variable values need to obtain. Model parameters were stored in province irregular grids, and the angles of rupture were stored in regional irregular grids. Magnitude and intensity values were input at the client. Then through a loop program, it could draw isoseismal until the intensity is equal to six. When the intensity is less than six degree, the earthquake damage is very small and can be ignored. Fig.5 shows the process to draw a seismic intensity attenuation map. Fig.6 shows the seismic intensity attenuation map of Yushu earthquake.

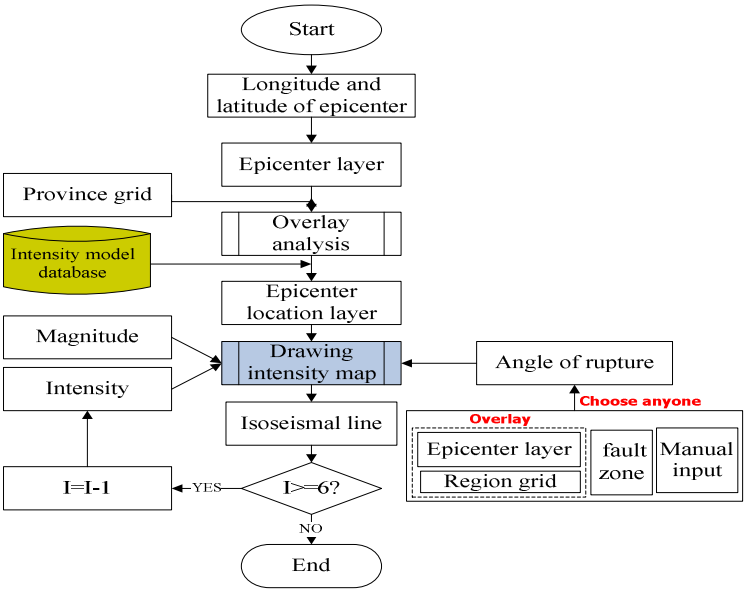


Fig. 5. The flow chart of drawing a seismic intensity attenuation map

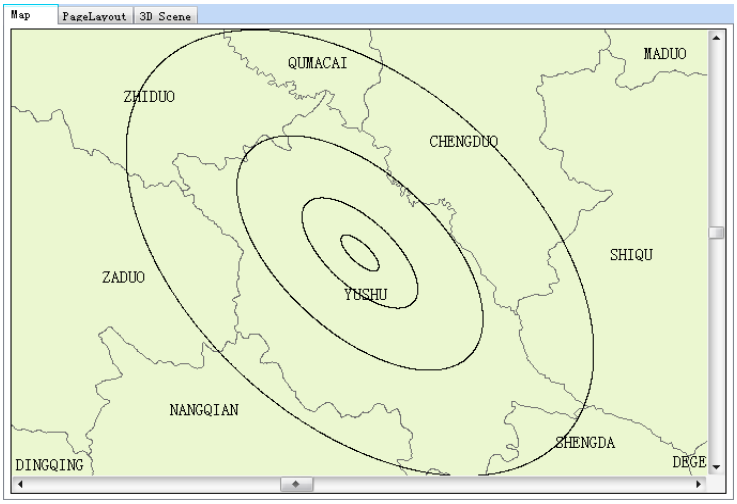


Fig. 6. The seismic intensity attenuation map of Yushu earthquake

4.3 Regular Grid Generation of Yushu Earthquake Disaster

According to Fig.2, drawing regular gird needs to know spatial reference, region range, grid size and storage file first. The spatial reference and region range can refer to layer template. As the Fig.6 has already drawn the seismic intensity attenuation map, the regular grid can use its spatial reference and be divided through the intensity map range. The regular grids of Yushu earthquake with ten rows and ten columns are shown in Fig.7.

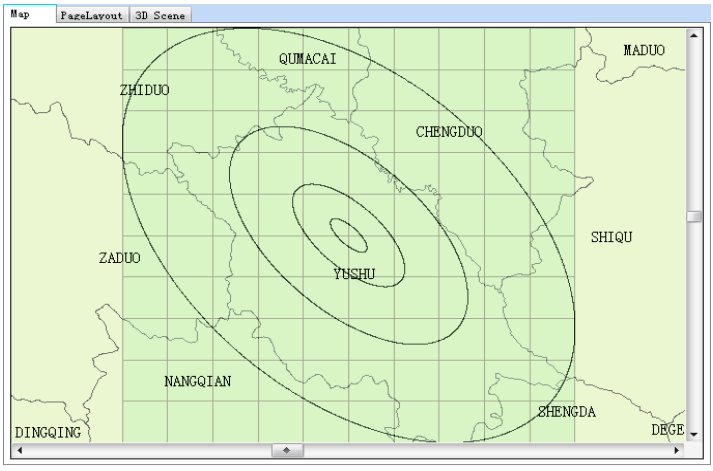
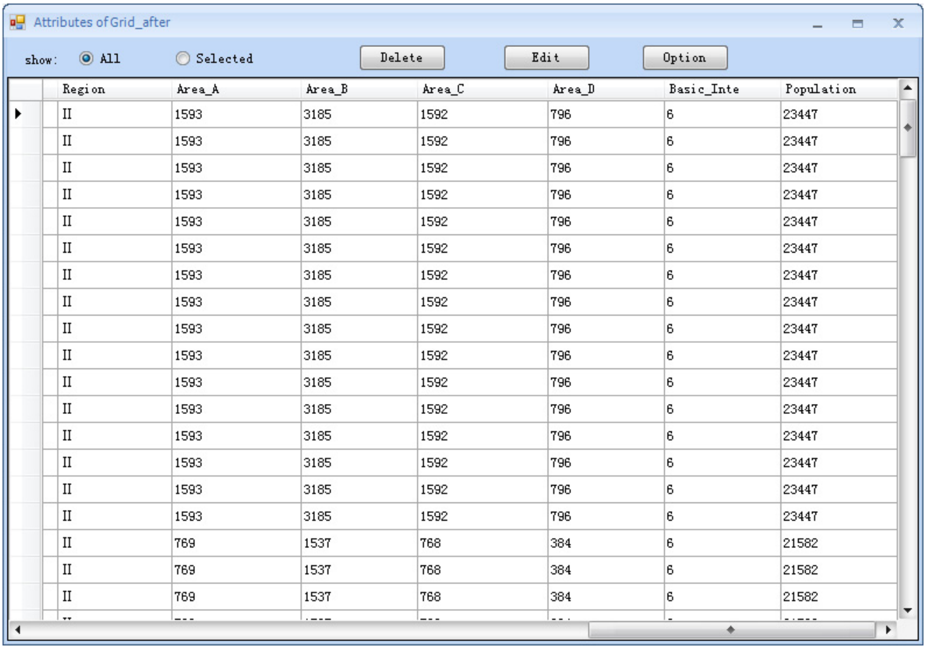


Fig. 7. Regular grids of Yushu earthquake with isoseismal lines overlaying

4.4 Data Acquisition

The above regular grid layer is a blank grid layer without any seismic thematic data. It should collect data to meet the needs of disaster management. Overlaying the blank grid layer and the county irregular grid layer with seismic thematic information can generate a gird layer with abundant seismic information. After the data acquisition, a lot of fields are added into the attributes table of the grid, which can be used to store the earthquake thematic data (see Fig.8).



Region	Area_A	Area_B	Area_C	Area_D	Basic_Inte	Population
II	1593	3185	1592	796	6	23447
II	1593	3185	1592	796	6	23447
II	1593	3185	1592	796	6	23447
II	1593	3185	1592	796	6	23447
II	1593	3185	1592	796	6	23447
II	1593	3185	1592	796	6	23447
II	1593	3185	1592	796	6	23447
II	1593	3185	1592	796	6	23447
II	1593	3185	1592	796	6	23447
II	1593	3185	1592	796	6	23447
II	1593	3185	1592	796	6	23447
II	1593	3185	1592	796	6	23447
II	1593	3185	1592	796	6	23447
II	1593	3185	1592	796	6	23447
II	1593	3185	1592	796	6	23447
II	1593	3185	1592	796	6	23447
II	769	1537	768	384	6	21582
II	769	1537	768	384	6	21582
II	769	1537	768	384	6	21582

Fig. 8. Attribute table of grid layer after seismic data acquisition

4.5 Earthquake Data Querying

For the demand of disaster management, it is necessary to query earthquake information in space and time. The querying interface provides a window to input the SQL statements. Interested readers are encouraged to read the relational database management system (RDBMS) documentation[11] in order to undertake more complex SQL queries. For an example, if a user wants to search the six intensity area in casualty, he needs to choose the target layer and then set "Intensity=6". The system would generate a SQL command and a new table of the interest. Then the results of the selected features corresponding to the query command are visually displayed with highlight in the map (see Fig.9).

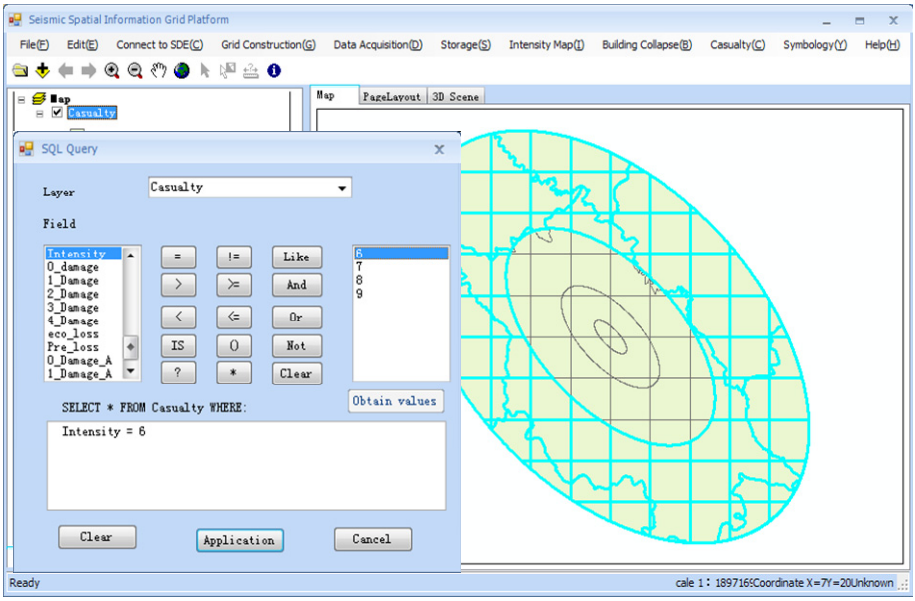


Fig. 9. Visual data querying from attributes to graph

5 Conclusion

This paper provides seismic spatial information grid that integrate various geo-informatics to manage the earthquake disaster. SSIG divides a large area into a plurality of small range to solve spatial heterogeneity of seismic data. By using spatial information technology, SSIG can effectively collect, store, manage and analyze various data in the earthquake disaster.

Although the platform is only exemplified in Yushu earthquake in China, it can be extended to a national or global level so as to manage the seismic data in an effective way and make best use of them. The proposed framework of SSIG in this paper is just an attempt on application of geo-informatics in the earthquake disaster management. As the seismic data is massive and complex, more suitable grid with different sizes and structures will be designed to adapt the special characteristics of the earthquake management in the future study.

Acknowledgments. The authors would like to appreciate the financial supports for this study from the National Natural Science Foundation of China (Grant#41072199, 41101258), the Natural Science Foundation for Outstanding Scholarship of Hubei Province in China (Grant#2008CDB364), the National Key Technology R&D Program of China (Grant#2008BAC36B01) and the Program for New Century Excellent Talent of Ministry of Education of China (Grant#NCET-07-0340).

References

1. Montoya, L., Masser, I.: Management of natural hazard risk in Cartago, Costa Rica. Habitat International 29(3), 493–509 (2005)
2. Karimi, I., Hüllermeier, E.: Risk assessment system of natural hazards: A new approach based on fuzzy probability. Fuzzy Sets and Systems 158(9), 987–999 (2007)
3. Haya, H., Nasu, M.: Effect of uneven thickness of soft ground on earthquake damage to building. Railway Technical Research Institute, Quarterly Reports 31(4) (1990)
4. Shijun, C.: Study on earthquake cluster in different tectonic environment. Institute of geology, China Earthquake Administration Ph.D. dissertation, Beijing (2004)
5. Yuxian, H.: Seismic safety evaluation technology tutorials. Earthquake Press, Beijing (1999)
6. Shijun, C., Li, M., Yuansheng, L.: Uneven spatial distribution of seismic activity in the multi-scale analysis. Progress in Geophysics (2005)
7. Deren, L., XianYan, Z., Jianya, G.: From Digital Map to Spatial Information Grid. Journal of Wuhan University 28(6), 642–649 (2003)
8. Deren, L., et al.: Research on grid division and encoding of spatial information multi-grids. Journal of Surveying and Mapping 35(1), 52–56 (2006)
9. Bakun, W.H., Wentworth, C.M.: Estimating earthquake location and magnitude from seismic intensity data. Bulletin of the Seismological Society of America 87(6), 1502–1521 (1997)
10. Kircher, C.A., Whitman, R.V., Holmes, W.T.: HAZUS Earthquake loss estimation methods. Natural Hazards Review 7(2), 45–59 (2006)
11. Wu, D., Begg, C., Houston, C., Miller, E.: Selecting a RDBMS for a multi-disciplinary clinical research database. Controlled Clinical Trial 13, 418–419 (1992)

Design on Early Warning System of CO2 Sequestration Leakage Based on Web GIS and WSN

Xin Liu[1,2], Shaoliang Zhang[2,3], Aihua Yan[3], and Huping Hou[2]

[1] School of Environment Science and Spatial Informatics,
China University of Mining and Technology, Xuzhou, Jiangsu 221008
[2] School of Medicine Information of Xuzhou Medical College, Xuzhou, Jiangsu 221004
[3] Low-carbon Energy Research Institute of China University of Mining and Technology,
Xuzhou, Jiangsu 221008

Abstract. For the leakage risk of sequestration CO2, we use the wireless sensor for monitoring data, such as CO2 concentration, the wind speed, and then the data are transmitted to the on-line monitoring system. Through the analysis of monitoring data, we may discover the leakage site. We use the Web GIS technology combined with Gauss point source diffusion model to simulate the CO2 leakage .According to the simulation results, we can divide the level of the accident, make the rescue plan formulation quickly, release early warning information, and the damage will be reduced to a minimum.

Keywords: WSN, Web GIS, CCS, Early Warning.

1 Introduction

As global warming, CCS (Carbon Capture and Sequestration, "CCS"), as a kind of new technology for CO2 emissions reduction, is presented in recent years, , CCS is injected CO2 into oil and gas reservoirs, deep salt water layer, and the dry coal seam geological reservoir, can achieve part of the permanent underground storage of CO2. As a new and has the massive carbon emissions reduction potential technology, CCS is expected to become the important link in future global response to climate change and realize low-carbon economic transition[1].

The CCS project has the risk of leakage[2].if the Carbon dioxide sealed in salty water layer or abandoned oil leaks and gas deposits, and other geological structure, through the fault、fracture or artificial drilling hole leak to the surface, it will offset CO2 contribution to mitigating climate change [3], what is more important, in view of the CCS projects of space scale, short term or long term leakage may likely make a significant impact to the local health, safety and environment[4,5].At the same time, because of China's fragile ecological environment, meteorological and geological disasters occur frequently, the risk of CCS in china is much higher than other parts[6]. In fact, for fear of the ecological environment problems caused by CO2 leakage, so many people against the CCS, this has become one of the important obstacles of CCS activities [7].

Therefore, in order to guarantee the implementation of CCS project, we should calls for the dynamic monitoring for the whole project implementation process, on the

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 407–413, 2013.

basis of the monitoring, we can implement early warring. Research suggests that study of CO2 sequestration leakage warning is still blank. So we mainly through the analysis of monitoring data obtained by wireless sensor technology(soil, near-surface atmospheric CO2, groundwater), find the leak point, and then through GIS and gauss point source diffusion model for large-scale simulation of leakage, take timely warning for possible damage, and make decision for the emergency.

2 The Basic Goal

CO2 leakage early warning, is to analyze and evaluate the change of a certain range CO2 concentration, for a period time, to predict its future development and determine its distribution and change trend in a particular time, and then forecast mass leakage's time-space range and damage degree. According to the need, we can give the alert information of the change or degradation, that is to say, we can make solution for the problems, or put forward measures or warning information for the upcoming problem. Ideal warning contains not only for a moment, but also include the early warning of mass diffusion leakage. It should be a proactive, predictive function [8].

3 Based on WSN and Web GIS Sequestration of CO2 Leakage Warning Process

The early warning of CO2 leakage, includes monitor, date analysis, leakage simulation, information release, is a complicated process because of its sudden characteristics. The specific process is showed as follow:

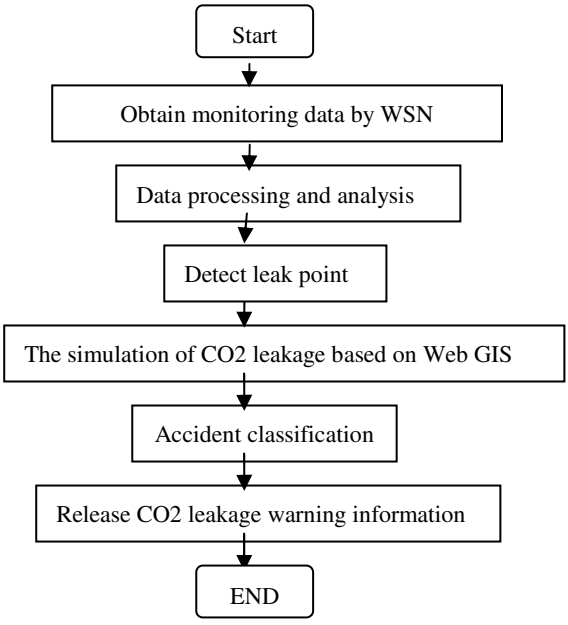


Fig. 1. Storing CO2 leakage warning flowchart

(1)Obtain monitoring data by using WSN
In order to ensure the implementation of the project under security control, it's necessary to monitor it. We focused on the monitoring of CO2 concentration, wind speed, wind direction, soil, PH value of water. In addition of groundwater, we use wireless sensor , which is designed by ourself to monitor automatically.

(2) Processing and analysis of monitor data
The monitor of background value has been conducted before the project is not implemented. But because of the carbon dioxide is existed in the air, soil and groundwater, the concentration of co2 is affected by so many factors, such as temperature, season, time and so on . Therefore we should process and analysis the obtained monitoring data, grasp its variation. If abnormal, we could alert message according alarm value, and to determine the location of the leakage.

(3)Leakage diffusion simulation based on Web GIS
According to the leaks found, combine the local spatial information with the selected atmospheric dispersion model, we can simulate the leakage diffusion based on Web GIS, grasp the tendency to spread dynamically.

(4) Accident Classification
According to the simulation results and the corresponding level of the accident, determine the accident level quickly.

(5) Release the warning information
Using the Web GIS platform to send warning messages to related person, the main contents include the leakage time, location, direction of diffusion, etc.

4 Design of Storing CO2 Leakage Warning System Based on Web GIS and WSN

According to analysis of the warning leakage of CO2 sequestration process, we establish a CO2 leakage early warning system based on WSN and Web GIS. The system has four main modules: (1) System Management Module. This module mainly organizes and manages the attribute database, spatial database, on-line monitoring database. (2) Analysis monitoring data module. The module is aimed to preprocess and analysis the obtained data, process the abnormal data and missing, and then discover the leakage.(3)Atmospheric dispersion modeling module. Combined with GIS and selected model for atmospheric diffusion to simulate CO2 leakage, based on simulation result to determine the level of accident.(4)Release warning information module. Using the network platform, combined with responsible person in the property database, put forward short message and warning information. The system architecture is shown as follows:

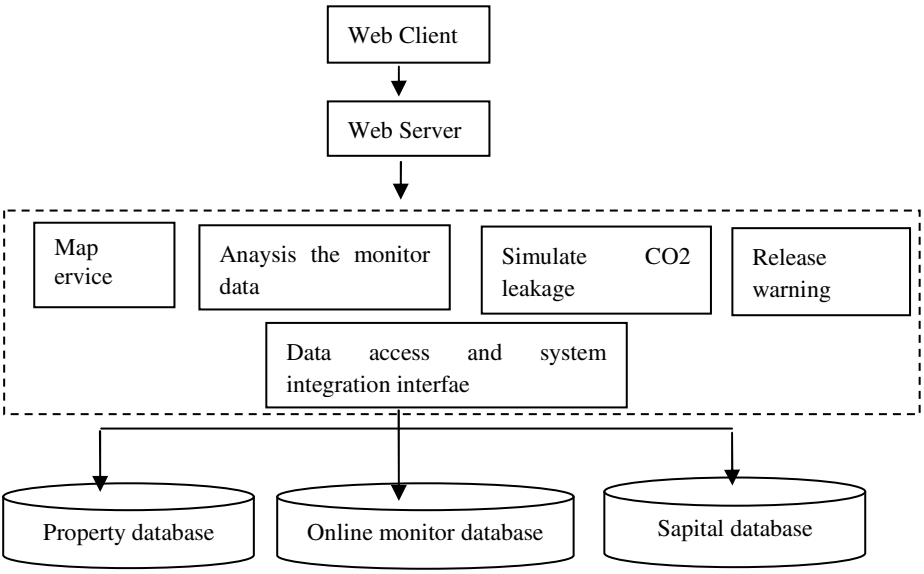


Fig. 2. System architecture chart

5 Storing CO2 Leakage Warning System Key Module Design

5.1 Monitoring Data Acquisition Module

Based on their possible impact the ecological environment, the near-surface region for CO2 sequestration monitoring, we focused on groundwater, soil and near-surface CO2 concentration monitoring, the specific content of the table as follows:

Table 1. Near-surface CO2 sequestration monitoring methods and content

Monitoring Methods	Monitoring contents	Quantity	Monitoring Frequency	methods
Groundwater	PH value, calcium and magnesium ions, nickel, manganese, iron, copper and other heavy metals	Two	Injection period：two months / per Non-injection period：two months / per	Artificial
Soil	CO2 concentration, temperature, humidity,	Three	Injection period：thirty minutes /per two months / per：one hour/per	Automatic
Near-surface concentration	CO2 concentration, temperature, humidity, wind speed, wind direction	ten	Injection period：thirty minutes /per two months / per：one hour/per	Automatic

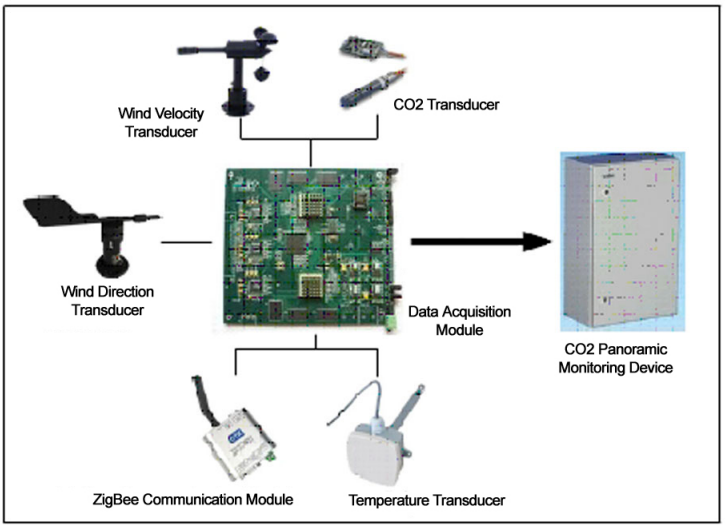


Fig. 3. CO2 concentration monitoring device exploded Chart

Automatic monitoring wireless sensors which can automatically sense the temperature, concentration, humidity, wind speed, wind direction and other rapid change paratemeters is developed by our self, with real-time and high precision monitoring advantages. It's good at unattended region to monitoring data. It's shown as follows. Through the ZigBee module, the data can be communicated by wireless network via GPRS DTU to the online monitoring system.

5.2 Storing CO2 Leakage Diffusion Simulation Module Based on Web GIS

Web GIS is a new technology combined with the Web Technology and GIS technology. The system will be developed by the .NET technology, which integrate the atmospheric dispersion models and Google maps.

Atmospheric pollution diffusion model variety, we choose Gaussian model, which is the "Environmental Impact Assessment Technical Guidelines-Atmospheric Environment(hereinafter referred to as "guidelines")" recommended model, has simple, less input data, calculation accuracy meet the requirements characteristics.

Refer to "guidelines", using the traditional point source plume (instantaneous) mode, using the following formula:

$$C(x,y,0,t)=\frac{Q}{(2\pi)^{3/2}\sigma_X\sigma_Y\sigma_Z}\exp(-\frac{(x-ut)^2}{2\sigma_x^2})\times\exp(-\frac{y^2}{2\sigma_y^2})\{\exp(-\frac{h^2}{2\sigma_z^2})+\exp(-\frac{4H^2}{2\sigma_z^2})\}$$

t is the time, the units is seconds, S;

Q is the speed of leakage, mg/s;

u,v,w represent the main wind direction, horizontal direction and vertical wind velocity vector values, m/s.

σ_x、σ_y、σ_z is the horizontal and vertical diffusion coefficient, is the standard deviation in x, y, z direction of the pollutants, m.

h is the high of the accident locations ,m.

H is the high of the mixed-storey chimney, m.

According to he formula, we can input various parameters, such as wind speed, wind direction, the leakage address, the time ,the leakage rate, etc. CO2 leakage based on Web GIS diffusion simulation process is as follows:

(1)The user input the atmospheric diffusion parameters, establish the Gaussian coordinate system which origin is the leak location.

(2)Using the coordinate conversion formula to convert the Gaussian coordinates to geographic coordinates.

(3) Make the leak location as the origin of coordinates, established fan-shaped area, and screen for calculating points.

(4)Using the C# programming atmospheric dispersion model, calculated the value of the discrete points.

(5) Interpolation of discrete points.

(6) Using the ellipse method to generate the approximate contours.

5.3 Release Warning Information Module

According to the leakage simulation results and the disaster classification criteria, we can determine the level of the accident. The Web GIS system can use the network and short messages to post the warning information.

6 Conclusions and Prospects

Compared with the traditional technology, wireless sensor has the incomparable advantage in unattended environmental monitoring, disaster relief field. On the basis of research results at home and abroad, we combine the GIS and WSN technology effectively, and establish a low-cost, practical early warning system for CO2 leakage. In the future, we will run the system to obtain relevant experimental result, and then improve the system.

References

[1] Xu, Z.: Monitor of the distribution of CO2 geological storage conditions and environmental impact. Advances in Climate Change Research 4(6), 363–368 (2008)

[2] Govindan, R., Korre, A., et al.: A geostatistical and probabilistic spectral image processing methodology for monitoring potential CO2 leakages on the surface. International Journal of Greenhouse Gas Control 5, 589–597 (2011)

[3] Zhang, Q.: Analysi of CCS monitor technology development. Shenhua Science and Technology 9(2), 77–82 (2011)
[4] Siirila, E.R., Navarre-Sitchler, A.K., Maxwell, R.M., McCray, J.E.: A quantitative methodology to assess the risks to human health from CO2 leakage into groundwater. Advances in Water Resourse, doi:10.1016/j.advwatres.2010.11.005
[5] Bogen, K., Burton, E., Friedmann, S.J., Gouveia, F.: Source terms for CO2 risk modeling and GIS/simulation based tools for rish characterization. In: 8th Greenhouse Gas Technology Conference, Trondjheim, Norway, Session E1-3 (2006)
[6] Newmark, R.L., Friedmann, S.J.: Water Challenges for Geologic Carbon Capture and Sequestration. Environmental Management 45, 651–661 (2010)
[7] Hortle, A., de Caritat, P.: Groundwater monitoring at the Otway Project site, Australia. Energy Procedia 4, 5495–5503 (2011)
[8] Xin, H.: The mathematical theory of early warning. Wuhan University of Technology 26(2), 195–198 (2002)
[9] SEPA. Technical Guidelines for Environmental Impact Assessment – Atmospheric Environment. China Environmental Science Press, Beijing (1993)
[10] Shao, C., Ju, M.: Sudden atmospheric pollution incidents environmental risk assessment and management. Environmental Science & Technology 329(6) (2009)

Urban Feature Extraction and Terrain Build with Large Scale Topographic Databases*

Xiaokun Zhu[1,2]

[1] Photogrammetry and Remote Sensing Centre, Beijing Insititute of Surveying and Mapping,
15, Yangfangdian Road, Haidian District,
100038 Beijing, China
[2] Institute of Remote Sensing and Digital Earth Chinese Academy of Science,
9, Dengzhuangnan Road, Haidian District,
100094 Beijing, China
zhuxk@bism.cn

Abstract. The urban terrain surface is much more complex to build than mountainous regions-discontinuous, easily covered by other objects as there are high density settlements in urban area. The feature datasets in topographic databases, usually acquired and updated both by photogrammetry and field work, are helpful to build the urban terrain surface. So in the paper, a workflow of urban feature extraction and terrain build with large scale topographic databases was proposed and designed. The two dimension (2D) buildings, road and water body features were extracted from the databases with three dimension (3D) terrain features, a elevation assignment and interpolation method was presented to construct the discontinuous urban terrain surface, the relative sub-modules including feature extraction, elevation points editing, 2D to 3D, elevation interpolation and checking were developed and introduced. At last, based on the urban flood disaster analysis requirements, an experiment was carried out to verify the feasibility of this workflow and the method.

Keywords: Digital Elevation Model (DEM), Topographic Databases, Feature Extraction, Terrain Build, Urban Terrain Surface.

1 Introduction

Digital Elevation Model (DEM) is one digital format of terrain surface. As one of the 4D products, it has been widely used. Terrain feature extraction and DEM construction methods include photogrammetry method, map vectorization method, airborne LiDAR measurement method, etc.[1]. No matter which method is used, all the terrain features including elevation points, contours, feature points and break lines are used to create TIN (Triangle Irregular Network), which is finally exported to DEM. Based on the former research and application, all the methods are practical for mountain areas. However, in the urban areas, there are much more complex terrain surface: discontinuous and seriously covered by artificial buildings and structures,

* This project is supported by 2013 Beijing Technology Foundation for Selected Overseas Chinese Scholar.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 414–423, 2013.

all the methods are difficult to produce the urban DEM, which is important and much useful to the digital city construction and urban terrain analysis.

The urban institute of surveying and mapping periodically carried out the large scale (1:500, 1:2000, 1:10000) topographic data updating, processing and database management. These databases contain more information including indoor interpretation and field work. So based on the feature datasets in topographic databases, the urban terrain surface build work will be much easier. In this paper, a workflow and relative methods of urban feature extraction and terrain build with large scale topographic databases will be given and based on the urban flood disaster analysis requirements, an urban terrain build experiment will be carried out.

2 Workflow

This urban feature extraction and terrain build workflow with topographic databases refers to the map vectorization DEM production method, that is, to extract the terrain features from the topographic databases, then to construct TIN and to create the final DEM grid [2]. The work flow draft is shown in Fig. 1.

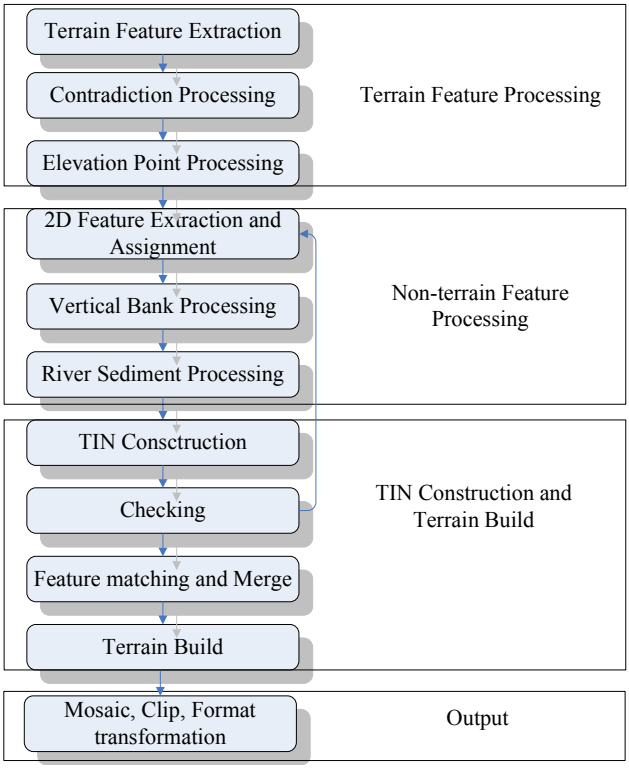


Fig. 1. Urban Feature Extraction and Terrain Build Workflow

2.1 Terrain Feature Processing

Terrain Feature Extraction. There are elevation points (including relative elevation points), contours (index contours, intermediate contours, supplemental contours, etc.), and break lines (or polygons) in the large scale topographic databases and all these useful terrain feature datasets with their 3D information of the research areas should be extracted to build natural terrain. The features extracted should be classified to points, hard lines, soft lines, polygons for TIN creation and adjoining sheets also should be merged according to same layers.

Contradiction Processing. Because different scale databases are used, so all the contradiction problems among these databases should be considered, such as different contour intervals, contradictive elevation points. In this step, all the redundant elevation points should be deleted, and contour and elevation point contradiction caused by generalization of different scale databases should be solved.

Elevation Point Processing. All the elevation points preserved should represent the urban terrain surface to construct DEM. The points on buildings, pedestrian overpass, overpasses, underpass, or other structures which are connected to the terrain surface with piers, should be deleted.

2.2 Non-terrain Feature Processing

2D Feature Extraction and Assignment. In the urban areas, buildings, roads and water bodies are the main non-terrain features constructing the urban surfaces. So in order to carry out urban terrain build and analysis especially in the urban settlements, these main features should be used. Normally, these feature datasets in the topographic databases are 2D, that is, no elevation information for these important features. In order to assign elevation to these features, firstly, they should be extracted. So all the 2D curb lines of highway, railway, ring road, etc., all 2D range lines of building block and all 2D river bank lines and hydrologic lines of water body areas including rivers, lakes, channels, pools, etc., are extracted from the large scale topographic databases.

All these 2D lines will be exported to 2D polygon features, which are named target areas below. With those terrain features extracted above, especially the elevation points along the 2D line features, the elevation values of the neighbor elevation points are projected on the 2D lines, which are later exported to continuous 3D feature lines after interpolation. There are two special situation we should consider: one is vertical bank, the other is the river sediment.

Vertical Bank Processing. Because Delaunay TIN is difficult to describe the vertical bank, two close feature polygons which are top and bottom of the river banks, are separately constructed and assigned. The method is also practical for the overpass vertical banks.

River Sediment Processing. For the River sediment, if there are some elevation points with the bottom elevation, the hydrologic lines should be assigned with the bottom elevation; if there are no elevation points labeled in the river, the hydrologic lines will be assigned with the bank points elevation minus relative elevation (such as 3.5m). All the vertices will be assigned the elevation value by interpolation.

2.3 TIN Construction and Terrain Build

TIN Construction. With the software ArcGIS[3], all the terrain features, and 3D non-terrain features are applied to create TIN. The elevation attributions of the elevation points and contours are used as height sources, while the Z values of the 3D features are used as height sources.

Checking. The TIN generated will be checked with Digital Ortho Map (DOM) and the contours in ArcMap and ArcScene, all the errors and contradictions will be located and corrected, and inverse computation from TIN to contours is also the useful method to check the final results.

Feature Matching and Merge. All the features from sub-blocks of the task should be matching and the errors for the adjoining sheets should be within the error limitation. Then the same layer should be merged and adjoining sheets should be mosaiced.

Terrain Build. After checking and feature editing, the large TIN could be exported to DEM, the final terrain result.

2.4 Terrain Surface Overlapped with Buildings

Sometimes, the building 3D models need to be added on the terrain surface for the further analysis, such as urban height analysis or catchment analysis. It is a kind of processing product of 3D Digital Surface Models (DSM). So the final terrain surface overlapped with buildings is acquired by adding the building height on the terrain TIN.

Building Height Estimation. The Building Height can be estimated with the floor numbers, which can be extracted from the buildings datasets in the databases. With the average building height, the building height can be estimated.

Building Elevation Assignment. All the building elevation can be calculated by adding the building height and terrain elevation. The final elevation data are added to the attribution of the building features.

Layer Overlap Analysis. The building 3D features will be added on the terrain surface by layer overlap analysis. This step can be carried out on TIN or DEM.

Checking. All the terrain surface overlapped with buildings can be checked the way same as the terrain results in 2.3.

3 Methodology and Realization

Based on the workflow mentioned above, the methods are realized with VBA in ArcGIS9.X [4] and the main interface is shown in Fig.2. It contains five sub-modules: Feature Extraction sub-module, Elevation Points Editing sub-module, 2D->3D sub-module, Elevation Interpolation sub-module and Checking sub-module.

Fig. 2. Main Interface of the Urban Feature Extraction and Terrain Build

3.1 Feature Extraction

The Feature Extraction sub-module, which is shown in Fig.3, is developed to extract the 3D terrain features and 2D non-terrain features from the large-scale databases and rename the useful layers. It can be set before feature extraction, which is much more flexible. Normally the code and elevation attributions are kept and the contradictions among different scale terrain features are automatically processed in this sub-module. From Fig.3, the 3D terrain features including contours, elevation points are extracted. And the 2D non-terrain features including water body, road and buildings are extracted too.

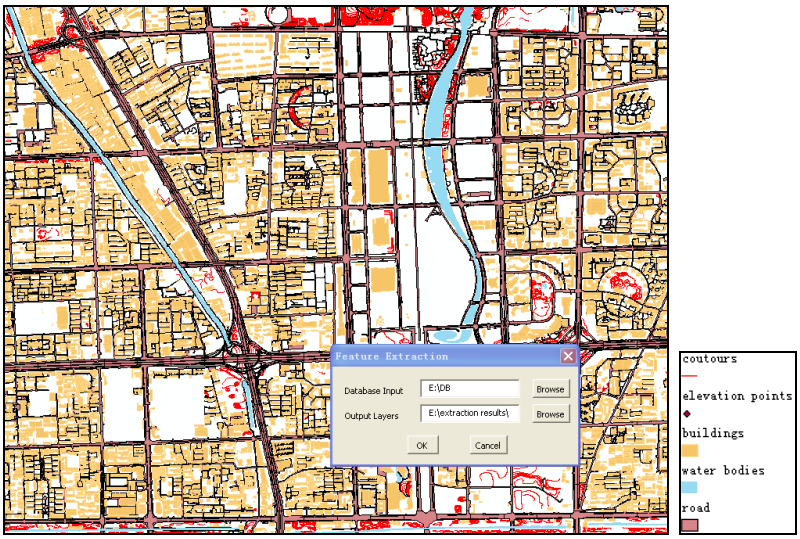


Fig. 3. Feature Extraction Sub-module

3.2 Elevation Points Editing

The Elevation Points Editing sub-module is developed to extract the elevation points on the artificial structures including buildings, pedestrian overpass, overpasses, underpass, etc. by overlapping analysis principles and separate these points from the natural terrain elevation points. These points are still useful to assign the elevation values. This sub-module is shown in Fig.4, in which there are some points are the automatically extracted points.

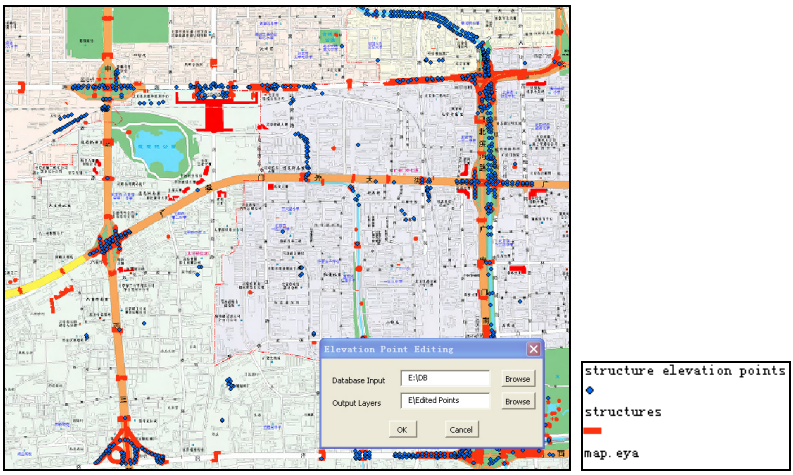


Fig. 4. Elevation Point Editing Sub-module

3.3 2D->3D

The 2D->3D sub-module is developed to assign the elevation values to the vertices of the building, road, and water body feature liens or polygons with the neighborhood elevation points. The vertical banks will be separated into two sides, inner side will be assigned with the inner points and outside will be assigned with the outside points. The elevation value is assigned by projection method, which is shown in Fig.5.

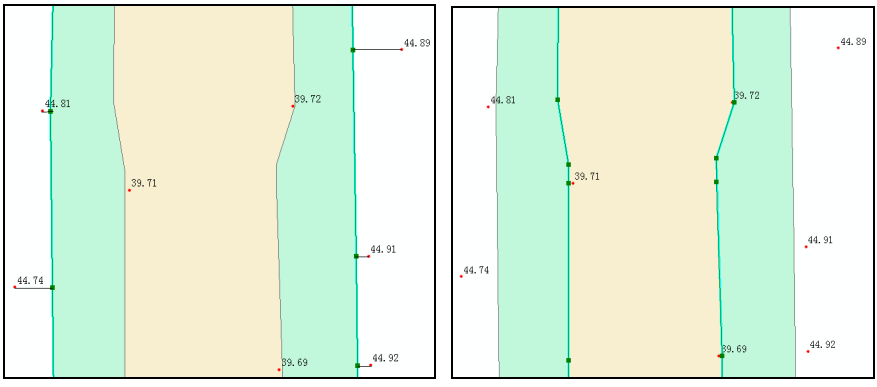


Fig. 5. 2D->3D Sub-module

3.4 Elevation Interpolation

The Elevation Interpolation sub-module is developed to interpolate the elevation values for the line vertices or polygon vertices, which are not assigned elevation

values by the neighbor elevation points [5]. The principle is shown in Fig.6. $A_0, A_1, \cdots A_n$ are the vertices of the feature lines, $d_1, d_2, \cdots d_n$ are the distances between different neighbor vertices, the weight of vertice k P_k and relative elevation value Z_k are calculated by equation (1):

$$P_k = \sum_{i=1}^{k} d_k / \sum_{i=1}^{n} d_i \quad Z_k = Z \pm P_k \Delta Z \tag{1}$$

ΔZ is the elevation difference between points A_0 and A_n which have had the elevation values; Z is the elevation value of point A_0.

By elevation interpolation, the 3D feature lines are much smoother and much more logical, which is shown in Fig.6.

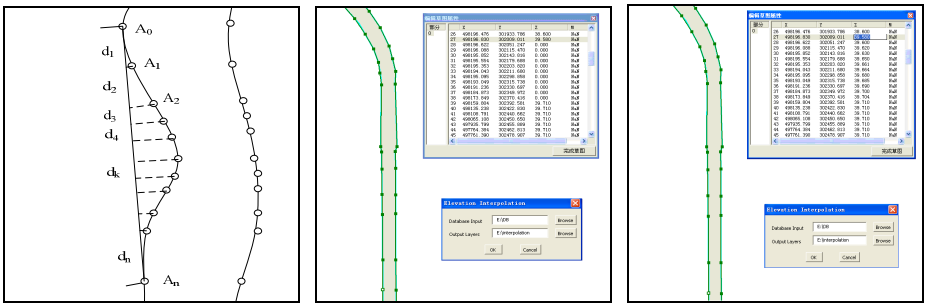


Fig. 6. Interpolation Sub-module

3.5 Checking

The checking sub-module is developed to compare the elevation values of the terrain feature points with those of DEM at the same location. The accuracy is evaluated with equation (2) [6]. The sub-module can also export the quality checking report.

$$M = \sqrt{\frac{\sum_{i=1}^{n} \Delta_i^2}{2n}} \tag{2}$$

M is the RMSE (Root Mean Square Error) of DEM product, n is the total number of the sample points, Δ_i is the difference between the two data sources.

4 Experiments

On 21th July 2012, there is a big rain storm in Beijing and there are many flooded places in urban areas. In order to analyze the reasons of the urban flooding, after the storm, the workflow of terrain build based on the large scale topographic databases is proposed and terrain of 22 flooded areas are analyzed and built with this workflow. Combining the DOM, around the flooded location, a 5km buffer ranges are set. The flooded areas are distributed in Fig.7.

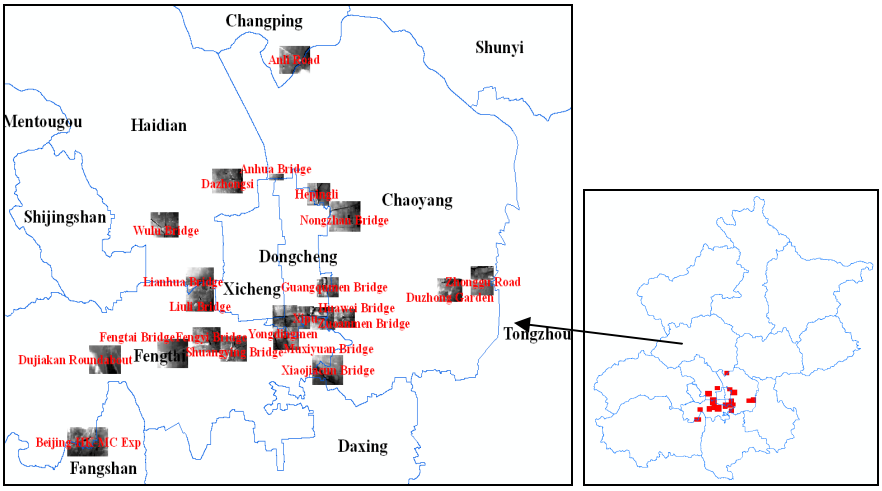


Fig. 7. Distribution of Flooded Places

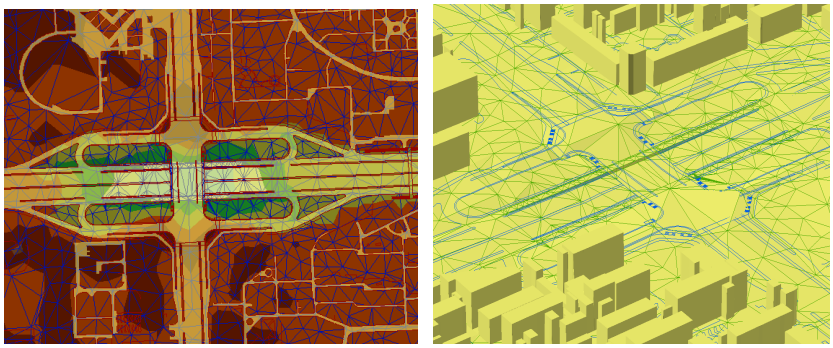


Fig. 8. Terrain Build with Buildings for Anhua Bridge

All the flooded places are within the plain areas of Beijing. 1:500 and 1:2000 large scale topographic databases cover these areas and are updated every year. With the workflow and the programs, we accomplished the task within 24 hours. The final

terrain surface of Anhua Bridge with buildings is shown in Fig.8 and the final accuracy evaluation results for the 22 flooded places are listed in Table.1.

Table 1. Accuracy Evaluation Results for 22 Flooded Places (Unit: meters)

Flooded Areas	RMS	Flooded Areas	RMS
Anhua Bridge	0.01	Yongdingmen	0.07
Wulu Bridge	0.02	Xipu	0.02
Duzhong Garden	0.02	Huawei Bridge	0.03
Lianhua Bridge	0.02	Dazhongsi	0.03
Liuli Bridge	0.03	Hepingli	0.02
Fengtai Bridge	0.04	Nongzhan Bridge	0.03
Fengyi Bridge	0.03	Zhonggu Road	0.02
Shuangying Bridge	0.02	Beijing-Hong Kong-Macau Expressway	0.01
Zuoanmen Bridge	0.02	Guangqumen Railway Bridge	0.03
Xiaojiacun Bridge	0.01	Dujiakan Roundabout	0.02
Muxiyuan Bridge	0.07	Anli Road	0.07

5 Summarization

The experiments verified the feasibility of the workflow and the methodology, which are later applied to 1414 km^2 plain area terrain build. This workflow and method making use of the data from the topographic databases, is much efficient especially for emergency and can construct the complex terrain surfaces of the urban areas. There are still some problems and some improvements should be carried out in future.

Firstly, the accuracy of the final terrain results depends on the location. The elevation points and contours from topographic map are the main elevation reference and the other elevation information is based on these data with assignment and interpretation. Therefore, the accuracy is higher for the location of terrain features but lower for non-terrain features.

Secondly, map generalization and interpretation leads to the difficulties to build the terrain surface of some complex bridges. It is much easier to construct 3D terrain surface for the concave-down flooded road. However, for the overpasses, such as overpasses of West Third Ring Road and East Third Ring Road, it is much difficult to judge and rebuild the terrain surface of the overpass area, which reduces the efficiency.

Thirdly, another problem is lack of elevation information in the experiments. There are not enough elevation points on some highway, the lowest elevation points under the bridge or elevation points along the hydrologic lines. The field work is still necessary for such situation. In this experiment, for the Beijing-Hong Kong-Macau Expressway, the field profile map is used to build the terrain. So information fusion is helpful to build the complex urban terrain in the future.

References

1. CJJ/T 8-2011: Code for Urban Survey. Ministry of Housing and Urban-Rural Development of the People's Republic of China (MOHORD) (2011)
2. Zhu, X.K., Li, M.: Method Research and Application of the Improved Work Flow of 1:10000 DEM Production. Beijing Surveying and Mapping 3, 1–14, 18 (2007)
3. Liu, X.S., Liu, C.L.: Build and Application of Digital Elevation Model in Arc/Info. Railway Investigation and Surveying 4, 8–10 (2004)
4. ESRI: Exploring_ArcObjects (2002)
5. Zhu, X.K., Yue, J.X., Liu, X.L.: Improvement of Slant Distance Measurement Method based on Detailing Surface Model. Journal of Geomatics 35, 3–5 (2010)
6. YBJ 05-2010: Beijing Digital Elevation Model Checking and Acceptance. Beijing Institute of Surveying and Mapping (BISM) (2010)

Analysis and Comparison between Digital and Smart Water Conservancy

Jinxin Wang[1], Liumin Zhang[2], Rui Hou[2], and ChengCai Zhang[1]

[1] School of Water Conservancy & Environment of Zhengzhou University, Zhengzhou 450001, China
[2] Engineering Institute of Surveying and Mapping of Henan Province, Zhengzhou 450003, China
{Jinxin Wang,Liumin Zhang,Rui Hou, Chengcai Zhang,jxwang}@zzu.edu.cn

Abstract. The digital water conservancy and smart water conservancy are the products of the contemporary mainstream IT applications in the water industry, and they are the different stages of development of water conservancy IT. This paper firstly discusses the concept, generation and development of digital water conservancy, and it's what has been achieved.Then it describes the background, concept and features of the smart water conservancy, and also the relationship between digital and smart water conservancy. Next the significance of digital across to the smart water conservancy is discussed in the paper. And finally a cool and rational thinking is put forward about the development of hydrological information technology. It conclude that construction smart water conservancy must be based on the laws of the technological development, combined with the actual situation of the regional socio-economic and capabilities.

Keywords: digital water conservancy, smart water conservancy, digital Earth, smart Earth, water conservancy informatization.

1 Introduction

Water is one of the most critical factors to life-sustaining. Human-water game has always been the theme of mankind to survive and reproduction in peacetime! On the one hand, it can be said that the history of human development is a struggle process of mankind chasing, using, resisting, and controling wate. On the other hand, from human water conservation history, we can easyly find that the level of human ideas and techniques of water-control and the current comprehensive national strength is directly proportional. In general, when a dynasty is in the period of wealthy and strong, which water-control level is high, and its achievements more prominent. And vice versa. Thus, the level of water conservancy development is a comprehensive reflection of socio-economic and other relevant factors of country in a corresponding period, especially the level of its science ane technology. Since the last century, with the digital earth strategy proposed, the concept of digital water covervancy established in people's minds; Just over a decade, the new thinking of smart water

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 424–434, 2013.

conservancy based on smart Earth techology walked into people's vision. From the digital Earth to the smart Earth is not just the conceptual speculation, it first is the inevitable result of technological development. Similarly, from digital to smart water convervancy, is not only the technological advances, more importantly, is water culture beyond. Correct understanding the of the relations between the digital and smart water conservancy will help us through the dazzling appearance, to rationally and scientifically grasp the inherent law of development of water conservancy science and technology, to avoid the confusion of overwhelmed appearance, and to promote water conservacy informatization in the right direction development. This paper discusses the concept and development of digital and smart water conservancy, the relations between the two consepts, revealing the technologies and cultural connotations from the digital to smart water conversavancy development.

2 Digital Water Conservancy and Its Achievements

Human water industry has gone through three stages: Ancient water conservancy project (approximately several thousand years, mainly in the agricultural society), modern water conservancy project (also known as traditional water conservancy project, about 300 years, mainly in the industrial society) and current hydraulic engineering (after World War II) [1]. At every ages, the hydraulic level and the level of prevailing science, technology and economic are closely related. In the 1990s, with the development of space, information and network communication technologies, National Information Infrastructure (NII, refers to the Information Superhighway) and National Spatial Data Infrastructure (NSDI, refers to the national spatial data and the various political, economic, cultural, technological and social information/data which is all regarded NSDI as their reference framework, that is, the goods on the Information Superhighway) strategies have been proposed, so from the overall integration, management and applications of global spatial information has become possible. In 1998, the concept of Digital Earth was made by the then U.S. Vice President Al Gore , and quickly evolved into a global strategy. Technically, Digital Earth is a massive spatial data which can be embedded, multi-resolution, and three-dimensional virtual representation of the Earth. It organizes the world's information which is scattered around the world or/and obtained from different channels by their geographical coordinates. Thus it not only reflects the inherent organic links among the variety of natural, cultural and social information, but also easily retrieves and uses those information by index of geographic coordinates. Digital Earth is the informatization of the Earth which includes the entire process of digitization, networking, intelligentizing, and visualization of all information on Earth. Its core idea is to use digital means to holistically solve the problems of Earth and to maximize the use of information resources [2].

Since the Digital Earth strategy was put forward, it quickly penetrates into a variety of industries of human society, more and more new terms beginning wih "digital" have been springing up rapidly. They are mainly divided into two categories: a class is the digital area, such as digital China, digital Henan, digital city, etc.; and the other

is the digital industries, such as digital water conservancy, digital land, digital transport, etc.. Digital water conservancy is one of many digital industries.

Briefly, digital water conservancy is a modern industy which uses the ideas and techniques of digital earth to construct and renovate traditional water industry, to achieve digitization, networking, intelligentizing, and visualization of various water conservancy projects of planning, construction and management. After more than ten years of construction and development, It has made the following major achievements(in China).

2.1 Digital Water Conservancy Has Achieved Digitization and Integration of Various Water Information

Digitization is the most basic requirements of Digital Earth. The geographical entities, events and processes in objective world all occur in Earth's surface and its vicinity geographical space. The geographical space belongs to the physical space, and can not go directly into the computer. Therefore, to build digital earth, first make geospatial information and other relevant information digitization, which is changing the information into encoded binary strings according to certain rules in order to be processed by computer, and then they are processed , modeled, and built the virtual geography space.

The traditional water related information and data, mostly based graphics, images, documents, files, audio and other forms categories stored in paper or other carrier. Use and analysis of these data is very tedious and troublesome. Digital water system makes all of the above information digitization, then uses the database to store and manage them. It firstly enables the basic spatial data, attribute data and thematic water data seamless integration based on their geographic coordinates, so the processing, modeling, comprehensive analysis and applications of water information become very convenient, fast and effective.

2.2 Digital Water Conservancy Has Built a Highly Efficient Technology Platform for Hydraulic Engineering Planning, Construction, Management and Application

Key technologies of digital water conservancy include 3S (GNSS - Global Navigation Satellite System, RS - Remote Sensing, and GIS - Geographical Information System), network communication, database, and mathematical modeling techniques, etc. [3], and it is a digital engineering [4] architecture of multi-technology platform seamlessly integration. In which, 3S technology is the core supporting technology [5]. Digital water conservancy overall technical architecture shown in Figure 1 [5].

From the figure we can see that the overall technical architecture of digital water conservancy is a seamlessly integrated, organic and unified large platform which is consist of a series of platforms including data acquisition and updates (system), data storage and management, information extraction and analysis, network and standard platforms. This great platform for all types of users (including all types of agencies, units and individuals) to achieve business process and logic application by means of digital water conservancy provides a convenient and efficient technology environment (Figure 2 [5]).

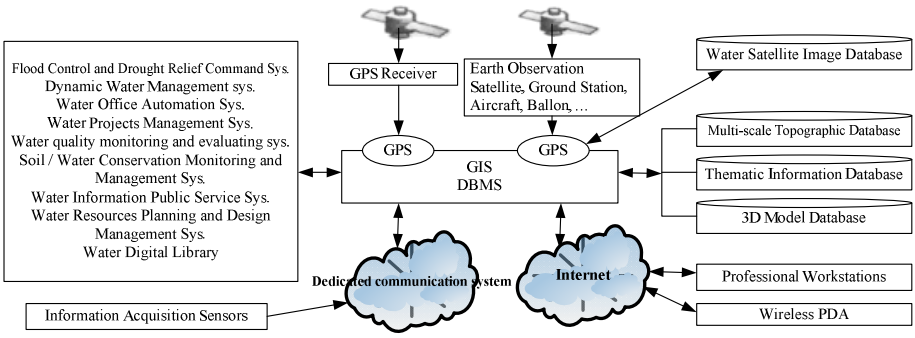


Fig. 1. Digital water conservancy overall technical architecture

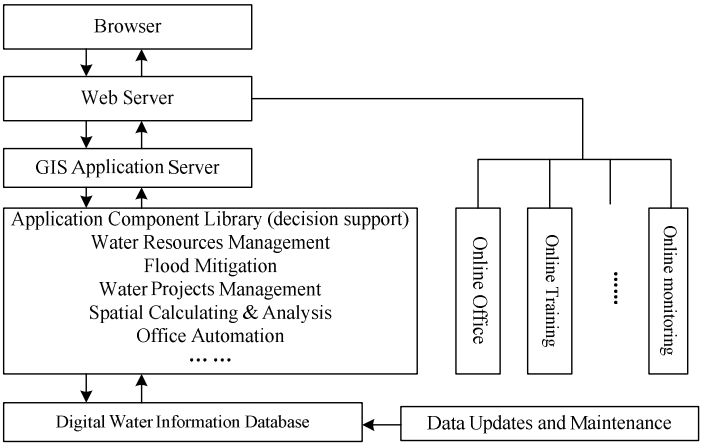


Fig. 2. The running mode of digital conervancy

2.3 Digital Water Conservancy Has Promoted the Great Development of Water Conservancy Cause

As the novelty of the information age, digital water conservancy greatly promoted comprehensive flourish of the theory, technology and application of contemporary water science with its strong power. It is at the background of digital water conservancy that a lot of new ideas and policy framework like digital valley [6], digital water resources[7], digital reservoirs [8], modern water conservancy [9], macro water conservancy [10] have been proposed, and they all deeply imprint the name of the times!

3S technology is the “drive engine” of the digital water conservancy “machine”. GNSS technology provides the basic geospatial reference framework for virtual space of digital water conservancy, and itself is also a remote sensing system - a satellite signal through the atmosphere, carrying a lot of temperature, humidity, particle size, density and other relevant information to the atmosphere. RS technology provides the

case of ground objects and landform, distribution of water resources, land cover, the ground temperature and humidity, water quality, sediment, flood distribution, water pollution and other natural environmental conditions scales from the entire planet to a small watershed by its high-angle, wide vision, multi-temporal, and multi-resolution holographic remote sensing images, and then it provides a full range of information for the study of surface hydrological processes, hydrology and water resources model, the planning, construction and management of various water projects, disaster prevention and mitigation logistics,etc.. If we say the RS and GNSS is two eyes of a mankind in the information age, then GIS is the his brain. Variety of geospatial, attribute and thematic information is processed and modeled in the GIS, then a virtual geographical space is built for simulation and deducing the real world by people. In detail, GIS can provide a three-dimensional virtual water space for people to build, simulate, and deduce various hydrological processes and water conservancy projects, to mine and and expand the variety of models of hydrology and water resources, to screen and optimize water conservancy construction program, to evaluate quality, ecological effects and economic benefits of water conservancy construction, to promote water conservancy science knowledge, to advocacy landscape tourism, and to provide a convenient and paperless office environment for water practitioners. Literature [5] and [11] have systematically summarized the results of the spatial information technology applied in China water industry in recent years.

2.4 Digital Water Conservancy Has Promoted Exchange and Sharing of Water Information Resources

Information/data is the basis for digital earth, and information sharing is the basic purpose of construction of digital earth. One outcome of Digital Earth is to achieve a comprehensive global basic geographic information sharing, shch as Google Earth system to achieve a global multi-resolution basic geographic information sharing, and China TIANDITU to achieve a bacic geographic information sharing of digital China. Nowadays, people can easily access the Internet to obtain the non-classified data they need.

The information sharing based on digital conservancy is reflected in the following aspects: 1) It transforms the traditional paper-based material into digital information, builds large-scale water conservancy information database for sharing, then firstly a lateral sharing can be achieved in a enterprise; 2) Based on Internet and WebGIS technology, secondly a longitudinal sharing can be achieved in entire national water conservancy system, such as from the National Ministry of Water Resources to the city or county water departments to share information; 3) Through research and project collaboration, thirdly a extensive information sharing can be achieved based on water conservancy between wr units and other water-related departments, shch as sharing among geology, meteorology, environment, land, agriculture and forestry sectors, and cross-regional large-scale water conservancy projects coordination and information sharing.

Information sharing is the theme of the information age. Geographic information and thematic information sharing not only avoids repetitive production, duplication,

greatly reduce production costs, but also promotes regional cooperation and sectoral cooperation, cooperation in scientific research, production cooperation, etc.. In summary, It can improve IT productivity, and promote rapid social and economic development.

2.5 Digital Water Conservancy Has Created a Digital Water Cultural Environment

A certain period of human culture is always suited to the level of prevailing social productive forces. As bronzes produced a farming culture, iron promote the development of industrial culture, the silicon (information) technology gave birth to the digital culture [12]. Similarly, digital water conservancy gave birth to the modern digital water culture.

In the concept of digital water, in addition to the construction of information systems, various other factors which are closely related to the information system, including management, application, effect, influence and infiltration, are called digital water cultural environment [13]. It is mainly composed of six elements as follows: information technology, policy system, spiritual culture, technology updates, digital lifestyle and digital work, and is also the the environment adapted to the digital water conservancy body (information systems) . The essence of the digital water culture is the pursuit the harmony amond practitioners, water information systems and water conservancy businesses [13].

3 Proposed of Smart Water Conservancy Concept

3.1 Understand the Concept of Smart Earth

The development of information technology is faster and faster. In 1999, next year of Digital Earth was made, the concept Internet of Things of occured; 2000, IPv6 protocol basically established; 2006, cloud computing was first proposed. So far, the key technology of auto-sensing, transmission and processing mass information has become mature [14], and the intelligent ideas are ready. November 6, 2008, Samuel J. Palmisano, the President and CEO of IBM, first proposed the concept of a smarter Earth / planet in a speech in New York, and in January 28, 2009, he explained “Smart Earth” to new President Barack Obama and proposed new government investment in a new generation of intelligent infrastructure [2,15]. Subsequently, like Digital Earth, the Smart Earth rapidly evolved into a global strategy.

Leading the wave of information technology of the beginning of the new century, Digital Earth has made great achievements [2], and build the overall framework of global information. However, due to the inertia of traditional techniques, a variety of natural social conditions, uneven economic development, people's work habits and other reasons, the Dugital Earth is still relatively weak in ubiquitous, automatic, accurate, intelligent, etc. For example, it is not a good solution to the current power grid inefficiencies, parking problems in heavy traffic, commercial weak link in the supply chain, some of the financial market crisis and other common things that the

people currently has done not good enough in quality, efficiency, precision and so on [14]. On the basis of Digital Earth, the Smart Earth will make an effective solution to those problems by using the most advanced information technologies. The main content of the Smart Earth is to make full use of the new generation of IT technology in all walks of life, that is, the electronic chip and sensors are mounted to the power grid, railways, bridges, tunnels, roads, buildings, water systems, dams, oil and gas pipelines, and even the various objects in daily life, linked up to form "Internet of Things", and integrated by supercomputers and cloud computing to realize the online integration of Digital Earth, human society and the physical system, so that humanity can be more refined and dynamic management of production and life, to achieve the "smart" survival state, to improve resource utilization and productivity levels, and thus to amend the relationship between man and nature [2,14]. Based on Digital Earth, Internet of Things and intelligent computing, Smart Earth puts the three elements (or ability)of perception, transmission and processing into the entire geographical space, make this space has humanlike self-perception, transmission and process information, make appropriate judgments and decisions, feedback to the corresponding receptors, and make it act accordingly. That is, the Smart Earth gives geospatial environment "humanlike intelligence." Their intellectual level depends on the accuracy of perception, transmission reliability, the richness of the knowledge, the speed of operation, and the contingency of processing method [14]. Among them, the perception of accuracy depends on the sensor, the transmission reliability depends on the communications technology, knowledge comes from the mining on Digital Earth's knowledge base, model base and sensor information data warehouse, processing speed depends on the network, grid and intelligent computing, processing methods' contingency depends on the cloud computing, knowledge mining and natural computing (neural networks, evolutionary computation, genetic algorithm, ant colony algorithm .....) and other intelligent computing.

Digital Earth concept includes the intelligentizing, therefore, precisely, Smart Earth is not a new concept and technology system. It belongs to the category of Digital Earth, is a "post-Digital Earth" era and the inevitable result of the development of Digital Earth.

3.2 The Concept of Smart Water Conservancy

After having introduced the concept of a Smarter Earth, instead of "digital", "smart" has become an omnipotent adjective. Smart city, smart grid, smart transportation, smart medical treatment, smart water management, smart finance …, a lot of new concepts springs out.

Smart water conservancy is the result of ideas and techniques of the Smart Earth used in water conservancy industry. It uses the Internet of Things technology, ubiquitous, automatic, real-time sensing of the data of the key elements, points, positions and link of water resources, environment, processes, and projects; By the information communication network, sents them to the online databases, data warehouses, and cloud storage; In the virtual space, it uses the technology of cloud computing, knowledge mining, natural computing, and other intelligent computing for data processing, modeling and deducting, to make scientific judgments, decision

optimization, and feedback to people or equipment to respond correctly. So it effectively solves various problems of water science, technology and industry, improves water utilization, benefit of water projects as well as practitioners working efficiency, effectively protects water resources and environment, makes disaster prevention and mitigation, and achieves harmony between man and water.

Digital water conservancy is the basis of Smart ones. In Figure 1, integrate the bottom of sensor networks and Internet to the Internet of Things, put the intelligent computing to GIS platform, and then the overall technical framework of smart water conservancy is built. Thus, the Smart Water Conservancy = the Digital Water Conservancy + Internet of Things + intelligent computing. Rather, the Smart Water Conservancy is a kind of flat structure based on Internet of Things (or the ubiquitous network), its logical hierarchy shown in Figure 3. It is a advanced stage of Digital Water Conservanct development.

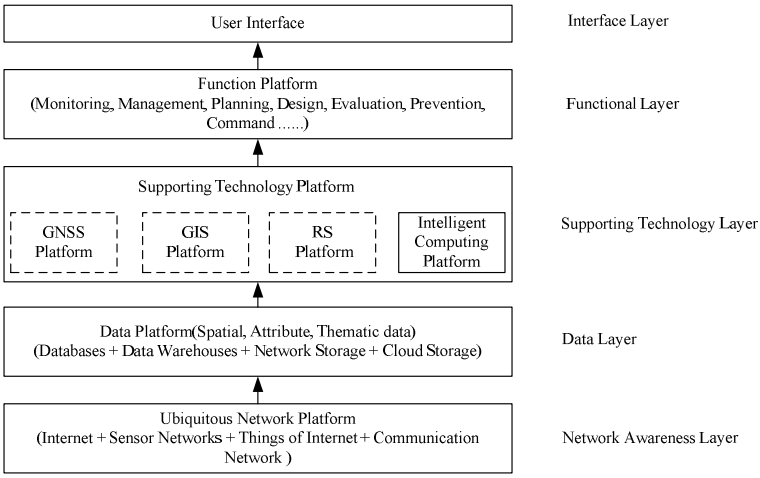


Fig. 3. Smart Water Conservancy Logical Hierarchy (dotted box are digital water conservancy supporting technology platforms)

4 Leaps from Digital to Smart Water Conservance

According to above analysis, from the digital to smart water conservance not only means technology, but also a lot of ideas and elements across and expansion:

4.1 From Static to Dynamic

Most digital water information is quasi-static. Its collection and maintenance updates mainly rely on manual operation, and their dynamic updates are not timely. However, smart water conservancy realized person - objects, things - people, objects - objects fully organic connection, and the perception and transmission of information is carried out automatically. The original problems caused by static management in digital water

conservance can now be resolved by sensor network including remote sensing, because the sensor network can achieve automatic and real-time information updates.

4.2 From 3D to Multidimension

Because the information in digital water systems is quasi-static, so a virtual digital water space built by them is also a quasi-static three-dimensional space; But the smart water information is automatically updated over time, and therefore a smart virtual space is a four-dimensional dynamic space of spatial 3D plus time. From another perspective, constituted by the water sensor information data warehouse itself is a subject-oriented, integrated, collection of data to support management decisions. By multidimensional analysis, synthesis and mining the data in the data warehouse, you can get a lot of implicit, previously unknown, potentially useful, and ultimately understandable water expertise and rules.

4.3 From Product to Service

On the current situation, the digital water conservancy also generally expressed as the relatively independent information systems. These systems have corresponding functions, to provide the specific product. The connection between each other, information sharing and interoperability capabilities are weak, basically in the island state. And in theory, smart water conservancy is a kind of network services (middleware) based on web, grid, cloud and other distributed computing, no longer showed a large information systems, and essentially is a flat architecture based on Internet of Things (or the ubiquitous network). Smart water conservancy prospects: blending tangible in the invisible, its services can be get everywhere and everytime!

4.4 From Correct and Scientific to Precise and Intelligent

What the digital water conservancy pursued is the correctness of the data results, the scientific of the project implementation. However, due to data limitations of data, technologies and ideas, this is basically a correct and scientific trends in broad outline, and is difficult to achieve accurate and intelligent high standards. While depending on the latest development of IT technology, the smart water conservancy can be achieved active and comprehensive perception, automatic and reliable delivery, sensitive and intelligent processing, and fast and precise response. It's results are not only scientific and correct and also accurate and intelligent; Differential control, make mankind achieve wisdom existence.

4.5 From Digital to Smart Water Culture

As mentioned earlier, digital water culture, in pursuit of practitioners, water resource management and water information systems harmony, can be summarized as water industry harmony. But the aim of Smart Earth is to treat, save and protect the earth,

and make the human home to get better [14]. Therefore, smart water coservancy should be modern water conservancy [9] or macro water conservancy [10], while smart water culture should be modern macro water conservancy culture. It purses the harmony between man and the natural water environment that also be called human-water harmony!

5 Conclusions

A prominent feature of modern information technology development is the combination of a country or region's technological development and its political and economic strategies at national level tightly! From the Information Infrastructure to the National Spatial Data Infrastructure, from the Digital Earth to the Smart Earth, those new concepts are not made by scientists, but by the heads of government or industry leaders raised. Thus, these concepts are not purely scientific concept, but implies a profound geopolitical, economic connotations. Faced with these exotic fascinated new terms, we should remain calm and rational! We should understand the following:1) From the digital to smart water conservancy, that is the inevitable result of contemporary social, economic and technological development, and it is not the people's will; 2) From the digital to smart water conservancy, that is not only the concept update and technology development, but also is the ideas progress and cultural leap; 3) From the digital to smart water conservancy, that is a gradual building process of accumulation. Smart water conservancy should be built on advanced and complete digital water infrastructure, then any conservative and aggressive practices are wrong. Scientific approach should calmly analyze the situation of contemporary technological developments, combine with the region's socio-economic reality, follow the internal laws of scientific and technological development, mainly rely on their own strength, and track the trend of technological development. Smart Water Conservancy is both an opportunity and a challenge!

Acknowledgments. This research is supported by Henan Innovation Talents Program (104100510003).

References

1. Xue, W., Rong, Y., Hu, X.: Progress of Recent Urban Water Conservancy Project. In: Tang, H., Li, G., Wang, L. (eds.) Progress of Hydraulics and Water Informatics 2007, pp. 128–132. Hehai University Press, Nanjing (2007)
2. Li, D., Gong, J., Shao, Z.: From Digital Earth to Smart Earth. Geomatics and Information Science of Wuhan University 35(2), 127–132 (2010)
3. Chen, Y., Wang, X.: Digital Water Conservancy and Chinese Water Conservancy Modernization. Perial River (1), 51–54 (2009)
4. Bian, F.: Digital Engineering Principles and Methods, 2nd edn. Surveying and Mapping Press, Beijing (2011)

5. Wang, R., Chen, B., Yang, Y., Zheng, W.: Application of 3S technology to digitalwater conservancy. Science of Surveying and Mapping Science 33(3), 210–212 (2008)
6. Lu, D., Li, R., Duan, S., Liu, J.: Digital Valley Architecture and Key Technologies. Soil and Water Conservation in China (5), 59–61 (2009)
7. Yan, J., Wen, H.: Preliminary Conception of Guangxi Digital Water Resources. GX Water Resources & Hydropower Engineering 1(1), 18–20+35 (2006)
8. Liang, R., Zhao, A., Zhao, M., Wang, C., Lin, X.: Studies on Digital Reservoir. Journal of Shandong Agricultural University (Natural Science) 36(2), 313–316 (2005)
9. Wang, L.: Modern Water Conservancy Prospects. Perial River (3), 1–6 (2002)
10. Liu, S.: Theory of Macro-water Resources Development and Scientific Outlook. Water Resources and Hydropower Engineering 40(8), 21–24 (2009)
11. Tan, D.: Application Research and Practice of Spatial Information Technology in Water Conservancy. Yangtze River 43(8), 1–6+17 (2012)
12. Chen, S.: Hundreds of problems of Digital Earth. Science Press, Beijing (2006)
13. Wang, Z.: Building Digital Water Cultural Environment: Harmony of People, Digit, and Business. China Water Resources News (March 8, 2003)
14. Lei, Z.: Interpretation Smarter Earth. Telecommunication Network Technology (1), 38–40 (2010)
15. Cai, D.: Smarter Earth, Internet of Things, Cloud Computing and Other. Secretarial Work (8), 47–48 (2010)

Evaluation of the Urban Land Intensive Use and It's Regional Differences in Shaanxi Province Based on GIS

Minning Zhao[1,2], Qingyun Li[3], Jianmin Feng[2], Lingxia Chen[2], and Huiru Li[1]

[1] Chang'an University Geology and Surveying College, Xi'an, China, 710054
[2] Department of Tourism, Resource and Environment Xian yang Normal College, 712000
[3] 61363 Unit of the Chinese People's Liberation Army, Xi'an, China, 710054
382825670@qq.com

Abstract. With the accelerating process of urbanization in China, the demand for urban land is increasing rapidly; the finiteness of land resources determines that urban land use must take the road of intensive utilization. This paper selected 10 cities in Shaanxi Province as the research object, by establishing an urban land intensive use evaluation index system, which included four aspects of the land's economic benefits, the investment intensity, the utilization intensity and the degree of sustainable, by the principal component analysis method and the statistical software SPSS, it studied the status of the urban land intensive use and spatial differences of 10 cities in Shaanxi province by the software SUPERMAP. Results show that the degree of the land intensive use of 10 cities in Shaanxi province is quite different, Guanzhong area shows higher, the northern Shaanxi is the second and the Southern Shaanxi is the lowest; the urban land intensive use degree in Shaanxi province is correlated with the level of economic development and location. Base on these, policy implications are derived that the city government should increase the land investment intensity, improve the economic benefits of land and land use intensity, speed up the economic development, improve the level of the urban land intensive use.

Keywords: urban land, land intensive use, regional differences, Shaanxi Province.

1 Introduction

Land resource should be used economically and intensively because of it's finiteness and the scarcity, and it is also the fundamental starting point of land use. In recent years, with the rapid development of the urbanization process in China, how to realize economy and society develop rapidly has become an urgent problem to be solved. Therefore, experts and scholars in the field of the urban land intensive use have carried out a series of studies, and made a comprehensive survey on some relevant researches, it can be seen that research methods such as the factor analysis, principal component analysis, entropy method, analytic hierarchy process method, multiple factors comprehensive method are used in the related research mostly [1-4]. In this

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 435–444, 2013.

paper, an evaluation index system of land intensive use was established, and Principal Component Analysis method was used, the status of the land intensive use of 10 cities in Shaanxi Province was researched, and map of land intensive use in Shaanxi Province was drawn by Supermap, the regional differences in Shaanxi Province on the land intensive were evaluated.

2 The General Situation of Research Area

Shaanxi province, is known as "Qin"or "Shan", it is located in east longitude 105 29′ ~111 15′ (E),latitude 31°42'- 39°35' (N),the land area is of 205600 square kilometers, the resident population is 37430000 by the end of 2011, 10 provincial jurisdiction City and 1 Agricultural High-tech Industries Demonstration Zone are set, the terrain in Shaanxi Province is complex, it is long in the north and south direction and narrow in the west and east direction, the Loess Plateau is located in the north of Shaanxi Province, The Guanzhong Plain is in the central known as Eight Hundred Li Qinchuan, the southern Qinba Mountains is located in the south of Shaanxi Province, Shaanxi province has a profound historical and cultural heritage, and it is also the source of Chinese culture.[5]

3 Research Technical Route and the Construction of Evaluation Model

3.1 Research Technical Route

This study evaluated the degree of the urban land intensive use in Shaanxi province based on the relevant research data, established an urban land intensive use evaluation index system, used principal component analysis method. Then the factor scores of 10 cities were calculated, finally, land intensive spatial map in Shaanxi province was drawn by the software Supermap, The research technical route is shown below in figure 1.

3.2 Construction of Evaluation Model

3.2.1 The Establishment of Evaluation Index System

Some principles such as comprehensive principle, leading principle and comparability principle should be followed in establishing land intensive use evaluation index system [6-7]. Guided by the principles above, this paper constructed a land intensive use evaluation index system base on analysis of the influence factors such as the natural environment factors, social factors, economic development and policy factors as well as the guidance of the experts and the actual of the Shaanxi province, the evaluation index system contains the following factors: land investment intensity, land use intensity, economic effect, social benefit, ecological benefits of land productivity, sustainability of land use. The evaluation index system established is shown in Table 1[8-12].

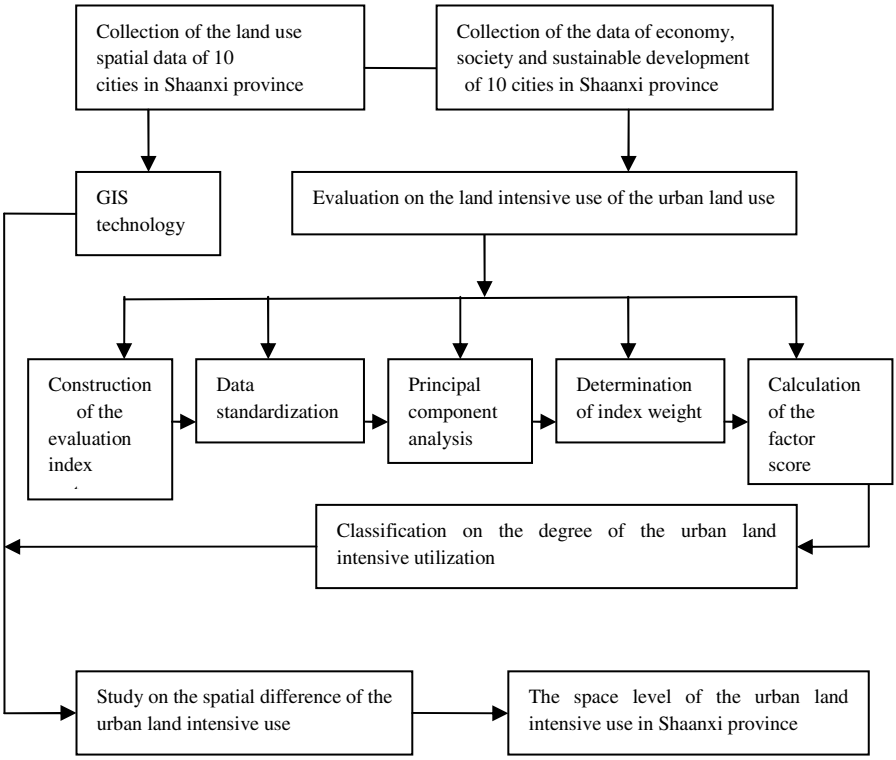


Fig. 1. Map of the Research Technical Route on Land Intensive Use

3.2.2 The Establishment of Evaluation Model

3.2.2.1 Data Standardization

In this study, 9 indicators were different in units and dimension, in order to make the data be comparable, data standardization was needed, through the range standardization of source data by using the SPSS16.0 software, the formula is shown below: [13]

$$\lambda ij = \frac{Xij - \overline{Xj}}{\sigma j} \tag{1}$$

λij is the value after standardization, σj is the standard deviation, $\overline{Xj}$ is the arithmetic mean value of the j index.

Table 1. Urban Land Intensive Use Evaluation Indexes System

Evaluation target	The first grade indicators	The Second grade indicators	Meaning of the indicators
The degree of the urban land intensive utilization In Shaanxi Province	Economic benefit of urban land	The per unit area GDP X1	The urban GDP/ Built-up area
		The per unit area total retail sales of social consumer goods X2	Total retail sales of social consumer goods/ Built-up area
		The per unit area tertiary industry output value X3	The tertiary industry output value/ Built-up area
	Investment intensity of urban land	The per unit area Investment in fixed assets X4	Investment in fixed assets / Built-up area
		The per unit area labor input X5	The labor input/ Built-up area
	Use intensity of urban land	The per capita living area X6	The total area of residential city/urban population
		The per capita Construction land area X7	Construction land area/urban population
	The degree of sustainable utilization	The per capita green area X8	Landscaping area/urban population
		The per capita road area X9	Road area/urban population

3.2.2.2 Principal Component Analysis

This study set 10 cities in Shaanxi province as research object, constructed variable matrix of 9 indicators shown in the Table 1, processed the data with the software SPSS16.0, and calculated the variance contribution of eigenvalues of the matrix and the rate of principal components, the factor regression coefficient were calculated based on components selected, finally, factor scores of 10 cities were calculate according to the calculation formula shown as follows:

$$Nik = \sum_{j=i}^{n} Wj\lambda ij \tag{2}$$

Nik is the factor scores of principal component k of the city I , Wj is index regression coefficient of the indicator j.

. Each city's factor score Qi was calculated based on the calculation of each factor scores, and variance contribution of each component which was used as the weight of each factor. The calculation formula is shown below.

$$Qi = NikBk \tag{3}$$

Qi is the score of the principal components of the city i , Bk is the variance contribution rate of each principal component.

In order to make the numerical value of urban land intensive utilization degree was intuitive ,percentage conversion was done on the comprehensive score of each city in formula(3), the conversion formula is shown below.[14]

$$Ti = \frac{Qi}{Q\max - Q\min} \times 40 + 60 \tag{4}$$

Ti is the percentile score of the city i, Qmax is the maximum value of the principal component scores. Qmin is the minimum value of the principal component scores.

3.2.2.3 The Classification of Grades on Urban Land Intensive use in Shaanxi Province

Grades of the urban land intensive use in Shaanxi province were classified with reference to the related studies on land intensive use, and it also referred to social and economic conditions of Shaanxi Province as well as the experts' opinions. It is shown in Table 2.

Table 2. Grades of the Urban Land Intensive Use in Shaanxi Province

Grades	The first level	The second level	The third level	The forth level
score	>75	75~65	65~50	<50
meaning	Highly intensive use	Moderate intensive use	Low level intensive use	Extensive use

4 Evaluation of the Urban Land Intensive Use in Shaanxi Province

In this study,10 cities in Shaanxi province are taken as research object ,which are Xi'an, Tong chuan, Baoji, Xian yang, Weinan, Yanan, Hanzhong, Yulin, Ankang, Shangluo, 9 indexes of 10 city formed the original data set (data source: Statistical Yearbook 2011, Shaanxi province, China City Statistical Yearbook 2011), the calculated Eigen values and the corresponding variance contribution rate are shown in table 3.

Table 3. Total Variance Explained

Serial number	Eigen values	Variance contribution rate(%)	Accumulated variance contribution rate(%)
1	3.669	40.766	40.766
2	2.494	27.714	68.479
3	1.164	12.928	81.407
4	1.013	11.251	92.659
5	0.298	3.313	95.972
6	0.219	2.429	98.401
7	0.126	1.402	99.803
8	0.017	0.190	99.994
9	0.001	0.006	100.000

Table 4. Component Matrix and Component Score Coefficient Matrix

Indicators	Factor extraction results			
	The first principal component	The second principal component	The third principal component	The fourth principal component
X1	0.017	0.969	-0.072	-0.102
X2	0.685	0.441	-0.366	0.299
X3	0.935	0.077	-0.209	-0.019
X4	0.837	-0.106	-0.292	-0.255
X5	0.010	0.984	-0.130	-0.030
X6	-0.343	0.069	0.890	0.111
X7	-0.211	-0.383	0.863	0.085
X8	-0.622	0.493	-0.513	0.032
X9	-0.081	-0.107	0.114	0.964
Indicators	Factor regression coefficient			
	The first principal component	The second principal component	The third principal component	The fourth principal component
X1	0.049	0.443	0.182	-0.090
X2	0.272	0.168	-0.030	0.332
X3	0.386	0.047	0.077	0.033
X4	0.305	-0.062	-0.018	-0.179
X5	0.037	0.436	0.138	-0.019
X6	0.040	0.210	0.528	0.005
X7	0.065	0.000	0.439	-0.011
X8	-0.361	0.079	-0.368	0.057
X9	0.012	-0.034	-0.058	0.876

According to the principles that the eigenvalues is more than 1, the first 4 eigenvalues in Table 4 were selected as the main components, the accumulative variance contribution rate is 92.659%, which means that the 4 new variables contain 92.659% information of the original variables. The selected principal components were orthogonal varimax rotation .it is shown the following table 4.

The result of the Table 4 showed that the first principal component had more load coefficient in the indicators as retail sales of social consumer goods ,the tertiary industry output value, the per unit area tertiary industry output value, the per capita green area, which reflected the input and output level of the land.

The second principal components had more load coefficient in the indicators as GDP and the per unit area labor input, which reflected the labor input and economic level of the land.

The third main component had more load coefficient in the indicators as the per capita living area and the per capita construction land area which reflected the land use intensity.

The forth principal component had more load coefficient in the indicator per capita road area, which reflected the sustainability of the land.

Scores of each principal component were calculated according to the regression coefficient in Table 4, and they multiplied the variance of each principal component contribution rate as weight, the weighted sum factor scores of 10 cities in Shaanxi province were calculated as follows.

Table 5. The Principal Component Factor Score and Comprehensive Score of Each City

City	Q1	Q2	Q3	Q4	Combined score	Percentile score	Ranking
Xi'an	0.7452	0.8799	1.0356	1.1218	0.8077	82.4659	1
Yulin	1.5330	-0.5768	1.0536	-0.0557	0.5950	76.5506	2
Weinan	0.7836	0.3011	-0.7333	1.3949	0.4650	72.9341	3
Xianyang	-0.4459	2.0594	-0.0910	-0.6857	0.3001	68.3459	4
Yan'an	1.0341	-0.7552	-0.5081	-1.7861	-0.0544	58.4872	5
Hanzhong	-0.2474	0.8817	-0.8513	-0.9141	-0.0694	58.0705	6
Ankang	-0.1073	-0.8213	-0.2537	0.4147	-0.2575	52.8376	7
Tongchuan	-0.5518	-1.2761	-0.6706	1.7538	-0.6442	44.6510	8
Baoji	-1.4941	-0.2993	0.0584	0.7125	-0.6043	43.1916	9
Shangluo	-0.5251	-0.9988	-1.4640	0.4418	-0.6304	42.4660	10

According to the results of Table 2 and Table 5, the urban land intensive use degree in Shaanxi province can be divided into four categories: (1) highly intensive use, Xi'an

and Yulin (2) moderate intensive use, Weinan and Xianyang (3) low intensive use, Yan'an,Hanzhoung and Ankang (4) extensive use, Tongchuan, Baoji and Shangluo.The spatial difference map of the urban land intensive use in Shaanxi province was drawn by SUPERMAP, it is shown in figure 2

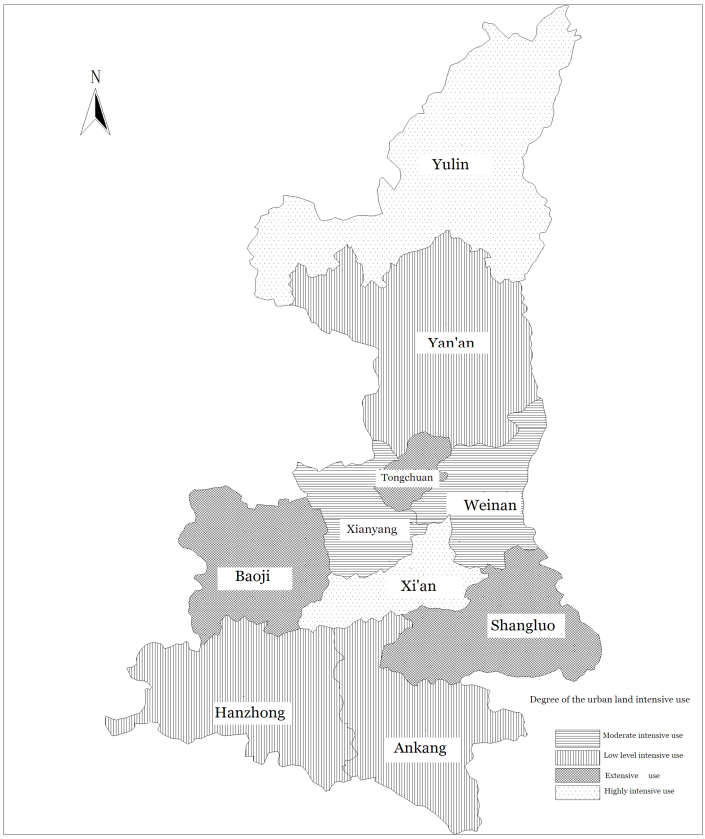


Fig. 2. The Spatial Ddifference of the Urban Land Intensive Use in Shaanxi Province (Note: For the sake of research convenience, the urban land is indicated with the city land)

5 Conclusions

Research shows that urban land intensive use in Shaanxi province has the following characteristics:

(1) The urban land intensive use among 10 cities in Shaanxi province is significantly different, the calculated percentage data display that the highest score is 82.4659(Xi'an City), while the lowest score is only 42.4660(Shangluo city), the highest value is almost 2 times of the minimum value.

(2) Seen from the score of each principal component factor, Xi'an, Yulin, Weinan and Yan'an have high scores in the land economic inputs and outputs principal component, while Baoji, Shangluo and Tongchuan have low scores in this item, and the weight of the component accounts for 40.766% of the overall weight, due to the impact of land economic inputs and outputs principal component, the former cities are high in the comprehensive score, while the latter cities are low in the comprehensive score. Xianyang, Hanzhong and Xi'an have high scores in the land of the labor input and economic level principal component, yet, Shangluo has low scores in the land investment, economic level, land use degree three principal components, therefore, the comprehensive score of Shangluo is the lowest .From the principal component scores of the land use degree Xi'an city and Yulin score the highest.

(3) It also shows that there is a certain relationship between the urban land intensive use degree in Shaanxi province and the level of economic development, The more intensive the urban land use is, the higher the economic development level is ,and vice versa.

(4) The degree of the urban land intensive use in Shaanxi Province shows the trend that Guanzhong area is the most intensive than Northern Shaanxi and Southern Shaanxi, which indicates that there is a certain relationship among the urban land intensive use and location of 3 regions in Shaanxi Province

(5) Several policy implications are derived after research that the city government should increase the land investment intensity, improve the economic benefits of land and land use intensity, speed up the economic development, improve the level of the urban land intensive use.

Acknowledgments. 1 Scientific Research Project of Education Department of Shaanxi Province (Program No. 12JK0137)

2 Shaanxi Provincial Key Disciplines Special Funds Construction Project --Key Supported Discipline "Historical Geography" (Program No. 10722，060103)

3 Scientific Research Program Funded by Xian yang Normal College (Program No. 09XSYK212)

4 The National College Students' Innovation and Entrepreneurship Training Project (201210722013)

References

1. Zhou, F., Chen, S., Zhong, L., Wu, X.: Evaluation of Urban Land Intensive Use Based on Factor Analysis- A Case Study of Qing Yuan City. Resources and Industry 5(11), 56–59 (2009)
2. Zhai, W., Huang, X., Zhang, Q., Zhong, T., Ma, Q.: Research on the Urban Development Zone Land Intensive Use Based on AHP—A Case Study of Jiangsu Province. Journal of Nanjing University 1(1), 78–84 (2006)
3. Luo, X., He, H., Ke, X.: The Urban Land Intensive Utilization Evaluation Based on Analytic Hierarchy Process- A Case Study of Zhongxiang city of Hubei province. Journal of Anhui Agricultural Sciences 37(36), 18114–18116 (2009)

4. Xu, S., Zhu, H., Zhou, S., Zeng, Y.: Research of Urban Land Intensive Utilization—A Case Study of Leiyang City. Hubei Agriculture Science 50(18), 3885–3888 (2011)
5. Introduction of the Shaanxi Province, http://baike.baidu.com/view/6510.htm
6. Yang, S.: Construction of Urban Land Intensive Utilization Evaluation Index System and Its Connotation. Exploration of Economic Problems (1), 29 (2007)
7. Cha, Z.: Construction of Urban Land Intensive Utilization Capability Evaluation Index System. Zhejiang Statistics (4), 9–11 (2002)
8. Li, M., Jiang, H.: Study of Intensive Land Use Theory and Policy. Chinese Science 22(2), 55–61 (2008)
9. Zhou, S.: Learning of land evaluation, vol. 283. Southeast University Press, Nanjing (2006)
10. Bi, B.: Land Economics. Renmin University of China Press, Beijing (1991)
11. Zhang, M., Zhou, J., Li, X.: Evaluation of Urban Land Intensive Use—A Case Study of Zhengzhou City. Anhui Agricultural Science 36(35), 15609–15611 (2008)
12. Luo, X., Zhou, Y., Nie, Y.: Evaluation of Wuhan Urban Land Intensive Use. Anhui Agricultural Science 35(31), 10040–10042 (2007)
13. Li, Y., Zhou, Y.: Principal Componment Analysis of the Spatial Differences of Urban Land Intensive Use in Jiangsu Province. Journal of Hunan Institute of Engineering 16(2), 25–27 (2006)
14. Li, Y., Zhou, Y., Wu, L., Jin, X., Wang, L., Hu, J.: Spatial Differences of Urban Land Intensive Use—A Case Study of Jiangsu Province. Journal of Nanjing University(Natural Science) 42(3), 310–315 (2006)

A Combinatory Framework of Geographic Information Services Integration Based on OWL-S

Peichao Guan[1,2], Guobin Zhu[3], and Xinwei Zhu[3]

[1] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China
guanpc@whu.edu.cn
[2] School of Information Management, Hubei University of EconomicsWuhan 430205, China
gpc@hbue.edu.cn
[3] International School of Software, Wuhan University, Wuhan 430079, China
{xinwei.zhu,gbzhu}@whu.edu.cn

Abstract. Along with the increasing advancement of geographic information services application, the requirement for semantic manipulation during the integration of different geographic information services has been also increasing under the Heterogeneous environments. Through studying current geographic information services applications, and analyzing the shortcomings of the implementations, this paper proposes a combinatory framework of Geographic Information Services Integration(GISI) with Ontology Web Language for Services(OWL-S). The framework is based on existing experiences on service ontology applications and relevant techniques. It provides an implementation process of semantics injection with geographic information service for the specific core framework module.

Keywords: Geographic Information, Service Integration, Ontology, OWL-S.

1 Introduction

The rapid development of internet and computer technologies in the last two decades has greatly improve the information sharing and relevant developments of a variety of basic information on internet. With the increasing demand for the information sharing in city management, environment government, and public safety related geographic information, geographic information services platform development has been in a track of rapid improving with the support from functional departments at all levels. The key problem in geographic information services platform is how to use advanced information technology from the characteristics of geographic information improving existing implementation framework and approach, along with effective integration. In the current mature technology environment, the integration of geographic information services are mainly based on Web services description language(WSDL) and Universal description, discovery and integration(UDDI) protocol, in order to make the

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 445–453, 2013.

interaction available on the syntax level. However, this implementation model is not able to solve the problem of semantic Interoperability, or satisfy the requirement of some precise spatial information sharing and Service Reusability.

In the past several years of development on semantics and ontology with geographic information and web applications [1-4], it has provided a new trend and model for geographic information description, discovery and integration implementation. Starting from studying the current situation of applications on geographic information services, this paper proposes a framework of Regional geographic information platform Geographic Information Services Integration (GISI) based on OWL-S, and integrate geographic service semantic information with traditional syntax-based web service bus.

2 Geographic Information Services and Service Ontology

2.1 Geographic Information Services

Traditional geographic information applications are almost exist in an independent and complete form [5], which depends on some platform and technology. Due to the difference between implementation techniques and spatial data model, it is hard to share resources and interact among applications. However, the goal of the geographic information services is to integrate the geographic distributed, autonomous management and Heterogeneous data sources(including spatial data and related Non-spatial data), which provides a method and a function that allows users to obtain any geographic information transparently in a standard service form and deal with geographic information.

Under the advancement of Service-Oriented Architecture (SOA) and Web related technologies, it is built in an open geographic information services under which secondary development has been increasingly becoming a mainstream model for geographic information services. Geographic information services have gradually switched to the standard direction. Current related industry standards include ISO19119 that defines the categorization of geographic information services (its characteristics are items are big and general, mainly for distributed geographic information services development providing categorization framework and guidance. It then builds the mapping between the categorization and physical implementation considering the GIS application systems demands) and OGC OWS services categorization (its characteristics include web-based, focused on HTTP and XML technologies for services categorization and service port definitions, which provides standard ports for each kind of service. Also, any satisfactory standard application can interact with the services defined in the service framework and therefore get more support from different companies). Combing information technology and the standard framework implementation, it provides the structure of geographic information services, as shown in Figure 1.

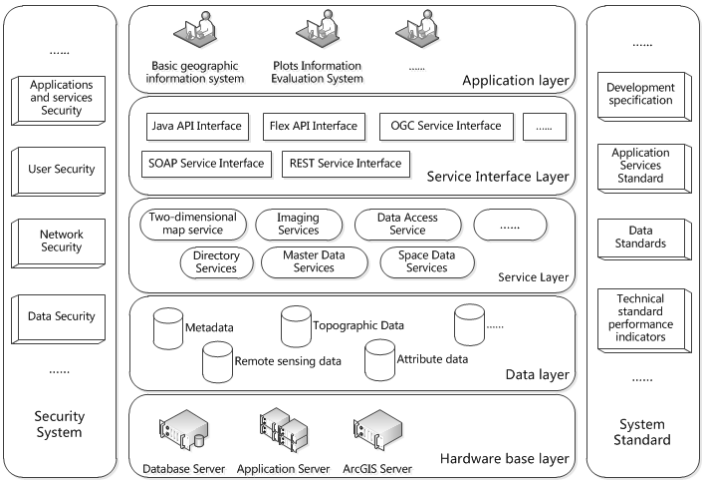


Fig. 1. A Hierarchy of Geographic Information Services

Under this kind of service-oriented structure, the key is the geographic information integration and sharing on web service technology. The technology foundation is based on the syntax service expression and its interface. However, for the expression of geographic information service semantics, integration has not provided enough restriction and corresponding solutions. On the other hand, foundational geographic information ontology study has been being used under some specific geographic information system environment. It is mainly offset the problems confronted in the traditional inquiries of geographical entities by leveraging geographic information's various semantic representations. It also focused on the discovery and indexing for the geographic information data and different geographic data on which the models are built. These approaches solved some of the practical problems, but they have not addressed the issues in the study of geographic information service framework as a whole.

2.2 Geographic Service Ontology

With today's fast internet advancement, information services are generally published by an information carrier web. These related technologies and standardshave been greatly improving the development of the web serviceson the other hand. However, in order to make the web-based server more automatic and organize web services more effectively, the problems in current service information format heterogeneous, information semantic multiplicity and the lack of information relationships have to be solved. The various semantics in ontology and its relationship can be used to address the above problems by bringing the concept of ontology and related technologies into web services techniques. With the support of ontology during the process of service mapping, the Precision rate and Comprehensiveness can be greatly improved comparing with the methods that are only based on simple keywords mapping algorithms. Service ontology takes care of service automatics. Meanwhile, it provides services with a shared semantic model in application area, and hence make the services interact on the semantic level.

The Semantic Web Services Framework(SWSF) proposed in the World Wide Web Consortium(W3C) in 2005 is the backup standard in the web service semantic[6], which combine the Semantic Web Service Language (SWSL) and Semantic Web Service Ontology(SWSO) that provides a process-oriented model. In August 2007, Semantic Annotations for WSDL and XML Schema(SAWSDL) that meets the recommended standard WSDL and XML schema by W3C tried to pass the WSDL and XSD elements cited in semantic model method to connect web service and semantics net[7]. These semantic models are defined outside WSDL, while the citation is assigned by marked, using WSDL, WSDL 2.0 and XSD extendable framework.

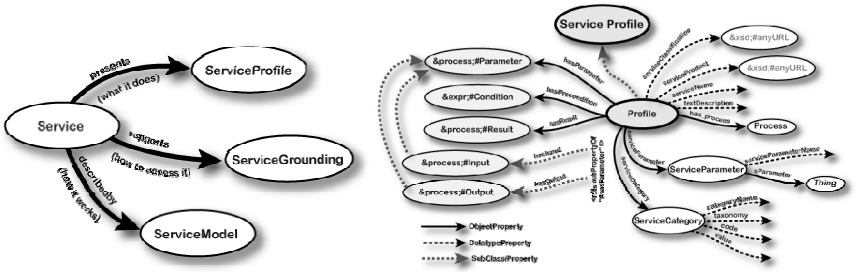


Fig. 2. Top level of The Service Ontology and Service Profiles

Some of the standard organizations have done parts of these kinds of work which introduced specific web-based model of ontology, such as OWL-S, WSMO, WSDL-S, and so on. Among them, W3C recommended OWL-S. As an ontology web language for services, OWL- is a combination of Web service and semantic web. The design mainly focuses on the problem of web service description and discovery, along with the semantic expression issues associated with the combination. Since May 2001, OWL-S has been recognized by many researchers. The current version is V1.2 which is released in 2008.

As shown in Figure 2, OWL-S generally includes three components: ServcieProfile, ServiceModel and Service Grounding. ServiceProfile describes the function of the service, namely what does the service do, service seeking agent through Service Profile mapping services to find the web service that satisfies the request. Service-Model describes how the service is done, namely the details in the service implemen-tation. ServiceGrounding describes how to access services. In [8], the paper discusses how to bring semantic into Web service based on OWL-S. In [9-11], the ontology applications based on OWL-S have been verified to some extent, which brings about certain practical merits. However, in the geographic information service, bringing in service semantic for semantic injection also needs to take the characteris-tics related to geographic information service into consideration, as illustrated by Table 1.

Table 1. Features of Geographic Information Services

Features	Geographic Information Services	Other Web Services
Spatiality	One of the most importantcharacteristic, Spatiality can describes the service location, shape, and spatial object topological relations.	None
Multi-scale	Different observations or clients require the use of thedifferent scales and different scale precision with service.	Has, different forms
Time and space	A strong characteristic to Geographic information service. A system based on GIS requires the use of different scales on geospatial for expression. A service is an integration of data sources with different time and different scales.	Scarcely
Abstraction	Describes real-world surface features and geomorphological features. Very complex andthe same natural features may have different semantic representation at the different levels of abstraction,	Has, different forms

For instance, a regular geographic information service based on ArcGIS, its service ontology can be described as follows:

```
......
<service:Service rdf:ID="GetLengthsService">
<service:describedBy>
<process:AtomicProcess rdf:ID="GetLengthsProcess"/>
</service:describedBy>
<service:presents>
<profile:Profile rdf:ID="GetLengthsProfile"/>
</service:presents>
</service:Service>
<profile:Profile rdf:about="#GetLengthsProfile">
<profile:hasOutput>
  ......
</profile:hasOutput>
<profile:hasInput>
  ......
</profile:hasInput>
<profile:textDescription>a geo-
service</profile:textDescription>
<profile:serviceName>GetLengths</profile:serviceName>
<service:presentedBy rdf:resource="#GetLengthsService"/>
</profile:Profile>
.....
```

In conclusion, it is reasonable to add service ontology semantic to the SOA implementation framework. On one hand, it provides more accurate service mapping and foundation for indexing. On the other hand, it essentially considers the various semantic relationships among geographic information service entities, geographic entities, and the integration of geographic information service.

3 A Combinatory Framework of GISI

Geographic information service integration is the key of geographic information service platform. Firstly, it should meet SOA structure. it means to develop and implement the services of multi-source and massive basic space information resources and related management under the stable network transmission, relevant policy and rules, and existing standard and information security restriction, Secondly, it provides secondary development interface that meets service standards to implements the data interactions based honeterogeneous platforms. Under the service integration of application environment, Enterprise Service Bus(ESB) is considered as an implementation method based on SOA structure. Because ESB has been an key method among the service communication. The figure below shows a geographic information integration framework implementation based on regional level ESB.

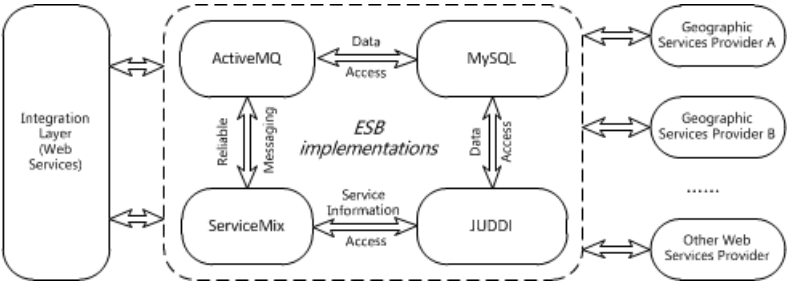


Fig. 3. A Framework of GISI Based on ESB

ESB adopts the open source ServiceMix, which provides hot deployment and Java Business Integration(JBI). It is compatible with similar JBI standard and ESB products to make it easy by using the third-party component libraries. ServiceMix is a ideal ESB implementation for a regional level service platform. In addition, as for the service integration requirement, systems usually use Java Messaging Service(JMS) broker to provide support implementation. Ands Apache ActiveMQ, as a widely used mature JMS, is also seamlessly integrated into ServiceMix. JUDDI as the implementation of UDDI, has provided service standard access interface under the support of the MySQL database.

In order to improve the existing model, the service ontology is integrated into the existing system (namely adding associated service semantics information to existing service syntax expression). The improved integrated framework is shown in Figure 4:

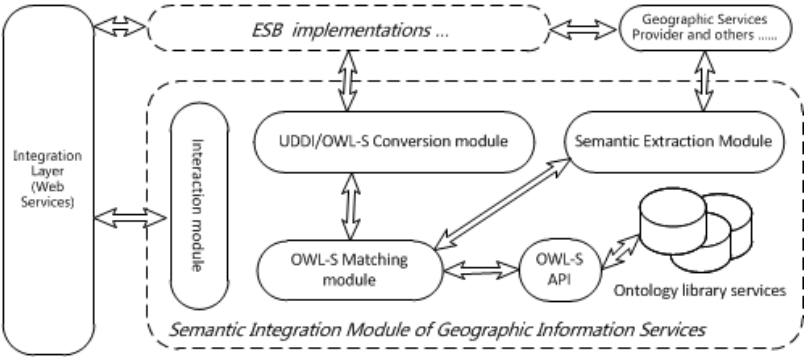


Fig. 4. A Combinatory Framework of GISI

4 Implementation of Semantic Integration Module

In Figure 4, a semantic Integration Module is responsible for adding semantics information to all heterogeneous geographic information services. In the process of implementation, ontology that meets the geographic information services standard (providing standard WSDL standard description) can be transformed in two ways.One is automatically acquired by using OWL-S based on an open source API coding. The other is manually manipulated by using the provided graphical tools (as shown in Figure 5) to get a specific geographic service ontology structure.

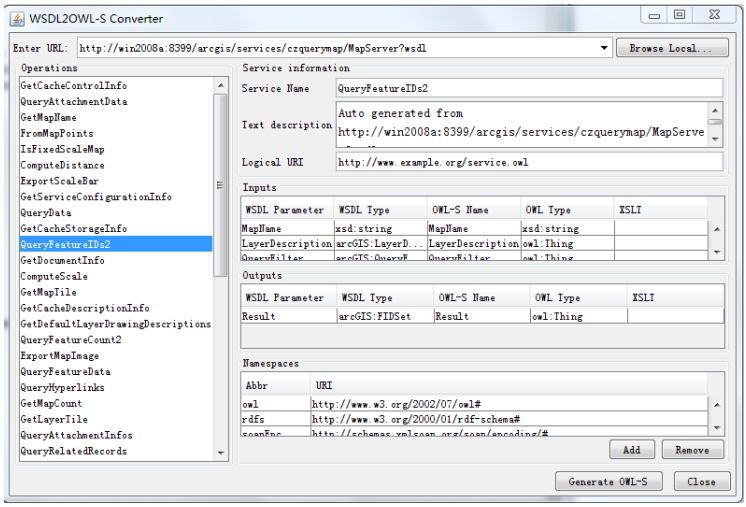


Fig. 5. Top level of The Service Ontology and Service Profiles

Considering the application environment of regional geographic information service integration, the levels and categorizations of services are not necessarily comprehensive and huge. It can be classified as three types based on the need: basic

geo-graphic information service, thematic geographic information service, and attribute information service. The integrated service mainly focuses on geographic information service. the association with attribute information service is carried out by the ID of geographic object (such as plot ID、point ID, etc.), which can be added by automatic service semantics as shown in the following steps:

Step1: Create a service ontology library to determine the storage solution. There are two ways to make it. One is with simple OWL files. The other is by using database method. OWL API provides some interface to implement it.

Step2: Construct semantic extraction module. Through extracting information from the description file attached by the existing geographic information services and related properties service, the extension module is developed to read the sematic information automatically. Because the implementation of geographic information services, the interfaces and the operation are more relatively fixed. It's beneficial for some of the services to be built by the appropriate body structure reflecting the service characteristics. In other words, specific implementation in OWL-S API is based on the body structure of the obtained file through adding semantic information corresponding matched.

Step3: Establish service entities and ontology matching by using the database and storing some one-to-many relationships of service entity and service ontology.

Step4: After the steps above, build interactive modules for the upper application with related interfaces of ontology-based semantic integration of geographic service.

5 Conclusion and Future Works

After the study, we make the following conclusion: with the help of semantics web and related technologies, the standard geographic information service expression will be evolved from the method of using the only traditional syntactic structure to the way by integrating with the additional semantics information. This new method attempts to improve the accuracy and precistion of the service manipulation and precision capability under web environment. Upon on this, the future work will focus on the next two directions: one is to make clear, how to summarize service ontology transformation unified model by studying the common and difference among geographic information service implementations. The other is to study how to use Web semantics and related technologies to make semantics inferences with the extracted geographic information service ontology exploring a more reasonable semantic interoperation on the platforms of geographic information services.

References

1. Yang, K., Wang, J., Peng, S.-Y., Heng, H.-P.: The constructing and practice of GIS ontology service mechanism. In: MIPPR 2005: Geospatial Information, Data Mining, and Applications, October 31-November 2 (2005), SPIE - The International Society for Optical Engineering; LIESMARS, China; Wuhan University, China
2. Huang, Y.Q., Cui, W.H., Zhang, Y.J., Deng, G.Y.: Research on Development of Agricultural Geographic Information Ontology. Journal of Integrative Agriculture 11(5), 865–877 (2012)

3. Khelifa, D., Mimoun, M.: Ontology based semantic integration of heterogeneous geographical information systems. In: 2008 3rd International Conference on Information and Communication Technologies: From Theory to Applications, ICTTA, April 7-11 (2008)
4. Luo, A., Wang, Y., Wang, L., He, Y.: Multi-level semantic matching of Geospatial Web Services. In: 2009 17th International Conference on Geoinformatics, Geoinformatics 2009, August 12- August 14 (2009), IEEE GRSS; NASA; NOAA; OGC
5. Perrin, M., Zhu, B.T., Rainaud, J.F., Schneider, S.: Knowledge-driven applications for geological modeling. Journal of Petroleum Science and Engineering 47(1-2), 89–104 (2005)
6. Web Services Description Language (WSDL) Version 2.0. W3C Note (2007), http://www.w3.org/TR/wsdl20/
7. Semantic Web Services Framework (SWSF) Overview (2005), http://www.w3.org/Submission/SWSF/#OWL-S-white-paper
8. Martin, D., Burstein, M., McDermott, D., McIlraith, S., Paolucci, M., Sycara, K., McGuinness, D., Sirin, E., Srinivasan, N.: Bringing Semantics to Web Services with OWL-S. World Wide Web 10(3), 243–277 (2007)
9. Bensaber, D.A., Malki, M.: Development of semantic web services: model driven approach. ACM, Lyon (2008)
10. Martinez, A., Patino-Martinez, M., Jimenez-Peris, R., Perez-Sorrosal, F.: ZenFlow: a visual Web service composition tool for BPEL4WS. In: 2005 IEEE Symposium on Visual Languages and Human-Centric Computing, pp. 181–188 (2005); Maximilien, E.M., Singh, M.P.: A framework and ontology for dynamic Web services selection. IEEE Internet Computing 8(5), 84–93 (2004)

The Analysis on the Coupling Characteristics of Ecological Environment, Natural Disasters and Poverty in Inner Mongolia Autonomous Region

Burenjirigala[1,*], Alatantuya[1,2], and Chunrong Guo[3]

[1] Natural Disaster Prevention Research Institute of Inner Mongolia Normal University, Hohhot, Inner Mongolia 010022
[2] Geographical Science Institute of Inner Mongolia Normal University, Hohhot, Inner Mongolia 010022
[3] Ordos Land Surveying and Planning Institute, Kangbashi City, Inner Mongolia 017000

Abstract. The high coupling of regional poverty and ecological environment degradation, natural disasters is the important reason for the unsustainable development of poverty-stricken areas .It has practical significance to correctly understand the interactions between the regional poverty, ecological environment and natural disasters, research the coupled relationship of the three. This paper chooses the vulnerability of ecological environment, natural disasters hazard and regional poverty indicators that are based on the ecological environment features and their results, the main types of natural disaster and their harm, regional economy and social welfare., etc. The vulnerability ecological environment, natural disasters harm degree and regional poverty degree are obtained in counties through spatial interpolation, spatial overlay and comprehensive evaluation process. And Then four types and thirty classes are divided by using coupling analysis method and coupling relationship judgment standard .Finally it is conclude that the risk class with high fragility, harmfulness and poverty degree of ecological environment, natural disasters and regional economy in Inner Mongolia occupy dominant position, and the relationship that the ecological environment, natural disasters and the regional poverty are not harmonious in about 71% of the counties and cities and there are differences between the coupling degree and the coupling type in different geographical areas.

Keywords: ecological environment vulnerability, natural disasters hazard, regional poverty, coupling characteristics, Inner Mongolia Autonomous Region.

Chinese library classification number: N945.1

---

* Date of reception:

Author introduction: Dr. Burenjirigala (1964-), male, Associate professor, the Mongol nationality, born in Etuoke Front Banner, in Inner Mongolia, mainly engaged in disaster prevention and mitigation, and regional geography research. E-mail: burenjiri@163.com

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 454–465, 2013.

1 The Introduction

Natural disasters, ecological environment and regional poverty are the hot issues that are widely concerned by people from all walks of life, and there is close relationship and mutual influence, restriction and dependence among three. Since the 1990 s, people have a more profound understanding of risk relationships among the poverty, vulnerability of ecological environment and natural disaster. And put forward that the investigation of risk and vulnerability is the key to understanding poverty [1]. A large number of studies home and abroad have also shown that poor areas often located in the lowest resilience the worst environmental damage, natural disaster frequent area in the world, [2-8] which shows that the high coupling of poverty, ecological environment degradation and natural disasters is the important reason for the unsustainable development in poverty stricken area in a sense. Therefore, it has practical significance to correctly understand interactions among region poverty, ecological environment and natural disasters by studying the coupling relations of the three as a breakthrough point so to seek corresponding anti-poverty rich and health development strategies.

2 Index System, Data Sources and Research Methods

2.1 Index System

The indicators such as heat, water, vegetation are chosen when vulnerable ecological environment is evaluated according to the causes; the indicators of pressure and response of human activity, land use and reclamation, industry and the investment structure are chosen based on the features of results; the index of natural disasters species and its condition and natural disaster damage in Inner Mongolia is selected to build the evaluation index system of the natural disasters in Inner Mongolia; Inner Mongolia region poverty evaluation index system is mainly composed of the reaction of poverty degree economic index and social welfare, which are shown in table 1.

2.2 Data Sources

The above index system involved in the meteorological data is from meteorological observation data each county of Inner Mongolia autonomous region 35 years in 1975-2009; vegetation index is based on MODIS 1B data source, using the maximum synthesis obtains the 2002-2012 NDVI average. Various types of natural disasters level data are obtained from the Atlas of climate in Inner Mongolia [9], Atlas of Land resources in Inner Mongolia [10]; natural disasters hazard index data is from civil affairs department of Inner Mongolia autonomous region according to the county level administrative units for statistical data in 1978-2009; Other social and economic data is from statistical yearbook data of Inner Mongolia in 2010, which is later sorted and calculated.

Table 1. Inner Mongolia Region Poverty, Ecological Vulnerability and the Evaluation Index System of Natural Disasters

Index types	Index	
Fragile ecological environment	The ecological environment characteristics	precipitation, evaporation, moisture, ≥10℃ accumulated temperature, annual range of temperature, precipitation variability, vegetation index
	results	Population quantity and structure index（rural populations, land population carrying, the first population bearing industry, food production population carrying, livestock amount population carrying）
		cultivation index, Food crop acreage proportion, effective irrigated area proportion
		The second industry proportion, Proportion of the tertiary industry, Population carrying of investment in fixed assets
Natural disaster	Main kinds of natural disasters	drought, wind damage, frost, floods, earthquake, Soil and water loss, grassland degradation
	Harm of natural disasters	flood-hit population, the number of casualties, damage area, financial loss
Regional poverty	economy of region poverty	GDP per capita, Financial income population bearing, Per capita fiscal deficit, Per capita net income of farmers and herdsmen, Population carrying food production, Livestock amount population carrying
	social welfare of region poverty	social retail sales of consumer goods Population carrying, population bearing farming mechanical power, electricity consumption Population carrying, traffic population carrying, Telecom population carrying, Ten thousand people have professional health personnel number, Middle school students in school population carrying, Pupils in school population carrying

2.3 Research Methods

2.3.1 Data Processing Method

2.3.1.1 Spatial Interpolation Method. The data of meteorological observation sites need to use spatial interpolation method to change point data into regional surface data information. There are many spatial interpolation methods, of which spline method is not affected by spatial scale and does not directly depend on the space smoothly covariance; therefore, it is more practical [11-14]. Spline method is applied to do

spatial interpolation, forming the spatial distribution of meteorological factors in Inner Mongolia.

2.3.1.2 Spatial Overlay Analysis Method. Overlay analysis method is to have the vector graphics of the spatial indicators overlay, generating new elements by division or merger. New elements combine the original properties of several elements. The aim of overlay evaluation index of ecological environment vulnerability and natural disaster risk evaluation index is to establish the same spatial attribute database, operation, statistical analysis and quantitative, orientation so as to identify the quantity characteristics of each index in different areas.

2.3.2 Comprehensive Evaluation Method

The principle of main component projection evaluation model for fragile ecological environment, natural disasters and poverty area is to change the original indexes into comprehensive index which is orthogonal to each other. Then each main component is projected to an ideal decision as a comprehensive evaluation index of one-dimensional [15]. Firstly, the specific method is to establish evaluation sample matrix, and to eliminate influence from which dimension and magnitude for all kinds of indexes, and range standardization method is adopted to improve the normalization. And then indicators are disposed by means of the orthogonal transformation, to filter out duplicate information between indicators. Finally according to the main component scores and characteristic value of the product and the projection values are calculated, and according to the size of the projection value different grades are divided. Thus, the regional ecological environment vulnerability, risk of natural disasters and poverty in the region are evaluated.

2.3.3 Coupling Analysis Method

Coupling degree is the quantitative indicators to measure coordination condition between system and elements. Coupling analysis can characterize regional natural environment, regional economy overall function. The coupling relationship of fragile ecological environment, natural disasters and poverty areas can be expressed as the following formula [16-22]:

$$Cxyz=\frac{(x+y+z)}{\sqrt{(x^2+y^2+z^2)}} \quad . \tag{1}$$

x, y, z is for ecological environment vulnerability, risk of natural disasters, regional poverty degree, The smaller the value is , the securer it is , the bigger the value is , the more fragile, dangerous and poorer it is ; Cxyz is the coupling degree for ecological environment vulnerability, natural disaster risk and the region's poverty, $-1.732 \leq Cxyz \leq 1.732$, The greater for the Cxyz absolute value is , the more synergy for the ecological environment, natural disasters and areas of poverty is, which have the roles of mutual coordination and promotion. And instead, the smaller for the Cxyz absolute value is, it is shown that the relationship is disjointed and restricted with each other between the ecological environment, natural disasters and the regional poverty.

3 Results and Analysis

According to standardizing for the evaluation indexes data, the comprehensive value results from the principal component projection transformation for <-1, -1-0, 0-1, >1. It will be divided into good (low), medium, heavier, heavy (high) four levels for the ecological environment vulnerability, natural disasters and poverty area of each county in Inner Mongolia. And the coupling relationship between fragile ecological environment, natural disasters and poverty area is divided into coordination (synergy) and not harmonious for Cxyz absolute value, $|Cxyz| \geq 1$, coordination, $|Cxyz| < 1$, not harmonious.

According to coupling degree of ecological environment vulnerability, natural disasters and poverty area in each county of Inner Mongolia, regarding coupling relationship as judging standard and four types and 30 coupling relationship subgroup are divided as table 2.

Table 2. Ecological environment vulnerability (x), natural disasters (y) and regional poverty (z) coupling relationship (Cxyz) criteria and its coupling relationship types and their counties and cities

coupling relationship types and standards	coupling relationship subgroup	coupling relationship subgroup standards	county name
coordinate safety class -1.732<Cxyz<-1	low vulnerability, light disaster, low poverty type	x, y, z<-1	Manzhouli city, Hailaer district
	low vulnerability, light disaster, middle poverty type	x, y <-1;-1<z<0	Aershan city, Yakeshi city, Genhe city, Eerguna city, Chenbaerhu county, ewenkezu autonomous county
	low vulnerability, middle disaster , low poverty type	x, z <-1;-1<y <0	Huolinguole city, Jining district
	low vulnerability, heavier disaster, low poverty type	x, z <-1;0<y <1	Baotou city, Hohhot City
	low vulnerability, light disaster, higher poverty type	x, y <-1;0<z<1	Elunchun Autonomous County

Table 2. (*Continued.*)

	middle vulnerability, middle disaster, low poverty type	0<x, y <1;z<-1	Baiyunebo mining area, xilinhaote city
	middle vulnerability, middle disaster, middle poverty type	-1<x, y, z<0	Shiguai mining area, Dongwuzhumuqin county, Erlianhaote city
	middle vulnerability, light disaster, higher poverty type	-1<x<0;y<-1;0<z<1	Xinbaerhuzuo county, Zhalantun city
	high vulnerability, middle disaster, low poverty type	1<x;-1<y<0; z<-1	Etuoke county
	higher vulnerability, heavier disaster, low poverty type	0<x, y < 1; z<-1	Keerqin District, Wuhai city, Yijinhuoluo county
uncoordinate safety class -1<Cxyz<0	higher vulnerability, heavy disaster, low poverty type	0<x<1; 1<y; z<-1	Dalate county
	middle vulnerability, heavy disaster, low poverty type	-1<x<0;1<y; z<-1	Dongsheng district, Zhungeer county
	middle vulnerability, middle disaster, higher poverty type	-1<x, y<0; 0<z<1	Xiwuzhumuqin county, Molidawa county
	higher vulnerability, middle disaster, middle poverty type	0<x<1;-1<y, z<0	Etuokeqian county, Daerhanmaoming anlianhe county
	high vulnerability, middle disaster, middle poverty type	1<x;-1<y, z<0	Alashanzuo county, Wuyuan county, Hangjinhou county, Dengkou county
	higher vulnerability, middle disaster, higher poverty type	0<x, z<1;-1<y <0	Xinbaerhuyou county

Table 2. (*Continued.*)

Uncoordinated risk class 0<Cxyz<1	middle vulnerability, middle disaster, high poverty type	-1<x, y<0; 1<z Zhengxiangbai county, Huade county, Xinghe county
	middle vulnerability, heavy disaster, middle poverty type	-1<x, z <0;1<y Honhshan district, Xianghuang county, Helingeer county, Wulanhaote city
	middle vulnerability, heavier disaster, heavier poverty type	-1<x<0;0<y, z<1 Arong county, Keerqinzuoyihou county, Zhenglanqi, Kulunqi, Kalaqinqi, Keerqinyouyiqianqi, Chahaeryouyiqian County, Balinzuo County, Balinyou County, Wuchuan County Fengzhen City, zhaluteqi, keshiketengqi, duolunxian
	middle vulnerability, heavier disaster, heavy poverty type	-1<x<0;0<y<1;1<z taipusiqi, chahaeryouyizhongqi, zhuozixian, zhalaiteqi
	middle vulnerability, heavy disaster, higher poverty type	-1<x<0;1<y;0<z<1 linxixian, liangchengxian, ningchengxian
	higher vulnerability, middle disaster, high poverty type	0<x<1;-1<y<0;1<z abagaqi, shangduxian, chahaeryouyihouqi
	higher vulnerability, heavier disaster, middle poverty type	0<x, y<1;-1<z<0 tumotezuoqi, tumoteyouqi, guyangxian, tuoketuoxian, wushen
	High vulnerability, middle disaster, higher poverty type	0<x, z<1;-1<y <0 ejinaqi, alashanyouqi, suniteyouqi
	High vulnerability , heavier disaster, middle poverty type	1<x;0<y<1;-1<z<0 wulateqianqi, linhequ
	Higher vulnerability, heavy disaster, middle poverty type	0<x<1;1<y;-1<z<0 Kauluxian

Table 2. (*Continued.*)

collaborative risk 1.726>Cxyz>1	Higher vulnerability, heavier disaster, higher poverty type 0<x, y, z<1	Qingshuihexian, Siziwangqi, Keerqinyouyizhongqi, Naimanqi
	Higher vulnerability, heavy disaster, higher poverty type 0<x, z<1;y<1	keerqinzuoyizhongqi, Alukeerqinqi, Wengniuteqi, Aohanqi
	Higher vulnerability, heavier disaster, high poverty type 0<x, y<1;z<1	Tuquan County
	High vulnerability , heavier disaster, higher poverty type 1<x;0<y, z<1	Wulatezhongqi, Hangjinqi
	High vulnerability , heavier disaster, high poverty type 1<x, z;0<y<1	Wulatehouqi, Sunitezuoqi

The fragility among Inner Mongolia ecological environment, natural and regional disaster, and the coordination between poverty coupling and security classes have five sub-categories, including 13 counties in the city, accounting for 14% of counties and cities of Inner Mongolia Autonomous Region, mainly in Inner Mongolia northeast regions and cities. Good ecological environment in these regions, only a few flags in the eastern Hulunbeier City's economic development level, is lagging behind the rest of the natural disasters in several cities less risk than other types of flags of the city. Overall ecological environment, natural disasters and poverty minimum pressure, has among mutual promotion and common to the health, well-direction.

Uncoordinated safe coupling classes have 11 sub-categories, including 23 counties in the city, accounting for 25% of the city of Inner Mongolian Autonomous Region counties, mainly in the coordination of safety classes around the periphery of the city and the flag of south-central regions. This type of fragile environment, natural disasters and poverty hazards degrees to moderate mainly on regional development in the overall risk is a relatively small, relatively safe state. In its interior is a regional differentiation in the eastern region of the ecological environment, which is relatively good in the west, natural disasters, hazards degree moderate, with some favorable natural basis, but regional socio-economic development lags behind. Its natural conditions are not commensurate with relatively favorable natural conditions which are not allows social and economic plight. And relative poverty is impossible for the socio-economic to provide effective long-term ecological and environmental protection, good cultural environment, among mutual restraint, mutual hinder healthy development. The class counties in western ecological environment is fragile and severe natural disasters, while poverty is low, the natural environment and the socio-economic development does not match the natural disaster vulnerability and more affluent social environment can not

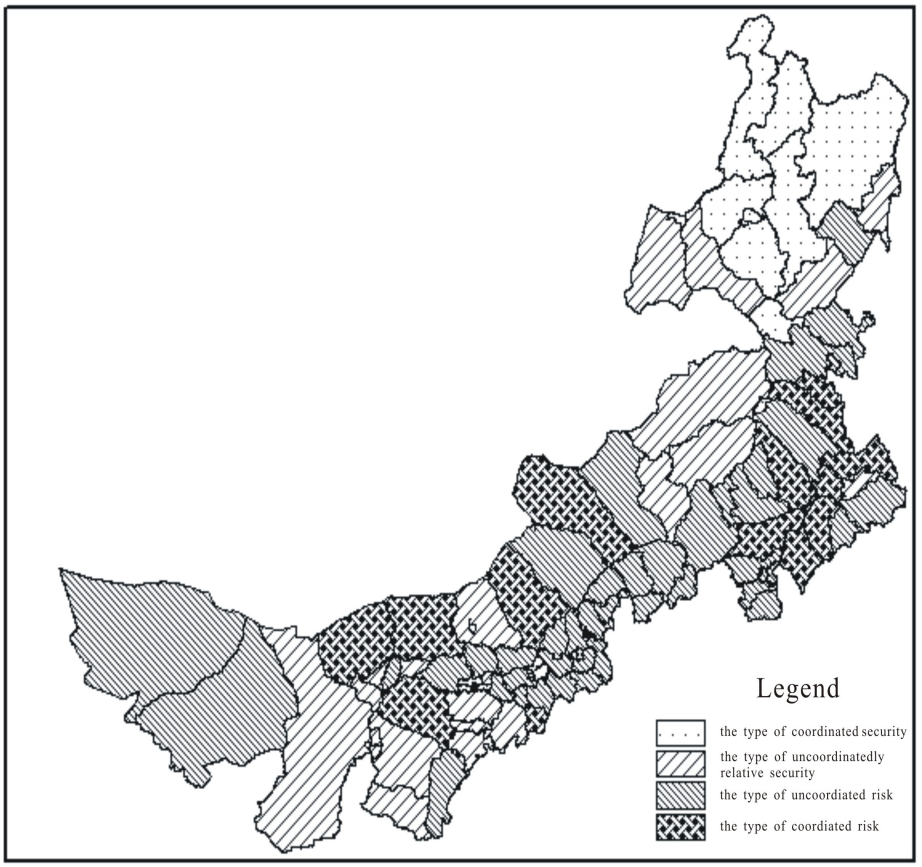


Fig. 1. Fragile ecological environment, natural disasters harm degrees and regional poverty degree of coupling type map in Inner Mongolia

provide comfort and support economic external environment. The same economic forces alone to improve the fragile ecological environment and prevention of natural disasters, but also difficult to achieve real problem, also has among mutual constraints hinder the healthy development of relations.

Uncoordinated dangerous class of coupling relations have 9 subcategories, a total of 42 counties in the city, accounting for 46% of the city counties in Inner Mongolian Autonomous Region. This distribution is very scattered, which is hills and desert-based. This type of Fragile Environment, natural disasters and poverty hazards degrees higher degree mainly on regional development in the overall relative risk status. Among them, Xilin Gol League South and Southern District Wulanchabu fragile ecological environment, natural disasters, heavier. What's more important feature is that the outer areas of poverty highlighted; western desert, steppe desert area and central desert steppe area in addition to severe disasters, and the socio-economic is more prominent than not rich ecological whose environment is extremely fragile;

southern hilly area of serious natural disasters. Overall area of this type of environment, natural disasters and poverty, pressure, regional development at risk is at a relatively large state, but did not belong to a higher risk of all these three highly coupled, at least among a class for medium, ecological environment, natural disasters and poverty with some regulatory role, who will not the system as a whole to malignant direction.

Synergistic coupling between risk categories have five sub-categories, including 13 counties in the city, accounting for 14% of the city counties in Inner Mongolia Autonomous Region, which is located in the southeast and north-central. This type of vulnerability of the ecological environment, natural disasters and hazards degrees are higher and the high degree of poverty, and the high degree of coupling between three forms of a fragile ecological environment lead to regional poverty, severe natural disasters exacerbate poverty and economic ecologically fragile, people living in poverty and disaster survival to obtain the immediate short-term benefits over predatory further use of environmental resources and destroy the ecological balance, damage ecological functions, resulting in deterioration of ecological environment, and thus the formation of poverty - ecologically fragile - the vicious natural disasters cycle of the cycle, regional development at risk.

4 Conclusion and Discussion

From the perspective of Inner Mongolia ecological environment and natural disasters and regional poverty, only a few cities in Northeast China and be safe outside the class, is more dangerous and dangerous classes have 55 counties in the city, accounting for 60%, showing that, Inner Mongolia ecological environment and natural disasters and regional economic vulnerability, hazard, high levels of poverty risk class occupies a dominant position. Ecological environment, natural disasters, poverty, coupled with regional relations, not simultaneously $x > 1$, $y > 1$, $z > 1$ though, and $1.726 > Cxyz > 1$. The coupling relationship is not completely out of control, but there are 14% of the flag counties in fragile ecological environment and natural disasters, poverty and regional disaster of a strong coupling and among the three ecological environment is fragile areas, areas prone to natural disasters and poverty areas in the distribution of geospatial highly additive. In addition, 71% of the counties in the city's ecological environment, natural disasters and regional poverty relationship disharmony between different geographical areas and the degree of coupling, the coupling type are differences. And between the natural environment and the local economy has contradicted each other, mutual restraint involved role. In general, the environmental and economic well-coordinated development is not only an important way to achieve sustainable development, but also the elimination of poverty, comprehensive and healthy development of the basic requirements. Inner Mongolia region resulting poverty and ecological environment system, natural disaster systems, regional economic system is inextricably linked, each system is the key point there cause of poverty, which is a breakthrough in anti-poverty measures and starting point. As long as a fundamental solution to one of the key links, we can lay the foundation for establishing a virtuous circle. In addition, the region's natural, social, economic overall effect depends on the

ecological environment, natural disasters, the coupling between regional economies. In order to rationalize poverty, natural disasters and ecological environment, and analyze the three coupling characteristics, and further study of their structure and function is needed. According to regional characteristics and causes of poverty, the actual situation, the most suitable model of development can be chosen and ways of good mutual promoting relations can established and safeguarding regional development and is the dynamic stability of poverty can be thus achieved , which is the necessary link of healthy development.

Acknowledgements. Financial support was from National Philosophy and Social Science Fund Project（05XJY004); National Natural Science Foundation of China (NSFC) (41261099).

References

1. Chistiaense, L.J., Subbarao, K.: Toward an Understanding of Household Vulnerability in Rural Kenya. World Bank Policy Research Working Paper No. 3326 (2004)
2. Dercon, S.: Growth and Shocks: evidence from Rural Ethiopia. Journal of Development Economics (August 2004)
3. Holzmann, R., Jorgensen, S.: Social Risk Management: A New Conceptual Framework for Social Protection, and Beyond. International Tax and Public Finance 8, 529–556 (2001)
4. Zhao, Y., Liu, Y.: Fragile Ecological Environment Distribution and Its Relationship with Poverty in China. Advance in Earth Sciences 11(3), 247–251 (1996)
5. Zhu, L.: On the Social Safety Net. Population Science of China 3 (1993)
6. Du, G., Wang, G.: The study About China's Agricultural Insurance and Rural Social Security System. The Capital Economic and Trade Study Press, Beijing (2003)
7. Wang, G.: Agricultural Natural Disasters and Rural Poverty Problem Research. Economists 3, 55–62 (2005)
8. Tong, Y., Long, H.: Fragile Ecological Environment under the Coupling of Poor Areas of Sustainable Development Research. China's Population, Environment and Resources 2, 47–51 (2003)
9. Meteorology Bureau of Inner Mongolia: Climatic Atlas In Inner Mongolia. Inner Mongolia Map Reproduction Institute, Hohhot (1998)
10. Territorial Control Office of Inner Mongolia Planning Commission: Atlas of Land and Resources of Inner Mongolia Inner Mongolia People's Publishing House, Hohhot (1998)
11. Wahba, G.: Spline Models for Observational Data. In: CBMS-NSF. Regional Conference Series in Mathematics, vol. 59. SIAM, Philadelphia (1990)
12. Collins, F.C., Bolstad, P.V.: A Comparison of Spatial Interpolation Techniques in Temperature Estimation. In: Proceedings of the Third International Environmental Modeling, Santa Fe, New Mexico (1996)
13. Dubrule, O.: Two Methods with Different Objectives: Splines and Kriging. Mathematical Geology 15, 245–257 (1983)
14. Laslett, G.M.: Kriging and Splines: An Empirical Comparison of Their Predictive Performance in Some Applications. Journal of the American Statistical Association 89, 391–400 (1994)

15. Li, M.: New Methods of Multiple Attribute Decision Making and Evaluation of the Principal Component Projection Method. Journal of Mathematical Statistics and Management 20(5), 45–48 (2000)
16. Bi, J., Zhang, S., Tang, Y., et al.: The Discriminant Model and Its Application of the Sustainable Development. China Environmental Science 19(1), 30–36 (1998)
17. Zhang, X., Chi, T.: China's Provincial Regional Economy and Environment Coordinate Degree Analysis in the 1990s. Journal of Geographical Research 20(4), 506–509 (2001)
18. Wu, D., Cao, L., Chen, L.: Synergetics Theory and Application. Huazhong University of Science Press, Wuhan (1990)
19. Song, X., Liu, Y.: The Coupling of Urbanization and Ecological Environment Model and its Application. Journal of Resources and Environment 23(5), 31–33 (2005)
20. Liu, Y., Li, R., Song, X.: China's Urbanization and Eco-environment Coupling Analysis. Journal of Natural Resources 20(1), 105–112 (2005)
21. Liu, Y., Li, X.: Fragile Ecological Environment and Sustainable Development. The Commercial Press, Beijing (2001)
22. Wang, S., Dong, S., Wang, X., et al.: The Interaction Mechanism Research on Economy and the Ecological Environment in Dingxi City. Journal of Resources Science 27(4), 106–111 (2005)

Spatio-temporal Analysis of Weibo Check-in Data Based on Spatial Data Warehouse

Liang Zhou, Mingye Bao, Nanhai Yang, Yizhen Lao, Yun Zhang, and Yangge Tian*

International school of software, Wuhan University, Hubei, 430072
tiandebox@whu.edu.cn

Abstract. With the increasing development of the application of location services, massive check-in data is produced by social media applications on the mobile appliances, which includes characteristics of spatio-temporal information, user-emotion information, and etc. Traditional analysis techniques cannot handle check-in data well because of the complexity of spatio-temporal information. Spatial data warehouse provided a good architecture for spatial data's storage and analysis. In this research, we designed a spatial data warehouse to store and manage the check-in data, used OLAP analysis technology to analyze it, and found many interesting results. It showed spatial data warehouse and OLAP provided a good frame to analyze check-in data.

Keywords: check-in data, spatial data warehouse, OLAP.

1 Introduction

With the increasing development of social media applications on the mobile appliances (such as Foursquare, Tweeter, Facebook, Weibo, QQ space), more and more people enjoy sharing their own location information, causing massive increase of check-in records in the Internet. Those check-in data shared by people, contains massive spatio-temporal tags and some rules of our daily lives to research. Nowadays, it captured lots of researchers' attention. For example, Linna Li presented the spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr[1]; Liangjie Hong presented their discover of geographical topics in the Twitter Stream in the proceedings of the 21st International Conference on World Wide Web[2]; Nilanjan Banerjee conducted an exploration research with micro-blogs for user interests in social media sites[3]; Eunjoon Cho used check-in data to study user movement in location-based social networks[4]; Salvatore Scellato exploited place features in link prediction on location-based social networks[5], and etc. They provided good references in this direction of research.

Check-in data is a new sort of network data, including spatial and temporal characteristics, also the user-emotion information. Spatial data warehouse provide a good base to store and manage spatial data. We can use OLAP to analyze check-in data and combine its result with GIS to provide a deep insight of check-in data.

Spatial data warehouse provide a good base to store and manage spatial data. We can use OLAP to analyze check-in data and combine its result with GIS to provide a deep insight of check-in data.

* Corresponding author.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 466–479, 2013.

There are many studies about spatial data warehouse, for example, L.Savery addressed a spatio-temporal data warehouse design for human activity pattern analysis[6]; Sonia Rivest introduced SOLAP technology which merged business intelligence with geospatial technology for interactive spatio-temporal exploration and analysis of data[7]; Sergio Di Martino integrated Google Earth within OLAP Tools for multidimensional exploration and analysis of spatial data[8]; Dimitris Papadias argued that the spatial and temporal dimensions should be modeled as a combined dimension on the data cube and present data structures, which integrate spatiotemporal indexing with pre-aggregation in their research[9], Gennady Andrienko addressed a method of spatio-temporal Aggregation for Visual Analysis of Movements[10]. These designs contain key techniques such as online analytical processing based on spatio-temporal data warehouse. Also, they can provide us with some references about designing spatial data warehouse.

Based on the existing researches, we chose check-in data of Wuhan as the study object, designed a spatial data warehouse to store and manage the check-in data, used OLAP analysis technology to analyze it, and found many interesting results. It showed spatial data warehouse and OLAP provided a good frame to store and analyze check-in data.

2 Data Warehouse Architecture and Data Organization

In this study, we built a spatial data warehouse for check-in data's storage and management. It provided a good basis for analysis of data.

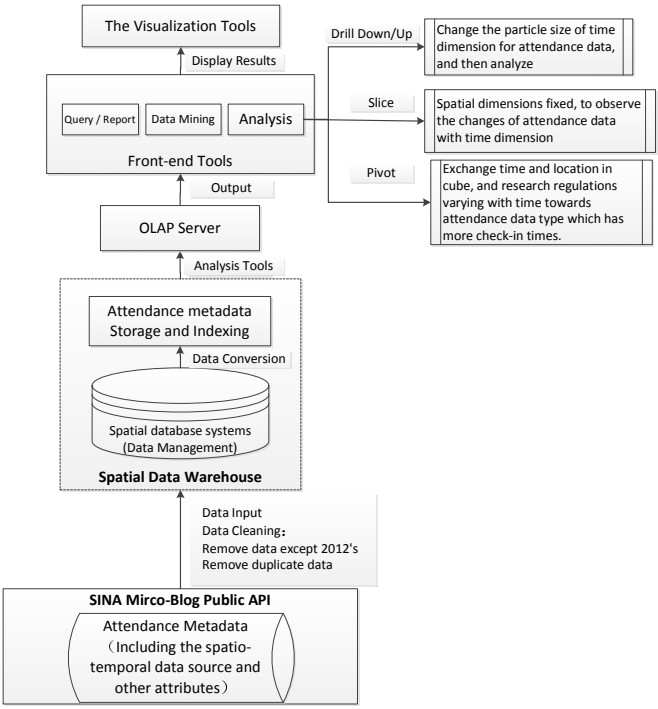


Fig. 1. Spatial data warehouse architecture diagram in check-in data analysis of the research

As shown in Fig.1, we use the public API interface of weibo(SINA) as data-acquisition system to acquire check-in data, and give them to spatial data warehouse for data cleaning. After transformation and integration of check-in data, we use technologies like OLAP and Data ming to analyze massive data. Finally, we display the results depending on the visualization tools and present corresponding conclusions.

In the specific process of data processing and organization, firstly we apply geo-coding to the spatial data and import them into spatial data warehouse. Then, we choose the check-in data in 2012 and basic geographic data in Wuhan City as the analysis object of our study. According to the geo-coding, we associate spatial coordinates of check-in data with spatial entities and integrate data into the data warehouse.

Based on the existing spatial data warehouse, we further classify the data of check-in time tags separately by month, week, day, hour, and other time granularities, simultaneously record the information in the spatial dimension such as name, attributes, characteristics and etc. Then according to the analysis granularity, we select the user ID, check-in time, check-in locations of longitude and latitude, specific address and other attributes, so as to lay a foundation for studying the spatial distribution of check-in data in different temporal granularities in the future.

For the convenience of analysis, we use the Star-model to organize the check-in data in the database.

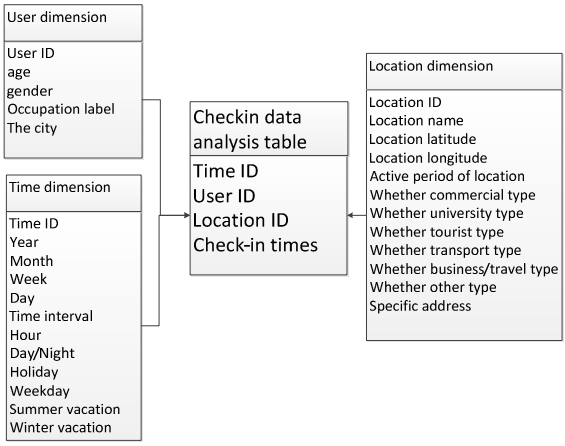


Fig. 2. Star-model architecture diagram

In the design of Star-model showed in Fig.2, check-in data analysis table is a fact table, the check-in times is the fact which should be observed from perspectives of location, time and user. Besides, the fact table provides corresponding indexing of

dimension. The other three dimension tables contain their own attributes and relative description information, which provide the study of check-in times with convenience. Before studying the check-in times, we set the time granularity. Then, we filter the check-in records by referring to the location-dimension table and the user-dimension table, and collect statistics about these dimensions' attributes like different location types. For example, when the time granularity is the month, we can count total check-in times of different places in different months. We can remove the duplicate check-in data depending on the user ID there, and acquire datasets for further analysis.

On this basis, combining the OLAP analysis methods like slicing, drilling, pivoting and GIS or other measures, we analyzed the spatio-temporal characteristics of check-in data and found some interesting conclusions.

3 Analysis

In the research, we found that it is an effective measure to analyze the spatio-temporal characteristics of check-in data by using OLAP analysis methods.

As mentioned in Fig.1, there is a close relationship between spatial data warehouse and OLAP. OLAP uses multidimensional data sets and data aggregation technology to organize and summarize the check-in data in the spatial data warehouse. Later on in the research, we apply several OLAP analysis methods like drilling-up/drilling-down, slicing, and pivoting. After completing the analysis work, we further use the OLAP visualization tools to display these results so as to acquire the information and connotation contained in the check-in data.

Our experiments showed that OLAP have a great help in the analysis of temporal and spatial variation towards the check-in data. We will use drilling and slicing techniques to explain in detail as follows.

3.1 Analyze by Using Technology of Drilling and Display Different Spatial Regulations of Check-in Data under Different Granularities in Time Dimension

In the studying of different time granularities, we found that drilling along the time dimension which put the check-in locations and check-in times altogether to acquire the spatial distribution of check-in point is an effective measure to study the spatio-temporal distribution regulations of check-in data.

We also found that it displayed different patterns in different time granularities. Observed from our experimental results, time granularities like the week, day/night, 6h(time period), holidays which have more obvious spatio-temporal distribution laws are the good perspectives to analyze.

We set an example diagram below which obtains the check-in data in December under the granularity of week to explain the principle of drilling(by rolling up).

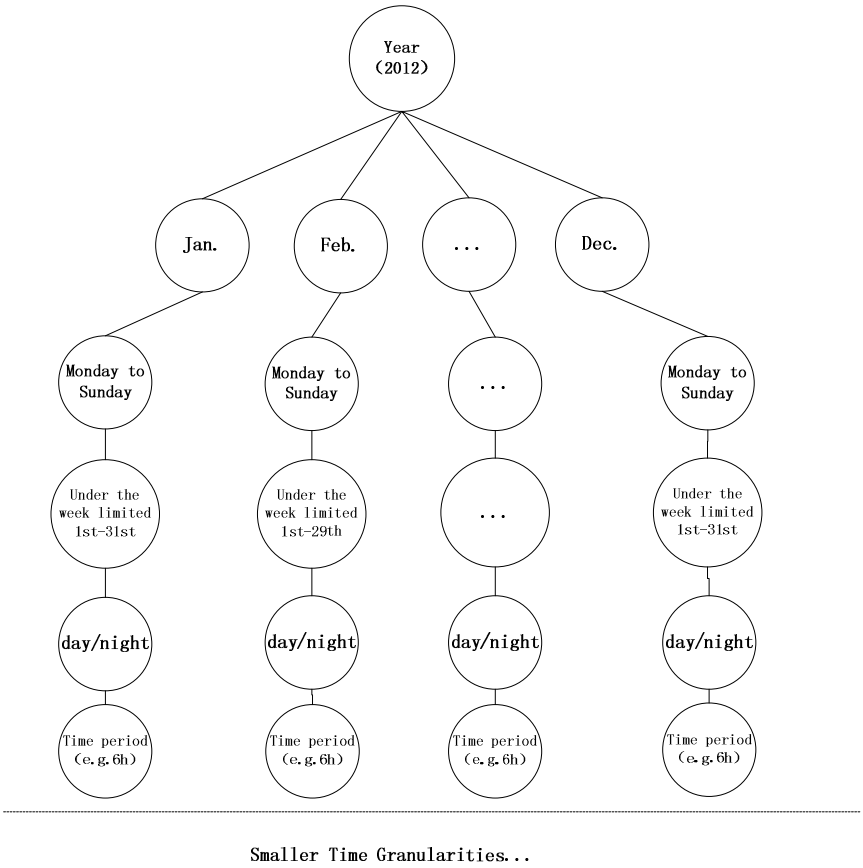


Fig. 3. Hierarchy diagram of drilling along the time dimension

There is the method description for the diagram in Fig.4: firstly, we change the hierarchy of the time dimension, from day(in December) to week(in December), then remove the location dimension, and get the summarized data of the week level in December.

After analyzing by using technology of drilling from different granularities, we found that the spatial pattern is more obvious under the time granularities including the week, the day, 6h (time period) and the holiday.

3.1.1 Week

Check-in data shows obvious fluctuations in a week. Set statistical data in December as an example, we use the methods like drilling-down and rolling up to acquire statistical check-in times from Monday to Sunday.

There is a chart of check-in fluctuations in Fig.4.

Time / Check-in Times / Place	Monday					Tuesday				Wednesday				...
	3	10	17	24	31	4	11	18	25	5	12	19	26	...
Wuhan University	3	3	8	4	7	5	13	9	17	2	13	12	22	...
Wuhan Institute of Physical Education	2	5	15	16	4	3	2	8	19	3	9	14	11	...
Huazhong Normal University	2	11	11	9	6	2	12	17	7	3	11	12	10	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Time / Check-in Times / Place	Mon.	Tue.	Wed.	...
Wuhan University	25	44	49	...
Wuhan Institute of Physical Education	42	31	37	...
Huazhong Normal University	39	38	36	...
⋮	⋮	⋮	⋮	⋮

Time	Mon.	Tue.	Wed.	...
Check-in Times	1084	788	859	...

Fig. 4. Example diagram of obtaining the check-in data by drilling

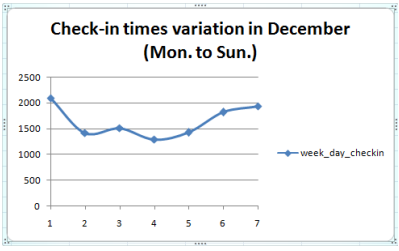


Fig. 5. Chart of check-in fluctuations in December

(x: from Mon. to Sun., y: sum of check-in times)

We can easily make out from Fig.5 that check-in times is relatively more in the start and the end of the week, especially Saturday and Sunday.

Besides, it displayed different regulations of distribution in space.

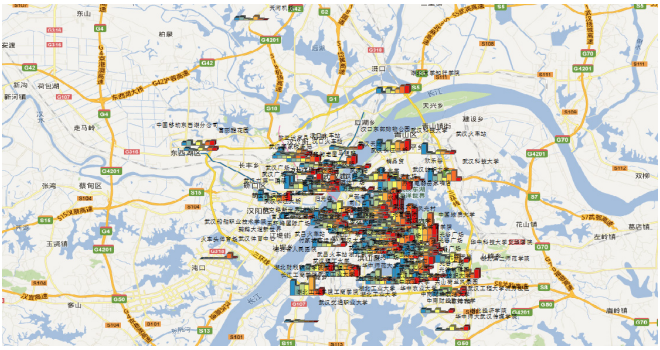


Fig. 6. Different regulations of distribution in space

3.1.2 Day/Night

A lots of points' check-in times show the obvious change from day to night, so we selected 6：00—18：00 as the day, also 18:00—24:00 and 0:00—6:00 as the night, and calculated respectively the proportion of check-in times in the day to the total quantity. If the ratio is over the 2/3, decided as day type; if the ratio is between 1/3 and 2/3, then decided as night type; or it's below 1/3, decided as day-night type.

Fig. 7. Distribution diagram of High-frequency check-in locations (day/night)

The study found that more business institutes are Day-type points, and more universities are Night-type, as listed in the following table:

3.1.3 6h

The study found it still too tough to use the granularity of the day/night for doing analysis in the check-in data. Therefore, we chose 6h as the interval, that is to say, divided a day into four periods(0:00—6:00,6:00—12:00,12:00—18:00,18:00—24:00), and study check-in records' spatial changing regulations in the different periods.

Table 1. Descending-order table of day-type check-in times

pid	sid	poi	kind	point_x	point_y	sum_checkin	day_night
89	43	AydHome Furnishings	Business	114.253	30.6144	60	day
54	70	New World Dept.Store,Avitation Road	Business	114.266	30.5794	87	day
148	74	Wuhan No.1 Hospital	Others	114.272	30.5723	39	day
128	52	Hanyang Marshland	Travel	114.283	30.5523	42	day
60	20	Yangtze River Bridge of Wuhan	Transport	114.288	30.5496	77	day
91	90	Hubei provincial people's hospital	Others	114.299	30.5352	58	day
141	57	Hubu Alley	Business/Travel	114.299	30.5478	39	day
50	78	Yanzhi Road	Business/Travel	114.309	30.5523	94	day
136	46	the 2nd Yangtze River Bridge of Wuhan	Transport	114.325	30.603	40	day
153	61	Fujiapo passenger station	Transport	114.327	30.5356	37	day
106	85	the CCNU's cultural street	Business	114.357	30.5144	50	day
41	22	Wuhan University	University	114.361	30.536	113	day
98	432	Huazhong Normal University	University	114.363	30.5166	55	day
193	100	Wuhan University	University	114.365	30.5366	29	day

Table 2. Descending-order table of night-type check-in times

pid	sid	poi	kind	point_x	point_y	sum_checkin	day_night
29	2146	China Mobile Dongxihu branch	Others	114.15	30.6289	139	night
195	80	Locomotive Stadium	Others	114.171	30.5001	29	night
90	1232	Wuhan State Taxation Bureau	Others	114.283	30.6043	58	night
59	149	Wanda Plaza	Business	114.292	30.5764	77	night
58	27	Port in Wuhan	Transport	114.301	30.5808	78	night
30	87	Wuhan University of Technology	University	114.344	30.5062	138	night
232	99	the FruitLake Children Park	Business/Travel	114.346	30.5459	24	night
45	35	Hubei Theatre	Others	114.354	30.5289	106	night
131	82	Wuhan University	University	114.365	30.53	41	night
56	75	Wuhan University	University	114.371	30.5371	81	night

- The Universities' check-in frequency is average in the four-6h periods a day.

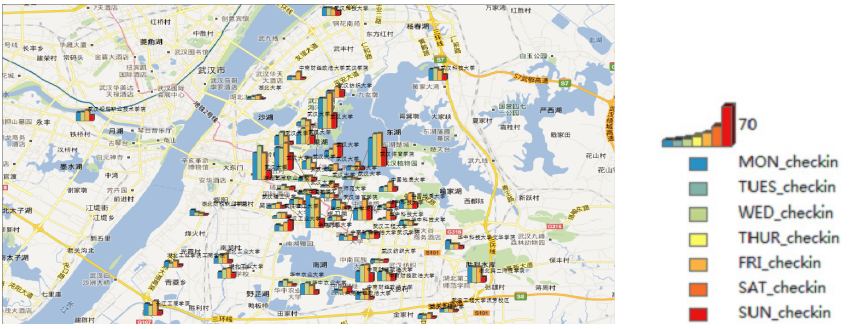


Fig. 8. The universities' check-in distribution (6h)

- The check-in peak of business locations is from 6 to 18h. While check-in times of other periods is less by contrast.

Fig. 9. The business locations' check-in distribution (6h)

- Tourist places' check-in records shows the similar spatio-temporal regulations with the check-in points of the business.

Fig. 10. The Tourist places' check-in distribution (6h)

Besides, to other types of check-in points or some places which has few amount of check-in data, the spatio-temporal regularities are not so strong.

3.1.4 Holidays

The holiday is a special period of time. During this period, people's activities have obvious differences with normal time. Depending on the holiday's fields of time dimension in the database, we can easily analyze and count the relevant check-in records in the holidays. We also found that check-in data gathered in December is really typical, so we set the weekend as an example in this study, and organize the check-in data on weekends and week-days in December by drilling-down and rolling-up.

We found that the most distinctive types are the business institutes and the Universities in the check-in records.

- Business institutes

Fig. 11. The Business institutes' check-in distribution (Holidays)

It's not difficult to find larger quantity and higher frequency of check-in than usual to those typical business institutes like Chicony Square.

- The Universities

Fig. 12. The Universities' check-in distribution (Holidays)

Unlike business institutions, check-in quantity of the University has less obvious differences between the weekends and the weekday.

From these experiments, we found these spatial regulations in different time granularities corresponded to the daily routines of check-in points, so as to demonstrate that using the technology of drilling is effective to analyze the different spatial regulations of check-in data in different time granularities.

3.2 Discussing the Long-Term Variations of the Different Check-in Points by Using the Technology of Slicing

Check-in data of different locations will display different change rules. When observing this regulation, we can use slicing based on the spatial data warehouse. That

is to say, fix the spatial dimension, and observe the changes of check-in data with time dimension. Because of the monthly growth of check-in data, for comparisons, we use this relative index named the trend parameter(TTP) during the analysis.

TTP of somewhere =
(check-in times of the month / city's total check-in times of the month)*10000

The indicator has been used to remove the error in the process of data acquisition, so we could compare the long-term variations more objectively between different places. The study found the check-in point's types such as the Universities, business institutes and transport agencies have different obvious cyclical law changes.

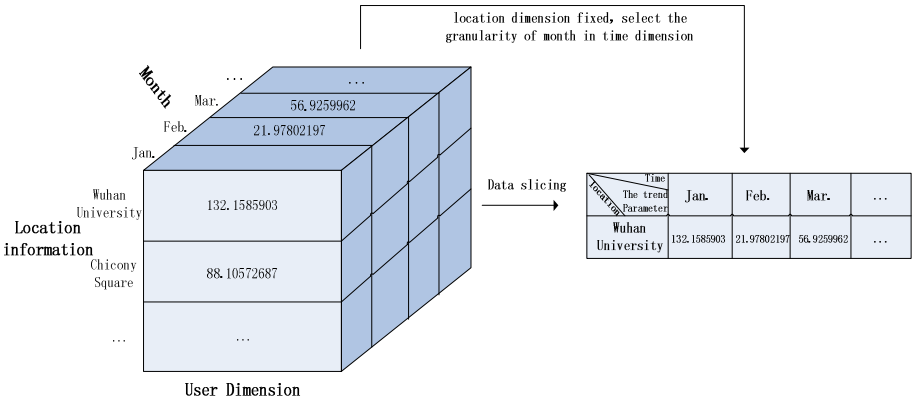


Fig. 13. Schematic diagram

As described in Fig.13, to the cube of check-in data, we firstly selected location dimension and user dimension, achieving data slicing. Then fixed the location dimension into a single check-in point, and chose the granularity of month to analyze, so as to study different check-in points' long-term variations over time.

The study found the long-term variations of the different check-in points.

3.2.1 The University

There will be a check-in peak after the summer to these points, especially in September when freshman enter into the universities with high-frequency activities organized by schools or communities. On the contrary, during the summer vacation and the winter vacation when lots of students will return home, the check-in trend parameters are relatively lower.

3.2.2 Business Institutes

Check-in peak of these points is about in Feb., July., or Nov.. Their distribution results is corresponding to periods like the summer vacation, the winter one and Spring festivals; the trough is about in Apr. and Sept., this period is also closely associated with the low and peak season of business institutes.

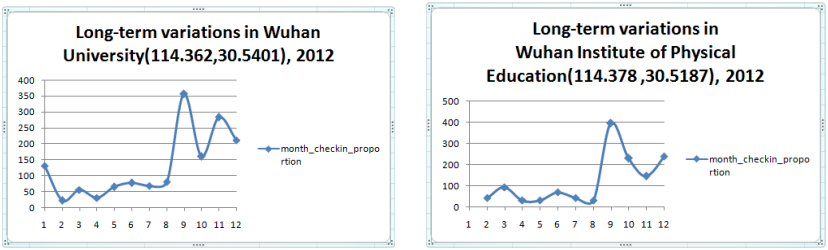


Fig. 14. Chart of long-term check-in variations of some universities

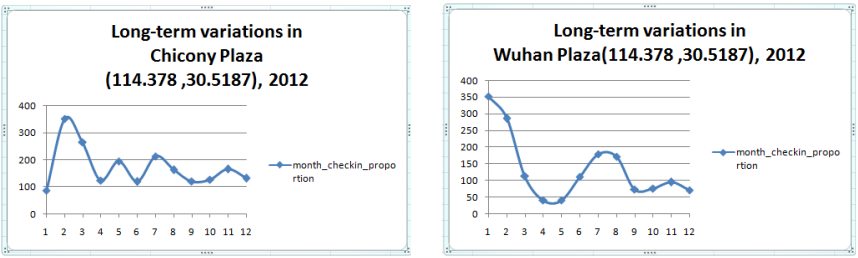


Fig. 15. Chart of long-term check-in variations of some business institutes

3.2.3 Transport Agencies

E.g., WuChang Railway Station, during summer and winter breaks, the check-in trend parameters is obviously higher and it's corresponding to the peak passenger flow of returning home or traveling.

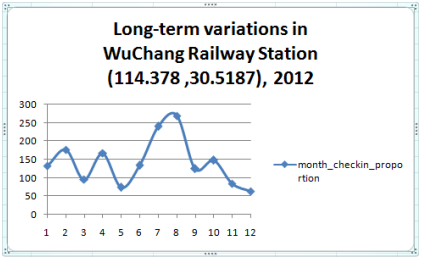


Fig. 16. Chart of long-term check-in variations of a transport agency

The experiment demonstrated that using the technology of slicing is effective to analyze the long-term check-in fluctuations of different places, and we found these long-term check-in regulations corresponded to the different check-in points' actual activity rules in a year.

4 Discussion of Results

Check-in data is a new kind of spatial data. In the research, we found it's an effective measure to analyze check-in data based on spatial data warehouse and technology of OLAP.

Spatial data warehouse is responsible for the storage of check-in data. It also monitors and manages the source data. Depending on the multilevel storage strategy and on-line analytical processing tools, spatial data warehouse can manage check-in data effectively, carry out the spatio-spatial analysis of check-in data reasonably, and make the research to proceed smoothly.

OLAP technology, as methods of analyzing multidimensional check-in data, provided effective data exploring by drilling, slicing, and pivoting. Based on this, we displayed obvious spatio-temporal variations of check-in data, coming to clear conclusions corresponding to regulations in our daily lives.

In the foreseeable future, we think application and development of spatial data warehouse technology will play a more important role in the analysis of webpage geographical contents, and we will do more exploration in the research of this respect.

Acknowledgments. This study is supported by Fundamental Research Funds for the Central Universities (No.216274012), Hubei Provincial of Education (No.B2013032 and No.13q100).

References

[1] Li, L., Goodchild, M.F., Xu, B.: Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. Cartography and Geographic Information Science 40(2), 61–77

[2] Hong, L., Ahmed, A., Gurumurthy, S., Smola, A., Tsioutsiouliklis, K.: Discovering Geographical Topics in the Twitter Stream. In: The Proceedings of the 21st International Conference on World Wide Web (WWW 2012), Lyon, France (April 2012)

[3] Banerjee, N., Chakraborty, D., Dasgupta, K., Joshi, A., Mittal, S., Nagar, S.: User interests in social media sites: an exploration with micro-blogs. In: CIKM 2009 Proceedings of the 18th ACM Conference on Information and Knowledge (2009)

[4] Cho, E., Myers, S.A., Leskovec, J.: Friendship and Mobility: User Movement in Location-Based Social Networks. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1082–1090. ACM (2011)

[5] Scellato, S., Noulas, A., Mascolo, C.: Exploiting place features in link prediction on location-based social networks. In: KDD 2011 Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1046–1054 (2011)

[6] Savery, L., Wan, T., Zeitouni, K.: Spatio-Temporal Data Warehouse Design for Human Activity. Pattern Analysis Database and Expert Systems Applications (2004)

[7] Rivest, S., Bédard, Y., Proulx, M.-J., Nadeau, M., Hubert, F., Pastor, J.: SOLAP technology: Merging business intelligence with geospatial technology for interactive spatio-temporal exploration and analysis of data. ISPRS Journal of Photogrammetry & Remote Sensing 60, 17–33 (2005)

[8] Di Martino, S., Bimonte, S., Bertolotto, M., Ferrucci, F.: Integrating Google Earth within OLAP Tools for Multidimensional Exploration and Analysis of Spatial Data. In: Filipe, J., Cordeiro, J. (eds.) ICEIS 2009. LNBIP, vol. 24, pp. 940–951. Springer, Heidelberg (2009)
[9] Papadias, D., Tao, Y., Kalnis, P., Zhang, J.: Indexing Spatio-Temporal Data Warehouses. In: Proceedings of the 18th International Conference on Data Engineering (2002)
[10] Andrienko, G., Andrienko, N.: Spatio-temporal Aggregation for Visual Analysis of Movements. In: IEEE Symposium on Visual Analytics Science and Technology, VAST 2008 (2008)

The Spatial Analysis of Weibo Check-in Data—— The Case Study of Wuhan

Mingye Bao, Nanhai Yang, Liang Zhou, Yizhen Lao, Yun Zhang, and Yangge Tian*

International school of software, Wuhan University, Hubei, 430079
tiandebox@whu.edu.cn

Abstract. With the popularization and development of mobile phones, more and more people share their spatial locations on social network, to leave their footprints. However, Studies in the patterns of the check-in data and its relation to the existing space are not enough. Using the method of the spatial analysis of the data direction distribution and hierarchical analysis, we found that the check-in data has the close contact with the real space. It is of great value for us to deeply explore spatial characteristics and extend the usage of check-in data.

Keywords: check-in, weibo, patterns.

1 Introduction

Check-in data is a kind of new web content with geographical information. As the smart phones with GPS function are becoming more and more popular, it is becoming simpler and simpler for people to locate themselves. People can locate their position through the mobile phone GPS or the Internet. Check-in data is the message that users use their mobile phone to publish on the Internet, and that is related to their current position. Now, softwares with the function of check-in message is being used, for instance, Foursquare, Jiepang, RenRen, Sina. At the same time, the positioning data will be shown on their homepage. Checking in on the Internet has formed a prevailing trend and the quantity of related data is growing very quickly. It is valuable for us to study whether there are certain space distribution patterns about the check-in data and which of the reality space factors do these patterns have connection with.

The check-in data has been used to study what is related with geographic spatial distribution model or the user behavior. Linna Lia establishes Spatial, temporal, and socioeconomic patterns by collecting data from Twitter and Flickr (Linna Lia*, Michael F. Goodchilda and Bo Xub). Lee uses twitter and other data sources for analysis. For example, he predicts election results according to the sentiment of user's comments on twitter and use user's comments to automatically monitor local events(Lee and Sumiya 2010); Li and Goodchild describe the locations of sites and

* Corresponding author.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 480–491, 2013.

space of the border through the results from the analysis of aggregation data with large amounts of location label(Hollenstein and Purves 2010; Li and Goodchild 2012); Cheng studies the moving model of the crowd though the data collected from position shared service(Cheng et al. 2011); Li and Goodchild study the patterns of geographical landscape and human activities though these data(Li and Goodchild, 2013). However, these studies did not combine check-in data with real spatial analysis in depth and did not pay close attention to the relationship between the space mode of check-in data and real space. Therefore, this paper takes Wuhan as the study area, by fetching check-in data of Wuhan to analysis the relationship between the city's environment and the places which check-in data gathers in, in order to investigate the causes of check-in data gathering in these sites.

2 Acquisition and Processing of Sina Check-in Data

We selected the check-in data as research object from the most popular Chinese social media——Sina. By the end of 2012, Sina has more than 503 million users. Sina provides more convenient API to help developers get data resource. By applying for an application development account on the web page and getting an "access_token" and authentication, you can call the API to get the data. Sina provides multiple API for us to get Sina weibo's check-in data. We can get the data based on each location's POIID. Through this interface, we can get all the published weibo of the site. As acquiring all records of the site though related data of each location POIID is too intricate and technically difficult to achieve, we choose the method of multi-point data acquisition to get the users check- in data according to the center of latitude and longitude and adjusting the radius.

The check-in data collected from Sina Weibo has different kinds of positioning accuracy because of different data sources, just like the smart phone's GPS positioning, the relative position of the user in the cellular location or users' selection in the map. As Sina open platform interface has number limitation (150 times per hour), we get data in multiple random. In order to avoid data loss, our circles of multi-points will be repeated from one to another. This leads to a small amount of duplicate data. Therefore, we must remove duplicate, then keep our data in database. By setting the time quantum of collecting data, through a month-long collection, we did not finish the job until the new historical data converged to zero. So we hold the view that the data related with that time quantum had been collected entirely.

We altogether collected 260843 of check-in data, from January 2012 to January 2013.186375 of them are different in latitude and longitude, 120764 of them are published by different user, 2.15994 per person. As showed in Table 1:

Table 1. The Top 10 points of check-in data

x	y	shape	count
30.44299	114.30962	point	51
30.45163	114.437279	point	51
30.452551	114.270974	point	49
30.45163	114.43728	point	34
30.445291	114.426354	point	30
30.45209	114.39051	point	16
30.45013	114.42921	point	15
30.451754	114.440559	point	10
30.4538	114.265801	point	10
30.44853	114.385971	point	9

3 Spatial Analysis of Wuhan's Check-in Data

We cover the figure of the points on the Google map. We find that data are mainly distributed along the road. The clear outlines are formed by the data points along the road.

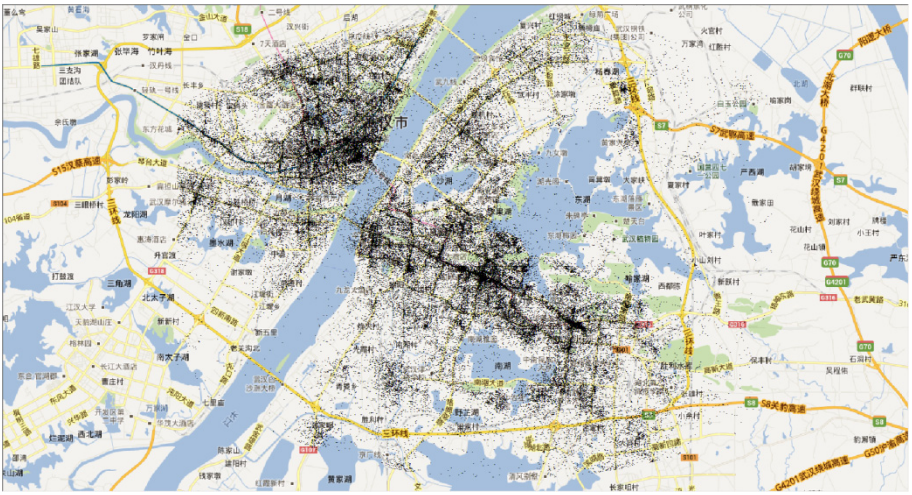


Fig. 1. The overall distribution map

In order to reflect the distribution patterns of hot check-in points on the space better, we will use these points to set point density statistical calculation. Point density statistical calculations is based on point elements that fall into neighborhood of every unit, in order to calculated the scale of per unit area level= $\frac{count}{area}$. Processing result is shown in Fig 2.

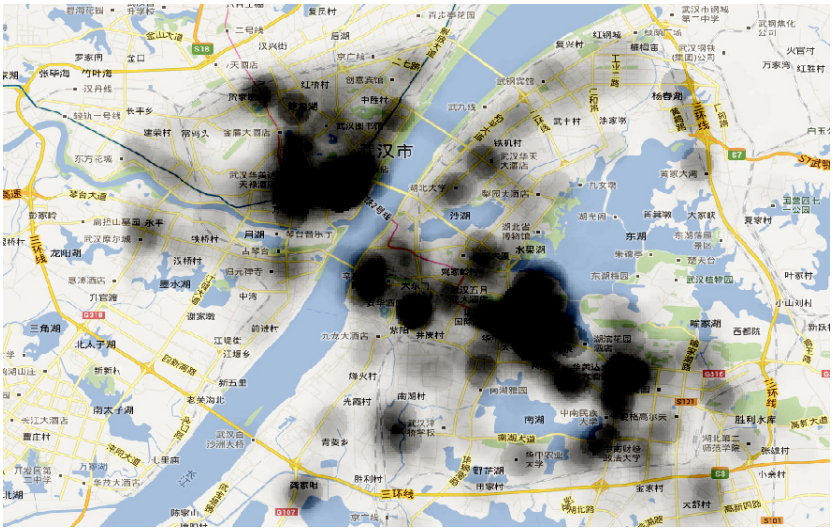


Fig. 2. The overall density map

Kriging difference is a kind of commonly used interpolation method. It can be used to plane interpolation based on the number of check-in people and analyzing the distribution of the whole. Use $Z(x_0)=\sum_{i=1}^{n}\lambda_i Z(x_i)$ to calculate. $Z(x_0)$ is the value of the unknown sample. $Z(x_i)$is the value of the known sample points around the unknown sample. λ_i is the weight of the ith known sample point of the unknown the sample point. n is the number of the known sample points. After processing, the check-in results of Wuhan are below:

In order to further evaluate the statistical pattern of the data, we have to use the method of half the variation index analysis to analyze check-in data which is the more than 10. Semivariable function which is also called semi-variation function or semivariable matrix is a specific function of geostatistical analysis. The semivariable function is $r(x,h)=\frac{1}{2}E[Z(x)-Z(x+h)]^2$. After calculating, we get the semivariable image is below:

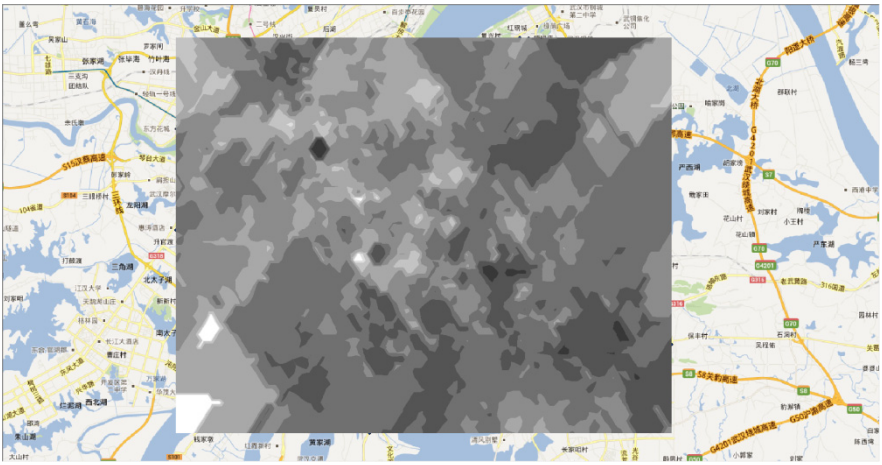


Fig. 3. Kriging interpolation results

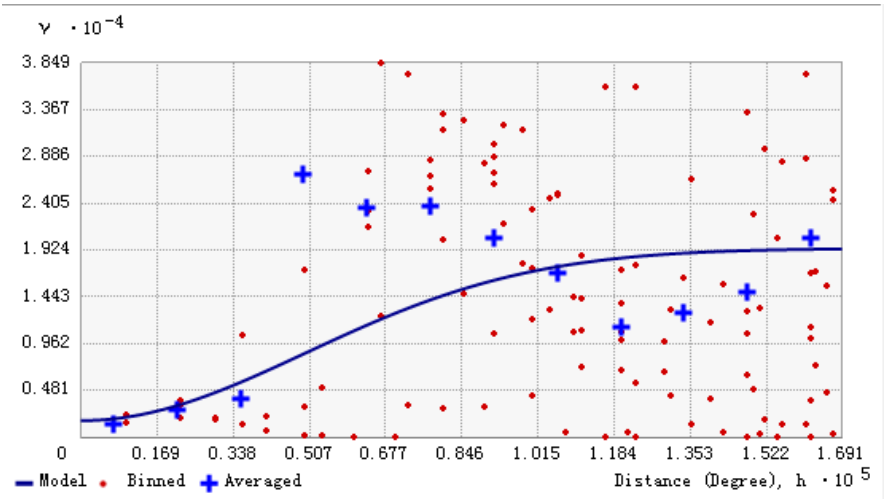


Fig. 4. Variation index check-in of times greater than10

Through the analysis of figure, we found that when the transverse reaching 1.522.The degree of correlation is regional stability. We will know that the check-in location of people presents objective patterns in the space.

4 Hierarchical Analysis of Check-in Data

Count data in the database, and we will get data with check-in times greater than 10 or 20. We export the data to the Google map and get the following three maps about check-in place distribution. At the same time, in order to better analyze the pattern of distribution

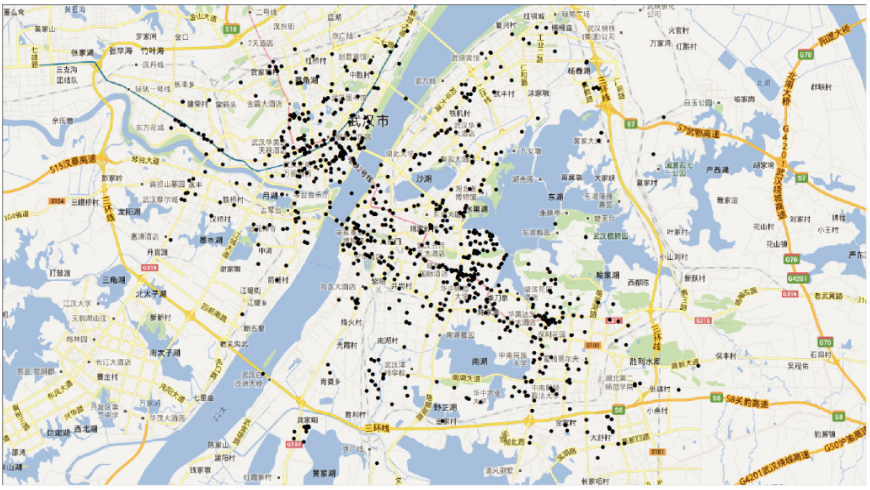


Fig. 5. Check-in times greater than or equal to 10

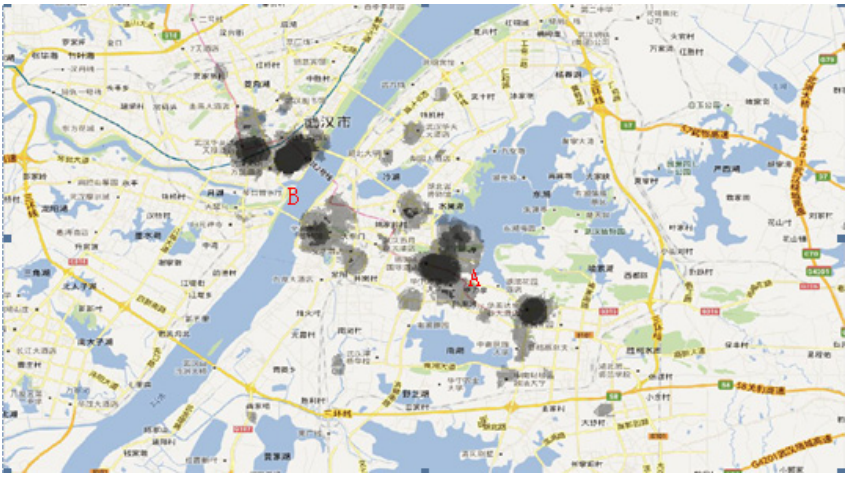


Fig. 6. Heat map of check-in times greater than 10

which is about high frequency check-in region, we will use heat map to represent it. Making statistics and conversion about the heat map with data which has check-in number greater than 10 or 20 times, we get the following two heat distribution maps.

In the figure, check-in times are greater than or equal to 10, the distribution of points covers most areas within the 3rd ring road in Wuhan. But it still present that the majority of distributions are Wuchang and Hankou and the points are density concentration.

In the distribution map with check-in points is over than 20 times, the distribution of points still has widely coverage. Some points are removed which are distributed at the edge of the part. The centers of the centralized distributions are not changed, still in Wuchang and Hankou.

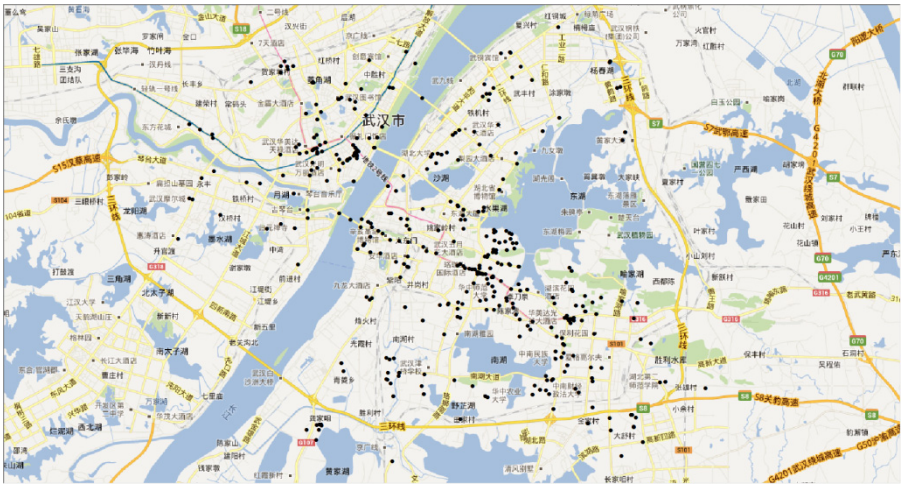


Fig. 7. Check-in number greater than or equal to 20

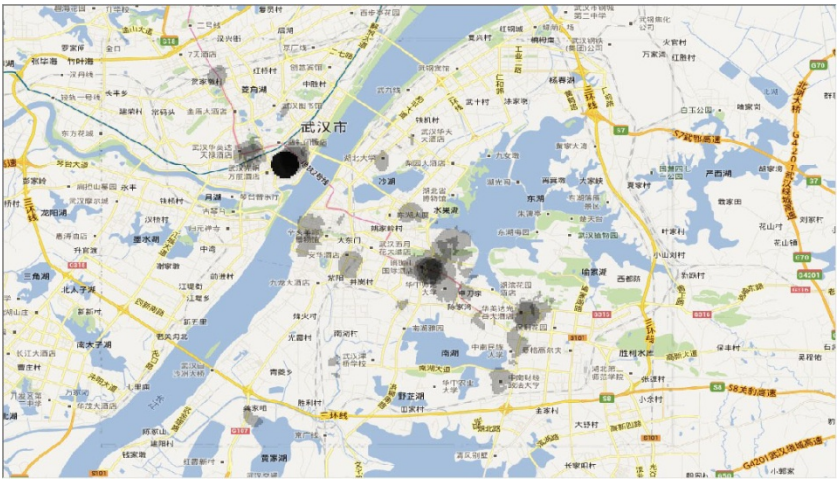


Fig. 8. Heat map of check-in times greater than 20

Comparatively analyzing Fig 3 and Fig 4, we find the main distribution center is a constant. Distribution of data presents certain stability. In combination with Google street map, positioning the center of density, we will find that the majority of points are distributed in the places surrounding the university, university itself, sightseeingplace, business center, and railway station. For the campus, check-in information is often something about where students take lessons or the self-study and where they may see something special. This is the reason why population floating in university is not big, but the number of check-in times is huge. As for sightseeing place, businesscenter and railwaystation, these places are the areas of dense population and have a large population flow.

Table 2. Statistical results of check-in data of top 50

rank	poiName	kind	s_sum
1	Chicony Plaza	Business	194
2	Wuchang Railway Station	Traffic	192
3	Wuhan University	University	175
4	Wuhan University	University	153
5	Wuhan Plaza	University	137
6	Wanda Plaza	Business	112
7	Wuhan Tiandi	Tourist Attraction	108
8	Wuhan Railway Station	Traffic	94
9	Wuhan University	University	89
10	Wuhan Sports University	University	87
11	Lingjiao Block Walmart	Business	86
12	Central China Normal University	University	86
13	Wanda Plaza	Business	86
14	Xudong Walmart	Business	80
15	Wuhan Plaza	Business	76
16	Asia Trade Plaza	Business	68
17	Optics Valley Plaza	Business	67
18	Tianhe Airport	Traffic	67
19	Optics Valley Plaza	Business	65
20	Yangtze River Bridge of Wuhan	Traffic	64
21	Hankou Railway Station	Traffic	64
22	Wuhan University	University	62
23	Chicony Plaza	Business	59
24	Happy Valley	Tourist Attraction	58
25	Wuhan University	University	55
26	Jianghan Road	Tourist Attraction	51
27	Port in Wuhan	Traffic	50
28	Han Street	Tourist Attraction	49
29	GUANSHAN HENANFENGQING park	Tourist Attraction	49
30	Wuhan University	University	48
31	Wanda Plaza	Business	47
32	Wanda Plaza	Business	45
33	Zuobiaocheng Business Center	Business	44
34	New World Dept.Store of Wuchang	Business	43
35	Hubei theatre	Other	43
36	Optics Valley Plaza	Business	42
37	Hubei university of technology	University	41
38	huazhong university of science and technology	University	40
39	Wuhan University	University	40
40	zhongnan university of economics and law	University	39
41	Jiqing street	Business	37
42	Public Park	Business	37
43	AydHome Furnishings of Hankou	Business	37
44	Wuhan University	University	36
45	Xiaopinmao Business Center	Business	36
46	the 2nd Yangtze River Bridge of Wuhan	Traffic	34
47	China University of Geoscience	University	34
48	New World Dept.Store of Hanyang	Business	33
49	Zhongnan University of Economics and Law	University	33
50	Wuhan University	University	33

In the picture, as for point A and B, through the combination with the Google map of streets, we find that the center of A is the Wuhan University. It is the hot check-in spot of Wuhan universities. And the center of the point B is market. It is the hot check-in spot of tourism in Wuhan. We sampling the age of users, finding that the younger generation is 65.53%, because they are the people who like contacting new things and loving travel. As a school which has a more than 100 years of history and having a plenty of creative students, Wuhan University is a hot check-in spot. Around scenic area of East Lake, it is in addition to check-in times of the Wuhan University.

Through statistics on site, we get the site of the top 50.

Through statistics histogram of the data of the top 100, we get the statistics figure as below: university 34%, business32%, both of them 68%, close to 70%.

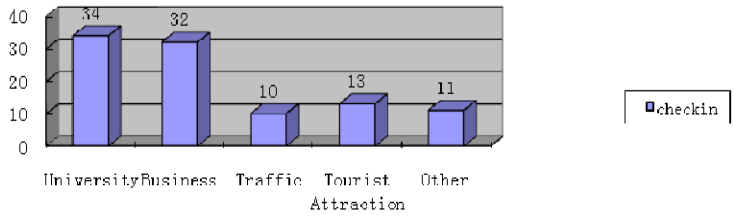


Fig. 9. Histogram Statistics of check-in data of top 100

5 The Partitioning Feature of Check-in Data about Wuhan

Wuhan's three towns are separated by lakes and rivers, so they are different from one another. At the same time, any town of the three has its special characteristics. Hankou is the commercial financial center. Wuchang is the place of technology and education. Hanyang is developed as exhibition leisure center. Different urban function orientation and regional culture make their characteristics diverse. And from the point of view about regional distribution, the characteristics of population distribution of Wuhan's three towns is zonal distribution along the Yangtze River and Hanjiang River. Along the both sides of the Yangtze River or Hanjiang River, density is the largest. The farther away from the river bank, the smaller density is. At the same time taking old city wall as the center, data is distributed into the surroundings in the form of concentric circles. The closer to the center, the greater the density is. In the same area the differences about average density exist. Due to these characteristics, so we discuss the partition of check-in data is necessary.

The number of distribution points of Wuchang is relatively larger, mainly distributed in and around Wuhan University, and the distribution of points of the Optical Valley and the Wuchang railway station is relatively dense.

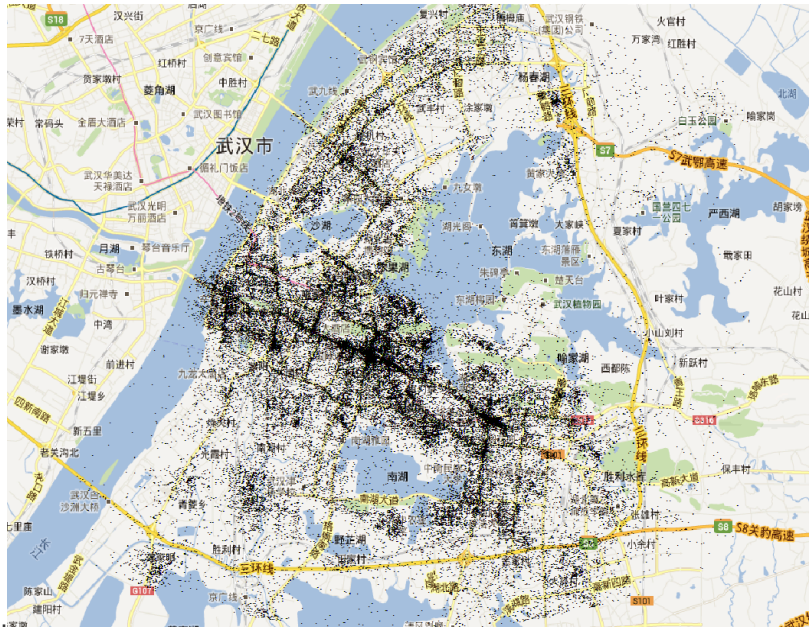


Fig. 10. Wuchang area

In the Hankou area, the points are mainly distributed in the Zhongshan Park and HanzhengStreet.

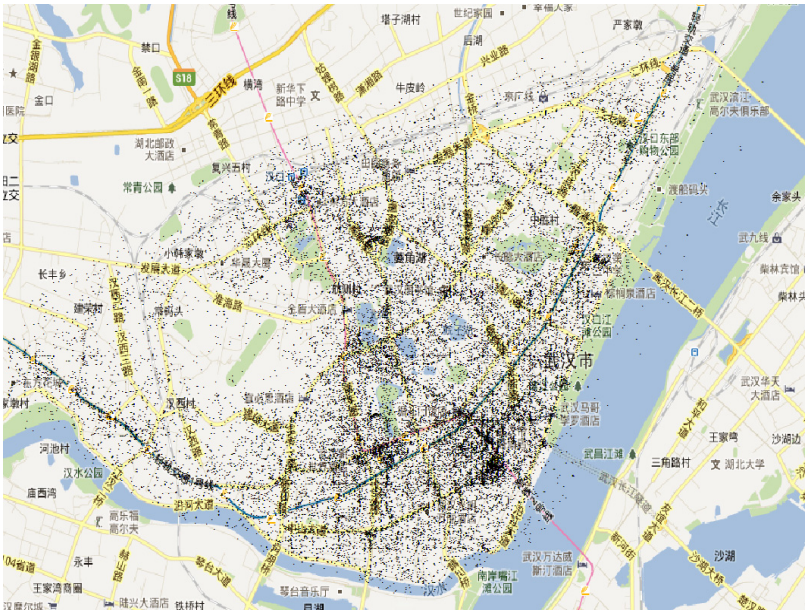


Fig. 11. Hankou

In the Hanyang District area, the points are mainly distributed in the Guiyuan Temple, GuQinTai and MoerCheng Mall.

Fig. 12. Hanyang

Through data comparison of three places, we found that the number of dense points in the Wuchang is more than the others because universities in the Wuchang are more and relatively widely distributed. However, Hankou is mainly for business. Especially in the Hanzheng Street, it is one of the most prosperous commercial centers in Wuhan. At the same time, Zhongshan Park is one of the famous scenic spot in Wuhan. As for the Hanyang, economic development is relatively backward, besides it has no universities getting together, so the number of check-in points is less. The check-in points primarily in Qingchuan pavilion and near bridge of the YangtzeRiver, but the distribution of check-in points is relatively sparse.

6 Conclusion

Our study shows the check-in data distribution has a close connection with real spatial pattern, which can provide help to other corresponding studies. Some points' high check-in frequency suggests we should give a deep insight into their urban function and population flow. Through dividing the check-in data according to region, we can see the check-in region patterns have a close connection with the region functions and the different region characteristics which have a deep impact on the check-in data. So the check-in data can support business and city construction planning for the enterprise and the government.

Though the studies of check-in data which have their value in theory and utility, there are still many problems needing to be solved, such as the influence factors of check-in frequency. For example, Wuhan University and HUST are well-known universities in China, but the differences of check-in data's characteristics between these universities are very large, and the reason needs a further research. In the future, we will make more studies on the reason of check-in diversity.

Acknowledgments. This study is supported by Fundamental Research Funds for the Central Universities (No.216274012), Hubei Provincial of Education (No.B2013032 and No.13q100).

References

1. Li, L., Goodchild, M.F., Xu, B.: Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. Cartography and Geographic Information Science 40(2), 61–77 (2013)
2. Lee, R., Sumiya, K.: Measuring\Geographical Regularities of Crowd Behaviors for Twitter-Based Geo- Social Event Detection. In: Proceedings of the 2nd ACMSIGSPATIAL International Workshop on Location Based Social Networks (LBSN 2010), pp. 1–10. ACM, New York (2010)
3. Hollenstein, L., Purves, R.: Exploring Place Through User-Generated Content: Using Flickr to Describe City Cores. Journal of Spatial Information Science 1(1), 21–48 (2010)
4. Cheng, Z., Caverlee, J., Lee, K., Sui, D.Z.: Exploring Millions of Footprints in Location Sharing Services. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM), Barcelona, pp. 81–88. AAAI Press, Palo Alto (2011)
5. Li, L., Goodchild, M.F.: Spatio-Temporal Footprints in Social Networks. In: Alhajj, R.S., Rokne, J.G. (eds.) Encyclopedia of Social Networks and Mining. Springer (2013)
6. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake Shakes Twitter Users: Real-Time Event Detection by Social Sensors. In: Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, pp. 851–860. ACM, New York (2010)
7. Silverman, B.W.: Density Estimation for Statistics and Data Analysis. Chapman and Hall, London (1986)
8. Soule, L.C., Shell, L.W., Kleen, B.A.: Exploring Internet Addiction: Demographic Characteristics and Stereotypes of Heavy Internet Users. Journal of Computer Information Systems 44(1), 64–73 (2003)
9. Taylor, W.J., Zhu, G.X., Dekkers, J., Marshall, S.: Socio-Economic Factors Affecting Home Internet Usage Patterns in Central Queensland. Informing Science 6, 233–246 (2003)
10. Tumasjan, A., Sprenger, T.O., Sandner, P.G., Welpe, I.M.: Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. In: Fourth International AAAI Conference on Weblogs and Social Media, Washington, DC, May 23-26 (2010)
11. Wold, H.: Estimation of Principal Components and Related Models by Iterative Least Squares. In: Krishnaiaah, P.R. (ed.) Multivariate Analysis, pp. 391–420. Academic Press, New York (1966)
12. Wold, S., Sjöström, M., Eriksson, L.: PLS-Regression: A Basic Tool of Chemometrics. Chemometrics and Intelligent Laboratory Systems 58, 109–130 (2001)
13. Zandbergen, P.A.: Accuracy of iPhone Locations: A Comparison of Assisted GPS, WiFi and Cellular Positioning. Transactions in GIS 13(s1), 5–25 (2009)

Application and Research of Subway Station Modeling System Based on VRML*

Lin Zhang

School of Electricity and Information Engineering
BUCEA
Beijing, 100044, China
zzcmeng0806@sina.com

Abstract. Virtual reality is a integrated information technology, which combines digital image processing, computer graphics, multimedia technology and sensor technology. It has a broad prospect for application in the fields of architecture, tourism, military and computer game design. This paper introduces the constituent of VR, VRML work principle and characteristics, gives a detailed analysis of the key technology of the subway station modeling system with the Huangzhuang station as the background.

Keywords: Modeling System, VRML, Subway Station.

1 Introduction

Virtual reality technology is the term used to describe a three-dimensional, computer generated environment which can be explored and interacted by a person. That person becomes part of this virtual world or is immersed within this environment and whilst there, is able to manipulate objects or perform a series of actions.

There are three characteristics: immersion, interaction and imagination. We also call them "3I"[1].

1. Immersion: when users enter into the vitual environment by all kinds of advanced sensors, they think all of the things they see, hear and feel are real and they believe they are in the real environment, that is called "immersion".

2. Interaction: at the same time, all users can interact with many objects in the vitual environment. They can operate the machine and they can lift the box, and so on.

3. Imagination: users can have many association with the immersion and interaction in the vitual environment.

* This work is supported by 2012 science and technology project of Beijing Municipal Commission of Education(the number is 051200315) and Funding Project for Academic Human Resources Development in Institutions of Higher Learning Under the Jurisdiction of Beijing Municipality(the number is PHR201108208).

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 492–498, 2013.

This essay, with the construction of Huangzhuang subway station as the background, introduces the operating principle of Virtual Reality Modeling Language (VRML), analyzes the key technology and the program of realization of the subway station of the virtual roaming system in detail, and realizes the virtual roaming system.

2 Vrml Overview

2.1 The Constituent of VR

Virtual Reality System is a new and integrated technology which involves computer graphics, human-computer interaction technology, sense technology and artificial intelligence. It can create some living three-dimensional scenes and people can operate the objects with some devices as if they are in the real scenes.The system consists of five parts:VR engine,I/O devices,software and database,users and tasks.The system structure is shown in Fig 1[2].

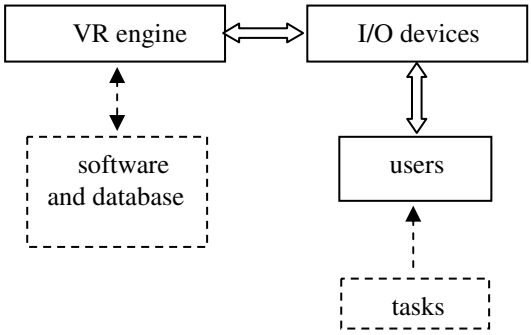


Fig. 1. VR system structure

1. VR engine is the key part in VR.It reads data from the input device,visits the database related to the tasks,excutes the real-time computation,updates the states of virtual world and outputs the feedback information to the output device.

2. I/O devices include input devices and output devices.There are many kinds of input devices such as tracker and roam operation interface.Tracker has many performance parameter:accuracy,jitter,drift and latency.Tracker includes mechanical tracker, electromagnetic tracker, ultrasonic tracker and optical tracker.Roam operation interface includes tracker ball and three-dimensional probe.There are three kinds of output devices in VR:graphic display devices, voice display devices and tactile feedback.

3. Software and database means modeling which includes geometric modeling, physical modeling, movement modeling, intelligent behaviour, model management and I/O mapping.

4. Users refer to human factor.It includes usability study,user performance study,user security study and VR Study of Sociology.

5. Tasks is the application of VR.This new technology is used in many fields such as architecture,medical treatment,education,art,entertainment military affairs and so on.

2.2 VRML Work Principle

VRML (Virtual Reality Modeling Language) is the second-generation web language following HTML, which used for describing the three-dimensional objects and their behavior.It describes the three-dimensional scene with text messages, and transmits it through the Internet, and then it is explained by the VRML browser to generate three-dimensional scenes in the local machine. Such a mechanism can avoid transmitting graphics files directly on the Internet, and leave the complex task to the local machine, thus the burden of the Internet is reduced, and make the three-dimensional interaction possible [3].

The visiting method of VRML is based on the client / server model, while the latter provides VRML with file and support resources, customers could download file they want to visit through the Internet, and could interactively visit the Virtual realm created by it through the VRML browser on the local platform, thus the independence from the hardware platform could be achieved.

2.3 VRML Characteristics

The basic features of VRML includes several aspects as follow [4]:

1. Platform independence

VRML access mode is based on Client/Server. Server offers VRML file (*. wrl), and the Client could download it through the Internet, and runs it through local VMRL browser, then the three-dimensional virtual reality scenes will be generated.

2. High speed transmission

VRML describes links with ASCII text which can be used on each common platform. The amount of recordable data is reduced and can be transferred on low bandwith network. VRML graphics rendering is “real-time”. When browsing the scene generated by VRML document, you can control the direction of browsing, be free to move, rotate and zoom models in the scene with the mouse, without any restrictions whatsoever.

3. Intelligence

It’s mainly reflected in the time sensor node, animation interpolator node, touch sensing node and perception node which VRML uses to perceive the dynamic interaction between users and the model, and then produces a predetermined response according to the set of the script node.

4. Extendibility

VRML is only a standard and can not satisfy all kinds of needs. Some applications need user interaction, some need high quality screen and some need complex virtual reality world. These applications sometimes restrict one another. So VRML applies extendibility to allow users to define objects and properties and execute them by browser with Java.

3 Modeling Huangzhuang Railway Station Based on VRML

3.1 System Analysis

1. Development environment

We have chosen a relatively better integrated development environment VRMLPad in this system. It could not only support the grammar tips, syntax coloring, bookmarks and other settings, but also the structure of scene graph tree, routing plans and resources graph and so on, thus could quickly query and orienting; in addition, it also provides a preview function, in which you can immediately see the effect of 3D scene rendering. It reduces the quantity of input process, and greatly enhanced the efficiency of the development.

2. VRML file format

VRML file has four elements: VRML file head, model, event and route. Not all of the file must include all of the elements, but the VRML file head is necessary[5].

(1)VRML file head: The standard file head is #VRML 2.0 V2.0 utf8 in VRML 2.0.It means this file is a VRML file and it uses utf8 character set.

(2)VRML model: scene consists of models and model consists of nodes.

(3)VRML event and route: event includes event value and time stamp and route is the channel which connects the event export of some nodes and the entrance of the other nodes.

There are 54 nodes in VRML 2.0:

1) Shape node: It is used to indicate all kinds of geometries, wireframes and Plane diagrams.

2) Attribute node: It is used to define the colour, material quality and grain of some objects.

3) Group node: It can divide nodes into groups and combined the interrelated nodes into one object.

4)Sensor nodes: It apperceives user's input operation and triggers the corresponding actions.

5)Other node: Such as script node, hyperlink node and so on.

3. Development process

Three-dimensional modeling should be technically divided into three steps:

Step 1: geometric modeling. Geometric modeling is the traditional method in computer graphics. First we need to build geometric model, after using hardware and rendering algorithm to complete the blanking, illumination and projection, the three-dimensional geometric model will be bulit.

Geometric modeling can be further divided into hiberarchy modeling and category modeling. Hiberarchy modeling uses tree structure to indicate the various components of the object. Category modeling allow the same kind of objects share the same owner, and the owner contains detailed structures of such kind of objects. The geometric modeling of the system is a combination of these two kinds of modeling method.

Step 2:physical modeling. In physical modeling of objects, the important technology is collision detection. When objectes are moving in a virtual space, they may collide or touch with each other. So the system must detect the collision. The principle of collision detection technology is that if two completely sealed polyhedrons come into collision, at least one surface of one polyhedron intersects with at least one surface of another one. If the intersection could be detected as soon as it occurred, and rearrange the position of the two objects, the collision will be eliminated.

Step 3: behavioral modeling. Behavioral modeling is the most difficult technology in all of modeling because the user can not control all of the interaction. So in modern VR system, one method is proposed which can model the behaviour by visiting external sensor objects. These objects have nothing to do with the user's action and it is intelligent.

VRML defines several basic types of sensors:

(1) Time sensor: it is used to control the animation.
(2) Touch sensor: it is used to detect the collision.
(3) Perception sensor: it is used to shape perceptions of users.

3.2 System Design

The designing steps include[6]:

1. Draft drawing

Draft can describe the location of Huangzhuang Railway Station and ascertain the coordinate of master surfacing, which can be used to build the subway model.

2.Basic object building

Applying the draft data can build the basic plane, such as wall and ground. But all these objects creating at the same time will overlap, we must adjust the objects' coordinate system to the predetermined location to form the basic model.

3. Object grouping and adjusting

Group the analogical objects and form some different object groups. Then adjust the coupled objects so that we can edit them.

4. Chartlet adding

Add chartlet and light to the object to form the final model.

Huangzhuang Railway Station is a transfer station which connetcts line 4 with line 10. It includes two layers. The under white cuboid is the platform of line 4 and the upper one is the platform of line 10. We call them platform A and platform B. The two black cuboids above the platform A are the entrances of line 4 and the two black cuboids under the platform B are the entrances of line 10. Since all of the inner structures are symmetrical, we named the entrances A、B、C and D. The structure is shown in Fig 2.

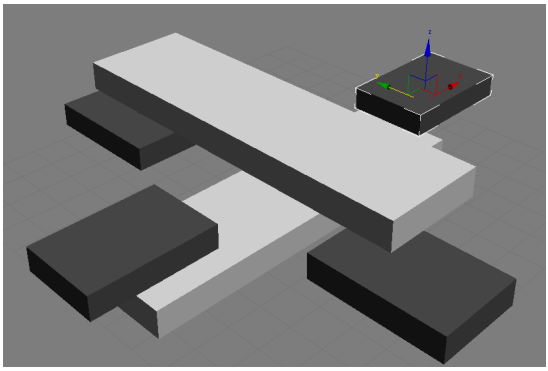


Fig. 2. The structure of Huangzhuang railway station

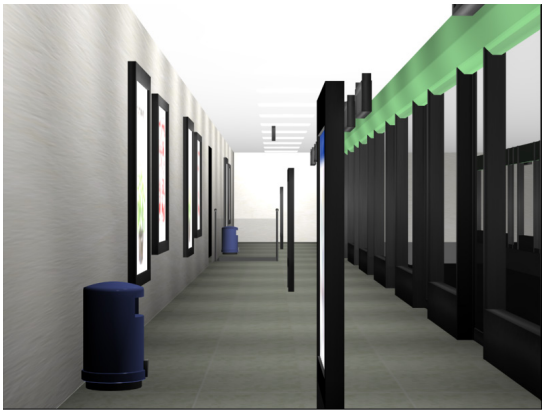


Fig. 3. Impression drawing 1

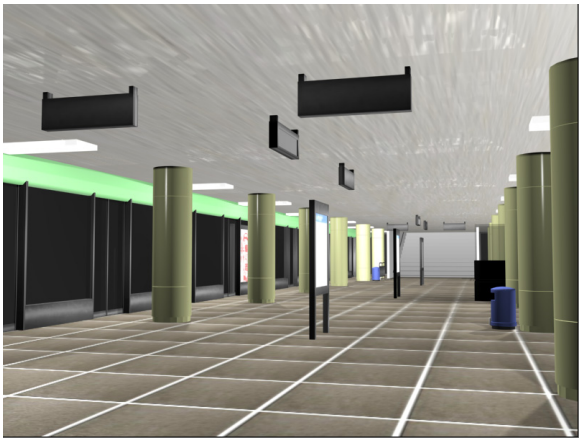


Fig. 4. Impression drawing 2

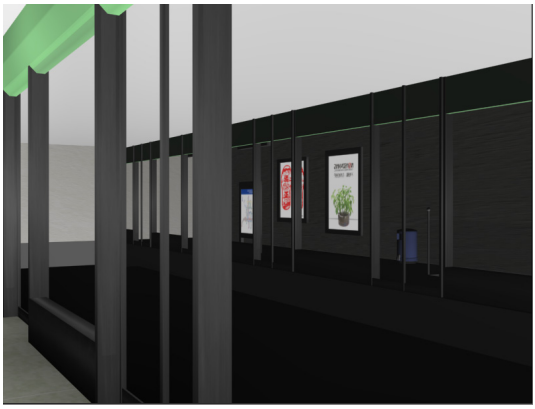


Fig. 5. Impression drawing 3

Render the model can make them have more sense of reality. In this project, we use light tracker to render the models. Light tracker is a advanced renderer and it is a kind of global illumination system. It uses some ray tracing technique to sample some points in the scene and calculate the reflection of light. It can make the illumination more real.From Fig 3 to Fig 5 are the impression drawings.

4 Summary

Three-dimensional modeling is the basis of VR system.It can get the data of real environment and build the virtual scene according to the actual needs. Three-dimensional modeling includes geometric modeling, kinematic modeling, physical modeling and behavioral modeling.VRML provides a better method to build lifelike three-dimensional scenes, many institutions have begun to use this technology for producing building.But the program still has large room for development because of its enormous quantity and technical limitations.This essay, with Huangzhuan subway station as the background, introduces the key technologies of virtual modeling and roaming system. It has a great significance to the study of architecture planning, designing and management.

References

1. Zhu, S.-X., Chen, L.-C.: Scene Modeling and Mode Optimization Technology in Virtual Reality. Computer Development and Application (18), 35–36 (2005)
2. Burdea, G.C., Coiffet, P.: Virtual Reality Technology. Publishing House of Electronics Industry
3. Shun, N.: The investigation of VRML principle and application. Journal of Inner Mongolia University for Nationalities (15), 17–18 (2009)
4. He, L.-K., Xu, Y.: Introduction and application of Virtual Reality Modeling Language. Journal of Hangzhou Teachers College (Natural Science Edition) (4), 157–160 (2005)
5. Ling, L.: The research of VRML modelling and animation based on VRML. Information Science, 55–56 (2009)
6. Yu, M.-H., Xie, L.: Modeling for virtual environment based on VRML. Software Space Time (26), 227–229 (2010)

Comprehensive Evaluation of Cultivated Land Quality Based on GIS in Tumote Right Banner of Inner Mongolia Autonomous Region

Ruiping Zhou, Yanru Wu, Chunxing Hai, Xiaojia Li, and Dandan Zhou

College of Geographical Science, Inner Mongolia Normal University.No.81, Zhaowuda Road, Hohhot, Inner Mongolia, P.R.C.
rpzhou@126.com

Abstract. It is one of important means to improve the cultivated land protection and management that the quality of cultivated land is evaluated synthetically in a county. The factors which influenced the quality of cultivated land were evaluated respectively in Tumote Right Banner of Inner Mongolia Autonomous Region including soil nutrient status, physical and chemical properties of cultivated land, and site conditions. Evaluation units were produced by overlaying land use map and the soil map. By Delphi approach and cluster the evaluation factors were chosen and their weights were determined by AHP method. Cultivated land quality was divided into four grades by "natural breaks" in frequency distribution. This result showed as follow. Firstly, soil nutrient contents were higher in Tumote Right Banner than that in other adjacent Banners. 71.77 per cent of cultivated land was the first and second grades. Physical and chemical properties of soil and biotope were suitable for crops. The first grade was 59.90 per cent of the cultivated land. The second grade accounted for 94.22 per cent. In general, the quality level of cultivated land was high. The first grade, the second grade, the third grade and the fourth grade cultivated land area was seperately 8.17 per cent, 44.00 per cent, 43.48 per cent and 4.36 per cent, which was consistent with the actual outputs.

Keywords: GIS, cultivated land quality, Comprehensive Evaluation, Tumote Right Banner.

1 Introduction

The various properties of cultivated land can be reflected by its quality[1-5]. The quality evaluation for cultivated land has had a significant development up to the present day[6-8]. On the whole, cultivated land resources evaluation has been developed from the past emphasis on the quantity of cultivated land, per unit area yield, fertility and basic natural state investigation to the comprehensive study for attention to cultivated land quantity, quality and ecology[9-12]. In order to provide basic decision support for the protection and management of cultivated land, the means of GIS and the existing model of comprehensive evaluation factor system were

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 499–508, 2013.

used to evaluate synthetically the quality of cultivated land in Tumote Right Banner of Inner Mongolia Autonomous Region.

Landform in Tumote Right Banner of Inner Mongolia Autonomous Region was composed of three parts which were northern mountains, central piedmont alluvial plain and alluvial plain of southern the Yellow River. The terrain of Tumote Right Banner tilts slowly from north to south. The soil is quite fertile. The altitude of the northern mountain is 1300 ~ 2338m. The altitude of piedmont alluvial plain cut by Yellow River is 1000 ~ 1050m, and that of southern alluvial plain of the Yellow River is 988 ~ 1000m. The annual average temperature is 6.9℃, and the accumulated temperature which is higher than 10℃ is 3053℃.The annual average precipitation is 359mm. The annual average relative humidity is 53%. Transiting the Yellow River is the main source of water for irrigation, and the average annual runoff is 24.76 billion stere. Tumote Right Banner, administer area 2323.64 square kilometer, has jurisdiction over 5 towns, 3 townships, 1 administrative committee, which belongs to Baotou city of Inner Mongolia Autonomous Region. The cultivated land is 111172.15 hectare. With the rapid development of social economy, great changes have taken place of land use patterns and soil property in Tumote Right Banner. It is necessary to a full investigation and comprehensive evaluation of land quality.

2 Research Methods

2.1 Evaluation of Access to Information

In this study, all kinds of relevant information including soil physical and chemical properties measured by sampling and experimenting such as organic matter, available nitrogen, phosphorus and potassium was widely collected and collated for research. Social and economic data taking the administrative area as the basic unit was gained from statistical yearbooks. The collected base maps included the land use map on a scale of 1:50000, agricultural land classification map, productivity calculation map, the map of land use zoning, geomorphologic map and soil map. The field survey data included the soil parent material, landform, soil thickness, crop yield, and so on.

2.2 Determination of Soil Sampling Point

According to the topographic maps, land utilization map and soil map, the research had a good sampling point layout. The author divide Tumote Right Banner into three parts on the basis of the topographic maps combined with the actual, namely, the northern Mount Daqing mountain, central piedmont alluvial plain and alluvial plain of southern the Yellow River. Cultivated land proportion of them was 1:2:7. Sampling according to land ratio in different geomorphological positions on the layout of sampling points, a total of 150 sampling points were determined. Soil sampling numbers were marked on the map.

The sampling was carried out in the spring of 2011. According to the map to determine the point which was chosen from the representative field of a village, the map position of a point was modified by means of the field position and was validated by GPS. The author selected 5-6 sample points in random in the sampling region and used the stainless steel auger to collect soil samples, which were fully mixed and put into plastic bags (Figure 1).

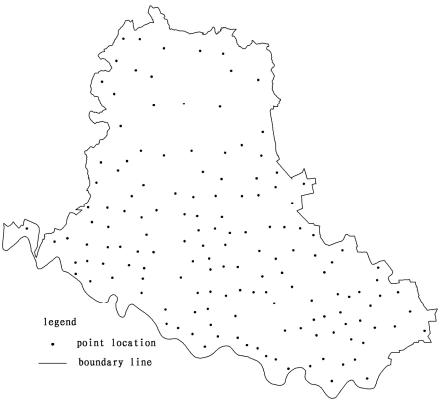


Fig. 1. Soil sampling points of cultivated land

2.3 The Division of Evaluation Unit

The geometric correction of the land use status map and the soil survey map were made for the same coordinate system to divide evaluation unit by spatial overlay method. Then the soil nutrient, site conditions and physicochemical properties were attributed to the overlay graph. By topology analysis the patches which were smaller than 4 square centimeter would be removed to establish the evaluation unit. The overlay graphs were divided into 306 evaluation units.

2.4 Selection of Evaluation Index

Cultivated land is a natural and historical complex. The quality of cultivated land is affected by many factors. This study followed sensitivity, diversity and dominant factor. Three indicators were selected including soil nutrient status, site conditions, and the physicochemical properties for cultivated land.

2.5 Evaluation of Farmland Nutrient Status

Evaluation Factors Were Selected. This study used the MATLAB software for cluster analysis on factors affecting the soil nutrient(Table.1).

The Weights of Evaluation Factors Were Determined. Multi-objective decision and AHP methods were a qualitative and quantitative combination, which were adopted to evaluate the quality of cultivated land. The judging matrix was established to weight coefficients of the factors. The sequence hierarchy and consistency check were operated. Then the level of total order and consistency inspection were run on the data. The results would be right if CI and CR were lower than 0.1(Table.1).

Table 1. The soil nutrient index system for cultivated land

Element	Weight	Factor	Weight
Primary factor	0.7052	Organic matter/%	0.3022
		Total nitrogen /(mg/kg)	0.0911
		Total phosphorus /(mg/kg)	0.0812
		Total potassium /(mg/kg)	0.0731
		Available nitrogen /(mg/kg)	0.0532
		Available phosphorus /(mg/kg)	0.0522
		Available potassium /(mg/kg)	0.0522
Other factors	02948	The pH value	02948

$$y = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

$$y = \frac{x_{\max} - x_i}{x_{\max} - x_{\min}} \tag{2}$$

Among them, Xi was the measured value for the factor. Xmin and Xmax were the smallest value and the largest value for the factor. By using the weighted sum method the soil nutrient index was determined (formula 3).

$$F = \sum_{i=1}^{2} G_i \sum_{j=1}^{7} Z_j M_j \tag{3}$$

Among formula (3), "F "was the soil nutrient index; "G_i"was weights of i factors; "Z_j"represented the standard value of the j factor; "M_j " represented the j factor weights.

Identify the critical points and ranked by total frequency curve method, the comprehensive index of nutrient status had 3 point mutations and was divided into four levels in Tumote Right Banner. The first level was larger than 0.671. The second level was between 0.621 and 0.675. The third level was between 0.577 and 0.621. The fourth level was smaller than 0.577.

2.6 Site Conditions, Physical and Chemical Properties Evaluated

The factors and weights were determined.Evaluation indexes were screened by DELPHI method. According to the influence index on the cultivated land quality, each evaluation index weights were detemined. The chosen factors affected site conditions included landforms, irrigation, drainage condition and the barrier layer depth. The weight values for factors were 0.25, 0.30, 0.22 and 0.23. Factors affecting the physicochemical property included surface texture, soil structure and soil salinization. The weight values for factors were 0.33, 0.25 and 0.42.

Using expert scoring method the various indexes classified and quantified (Table.2), the indexes were standardized using the formula (1) and formula (2).

Table 2. The site conditions and the physicochemical properties of cultivated land grading standard

Element	Factor	Classification and score			
The site conditions of cultivated land	landforms	Plain flat	plain hilly slopes	flat slopes	plain low
	value	100	90	80	70
	irrigation	Triple water	Twice water	Once water	No water
	value	100	80	60	30
	drainage condition	Better	Good	General	Worst
	value	100	70	40	10
	the barrier layer depth	>5m	4-5m	3-4m	<3m
	value	100	90	70	50
The physicochemicapropertries of cultivated land	surface texture	medium loam.	In light loam	light clay	sandy soil, heavy loam
	value	100	90	70	30
	soil structure	The soil, soil / clay / loam	Soil / viscose / viscose, soil / sand / soil	The sand, clay / sand / soil / sand, sand / sand	The gravel
	value	100	60	30	10
	soil salinization	No	Mild	Moderate	Severe
	value	100	70	40	10

With the weighted average method the unit value was calculated and grade was classified by cumulative curve. The site conditions were divided into four levels. The first level was larger than 0.629. The second level was between 0.561 and 0.629. The third level was between 0.500 and 0.561. The fourth level was smaller than 0.500. Physical and chemical properties of soil were also divided into four levels. The first level was larger than 0.880. The second level was between 0.780 and 0.880. The third level was between 0.650 and 0.780. The fourth level was smaller than 0.650.

2.7 Comprehensive Evaluation of Cultivated Land Quality

The weight of evaluation index was determined. In Tumote Right Banner, the three main influence factors for the cultivated land quality were considered to determined their weights, soil nutrient, the site conditions, and the physicochemical properties whose weights were 0.38,0.34 and 0.28.

Using Arcgis software and the weighted sum calculation, the maps which were soil nutrient status of cultivated land, site conditions and soil physicochemical properties, were overlayed to be gained the comprehensive index of cultivated land quality. The natural curve breakpoint method was adopted to divide the quality of cultivated land into four levels. The first level was larger than 0.9438. The second level was between 0.9208 and 0.9438. The third level was between 0.9012 and 0.9208. The fourth level was smaller than 0.9012. The distribution map of the cultivated land quality grade was made(Figure.2(4)). At the same time, the area of each grade was counted(Table.3).

3 Evaluation Results Analysis

3.1 The Grade Distribution and Area of Soil Nutrient Status of Cultivated Land

The soil nutrient status of cultivated land, site conditions and physical and chemical properties in Tumote Right Banner, the distribution of which was shown in Figure 2, and the level and the area percentage in Table 3.

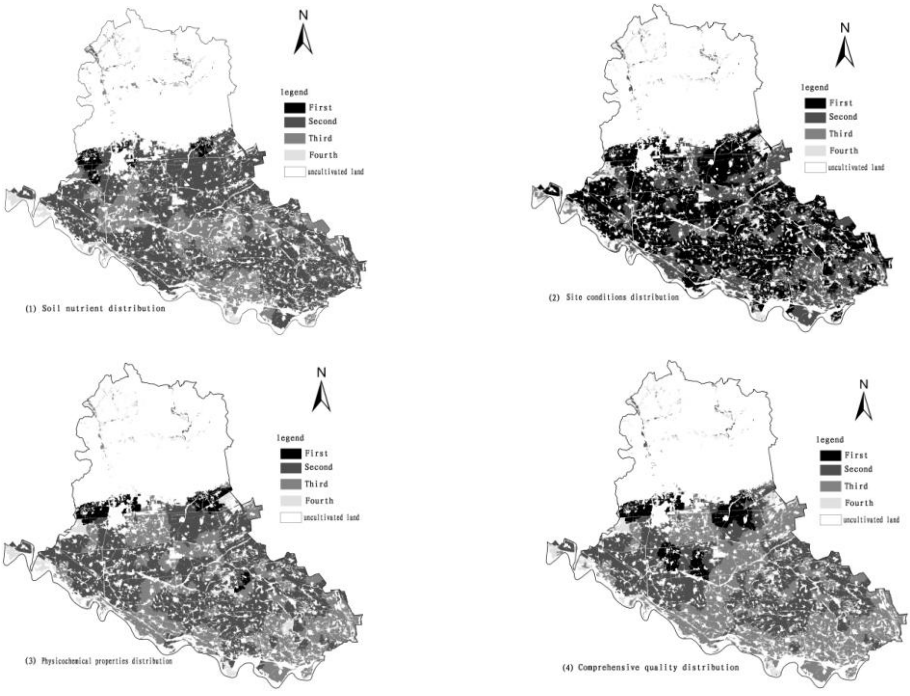


Fig. 2. Distribution level of cultivated land quality

Table 3. The level area and the area percentage of cultivated land quality

Factor	Soil nutrient		Site conditions	
Level	Area/hm^2	Percentage	Area/hm^2	Percentage
First	3306.94	3	74846.14	63.34
Second	81376.88	68.77	36485.92	30.88
Third	30546.83	25.75	3777.94	3.2
Fourth	2929.03	2.48	3049.68	2.58
Factor	Physicochemical properties		Comprehensive quality	
Level	Area/hm^2	Percentage	Area/hm^2	Percentage
First	5744.05	4.86	9649.76	8.17
Second	65036.28	55.04	51987.19	44
Third	40839.75	34.56	51369.95	43.48
Fourth	6539.6	5.53	5152.78	4.36

As could be seen from table 3, soil nutrient status of cultivated land was in the upper level in Tumote Right Banner. The first level area was 3306.94 hectare which was 3 per cent of the total cultivated land area. The second level area was 81376.88 hectare which accounted for 68.77%. The third level area was 30546.83 hectare, accounting for 35.75%. The fourth was 2929.03 hectare, accounting for 2.48%. Soil fertility of the first grade land was mainly distributed in the middle north in Tumote Right Banner, which of the distribution is the least in Gou men town, and that in Meidaizhao town was the most. This region is in the middle of alluvial fan had fertile soil which has rich organic matter, available nitrogen, available phosphorus, available potassium and trace element abundance to supply crop growth.

Both the second level land and the third level land were distributed each town in Tumote Right Banner. These cultivated lands were in the cross section of the Yellow River alluvial fans and Mount Daqing diluvial fans, the main regions of corn production, high nutrient content of soil, where irrigation canal system matched condition was quite good to guarantee the need of crop growth for nutrition. Four level lands scattered along the coast of the Yellow River and Mount Daqing, is mostly the saline-alkali soil, which have heavy salinization, poor soil, more serious flood disaster, less capital investment. It was difficult for crop growth.

3.2 Grade Distribution and the Area of Site Conditions

Grade distribution and the area of site conditions were showed in table3.The first grade land area was 5744.05 hectare for 4.86 per cent of total area of cultivated land. The second grade land area was 65036.28 hectare accounting for 55.04 per cent of total area. The third grade was 40839.75 hectare, accounting for 34.56 per cent, and the fourth grade land was 6539.60 hectare, accounting for 5.53 per cent of total area of cultivated land.

As you could see from figure 2(2), good site condition of first or second grade was mainly distributed in the central and southern, including Mingsandnao, Meidaizhao, Goumen, General Yao, Shuanglong and Haizi. This region had rich water resources, lower soil salinization, flat terrain, convenient irrigation, pertocalcic horizon for hindering crops growth deeply buried. All of these conditions indicated that it is ideal to agricultural production. The third or fourth grade was mainly distributed in the surrounding area along the Yellow River and the main canal, where salinization was most serious. So it was most difficult for agricultural production.

3.3 Grade Distribution and Area of Physical and Chemical Properties of Cultivation

The first level was 74846.14 hectare, 63.34 per cent of total cultivation. The second level was 36485.92 hectare, accounting for 30.88 per cent. The third level was 3777.94 hectare, accounting for 3.20 per cent. The fourth was 3049.68 hectare and 2.58 per cent of total cultivation.

As you could see from Figure 2(3), the first or second level which had better physical and chemical properties of cultivation was mainly distributed in the southern towns of Mount Nine Peaks, including Salaqi, Mingshanao, Meidaizhao, Goumen, General Yao, Shuanglong and Haizi. The soil physical properties of this region were helpful for crop growth. There were many aggregate structures in deep soil which belonged to medium loam or light loam. The third and fourth grades were serious for soil salinization. There were the thin soil layer profile which was fluffy sand or gravel was poor for the ability of fertilizer and water conservation, were clay plate or block structure for water permeability and air permeability to be difficult, and calcium in soil layer for a great adverse effect on the crops growth. So the output rate of cultivated land was quite low.

3.4 Analysis Results of Cultivated Land Quality Assessment

The cultivated land quality level is good on the whole in Tumote Right Banner. The first grade was 9649.76 hectare, 8.17 per cent of the total cultivated land area. The second grade was 51987.19 hectare, 44 per cent of the total land area. The third grade was 51369.95 hectare, 43.48 per cent of the total land area. The fourth grade was 5152.78 hectare, accounting for 4.36 per cent of the total cultivated land.

From Figure 2 (4) could be seen, the first or second level of cultivated land had higher quality. The soil physical and chemical properties were suitable for agricultural production. There was high soil nutrient content, good site conditions, low salinization, irrigation facilities, good drainage condition. Meantime, the terrain was flat and agricultural infrastructure was perfect. The third or fourth level cultivated land was mainly distributed in the low-lying along the Yellow River coast or the slope of Mount Daqing foothills. Comparatively, soil nutrient content was lower and the sand, gravel or viscous component was larger. Add in the low investment in agriculture, the flood and drought disaster, incomplete infrastructure and the shallow the barrier layer from the surface of the earth, these adverse factors brought about soil

potential fertility to play ineffectively and to be low agricultural productivity. In brief, the cultivated land quality was in middle level in Tumote Right Banner. High quality cultivated land was only 8.17 per cent of the total cultivated land area. The medium quality cultivated land accounted for about 77.48 per cent.

Therefore, China would built four hundred million acre high-standard basic farmland in 2020. In Tumote Right Banner, we should seize this opportunity to keep the level of cultivated land quality and not be reduced, increase the intensity of the transformation of the second or third grade cultivated land, and improve agricultural infrastructure construction for the fourth grade cultivated land.

4 Conclusions

The cultivated land quality is irrelevant to food security of a country and social stability. The evaluation of cultivated land quality is great significance to the protection of arable land and the improvement in low-yielding fields. Tumote Right Banner in the Yellow River alluvial and pluvial plain on the south of Mount Daqing, a long history of agricultural exploitation, is a typical agricultural county in Hetao plain. Water and heat conditions are good for agricultural development. Soil profile configuration is suitable for agricultural development, and agricultural infrastructure is relatively good. The overall quality of cultivated land is on the high side. Based on the field investigation and sampling, using geographic information system as a mean, the quality of cultivated land was evaluated using soil nutrient, soil conditions, physical and chemical properties in Tumote Right Banner in order to expand the cultivated land quality evaluation ideas and to provide reference information for the right of agricultural development.

Acknowledgments. This work was financially supported by the Inner Mongolian Natural Science Foundation (2012MS0618) and the National Natural Science Foundation of China (41261054).

References

1. Nong, X.-X., He, Z.-W., Wu, B.-Q.: Application of ARCGIS Spatial Analysis Model in Evaluating Cultivated Land Quality. Research of Soil and Water Conservation 16, 234–236 (2009)
2. Shi, S.-Q., Chen, Y.-Q., Yao, Y.-M., Li, Z.-I., Hey, I.-B.: Assessing Natural Quality and Use Quality of Cultivated Land in Northeast China. Resources Science 30(3), 378–384 (2008)
3. Xing, S.-H., Huang, J., Huang, H., Mao, Y.-L.: Quality evaluation of regional cultivated land based on GIS. Journal of Fujian Agriculture and Forestry University (Natural Science Edition) 31(3), 378–382 (2002)
4. Hu, K., Shi, P.-J.: Evaluation and Study on Quality of Cultivated Land in Gansu Province. China Land Science 22(11), 38–43 (2008)

5. Zhang, F.-R., An, P.-L., Wang, J.-Y., Zhang, J.-L., Liu, L.-M., Chen, H.-W.: Soil Quality Criteria and Methodologies of Farmland Grading. Resources Science 24(2), 71–75 (2002)
6. Han, M., Zhang, H., Chen, X.-H., Peng, J., Lou, B., Xie, H., Wang, D.: Study on Agricultural Land Quality Assessment and Its Application in Guizhou Province. Guizhou Agricultural Sciences 38(4), 88–91 (2010)
7. Cheng, J.-N., Zhao, G.-X., Zhang, Z.-X., Wang, J.: GIS Supported Comprehensive Evaluation of Cultivated Land Quality at Small Scale —A Case Study in Dingzhuang Town of Shandong Province. Journal of Natural Resources 24(3), 536–543 (2009)
8. Wang, L.-J., Zhao, Y.-G., Guo, M., Zhang, G.-L.: Evaluation of Farmland Productivity Based on GIS and Fuzzy Mathematics Theory at County Level. Soils 42(1), 131–135 (2010)
9. Xu, Z.-Y., He, J.-L., Liu, W.-B., Hu, Z.-L.: The Different Degree of Arable Land Quality in Hunan Province Based on Gini Coefficient Analysis. Syatem Sciences and Comprehensive Studies Inagriculture 24(2), 208–213 (2008)
10. Zhang, S.-H., Liu, Y.-Z., Zhang, K.: Study on Quality Evaluation of County Cultivated Land Based on Information Technology. Journal of Anhui Agri. Sci. 36(19), 8202–8204 (2008)
11. Wang, Y.-B.: Exploration of Comprehensive Quality Evaluation Method of Cultivated Land for County Areas. Resource Development & Market 25(7), 598–602 (2009)
12. Yuan, T.-F., Qiu, D.-C.: Correlative Analysis of Coupling Relationship between Distributing of Cultivated Land Quality and Economy in Chongqing. Economic Geography 28(3), 484–496 (2008)

Research of Scheduling Strategy Based on Fault Tolerance in Hadoop Platform

Zhengwu Yuan and Jinli Wang

College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Abstract. The scheduling problem is a hot issue in the current cloud computing, and node failure conditions should be taken into consideration. Firstly the disadvantages of current Hadoop task scheduling algorithms and fault tolerance in Hadoop platform are discussed in this paper. Then a scheduling strategy based on fault tolerance is presented. According to this strategy, the cluster detects the speed of the current nodes in a cluster, and makes some backups of the intermediate MapReduce data results to a high-performance cache server, and the data is produced by the node that may go wrong soon. Thus the cluster may resume the execution to the previous level quickly when there are several nodes going wrong, the Reduce nodes read the Map output from the cache server or both of the cache and the node, and keeps its high performance. Finally the computer simulation is done. It shows that the strategy presented in this paper is effective under failure tolerance conditions.

Keywords: Failure tolerance, Task scheduling strategy, Hadoop platform, Backup, MapReduce.

1 Introduction

Computing[1] node failure is very common in the Cloud Computing model, and there are a mass of tasks running in the same cluster at the same time. And a task will definitely go wrong if there are problems in the running node, thus it will reduce the performance of the cluster. Besides, the phenomenon that the nodes in the cluster will go wrong after running in a certain time cannot be avoided. The Hadoop[2] Cloud Computing platform is an open source in the Apache project, and it has been widely used for development and research of Cloud Computing by a majority of manufactures and universities. The core technologies of Hadoop platform are mainly composed of MapReduce[3] functional programming model and HDFS[4](Hadoop Distributed File System)[5]. In the Cloud Computing model, the computer servers that are being used are usually ordinary ones, and the server are almost running all the time in case they are invited frequently, which is a fact. Cluster server nodes and disks go wrong are common, a study made by Yahoo shows that there are about 2 to 3 percent of server nodes going wrong every day in a thousand of nodes that are running for a very long time, and the failure rate of new nodes could be about three times more. Research of task scheduling algorithm in Hadoop platform is a hot issue currently, and it is also the key to improve the cluster performance.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 509–517, 2013.

There are mainly three built-in job scheduling algorithms[5] in Hadoop platform. They are FIFO(First In First Out) scheduler, Fair Scheduler[7], Capacity Scheduler[8], each has its own advantages, and there are also defects in them. And the schedulers mentioned above have not taken fault tolerance into consideration. Thus, the scheduling algorithms can still be improved to improve the performance of the cluster. Some scholars put forward a few improved scheduling algorithms to improve the performance of the cluster, such as delay algorithm, LATE(Longest Approximate Time to End) algorithm and improved LATE algorithm, etc. But there are few that have taken fault tolerance into consideration. Most of the time, the cluster will re-execute or make a backup of the tasks that go wrong or runs very slow compared to other tasks. And in such conditions, there will be large numbers of backups and re-executed tasks in the cluster, and it will use more of the resources of the cluster, thus giving much pressure to the cluster as well as reducing the utilization of cluster resources.

First the disadvantages of Hadoop built-in scheduling algorithms is analyzed in this paper, then a new scheduling strategy based on fault tolerance in Hadoop platform is presented. The cluster will detect the speed of the current nodes that are running in the cluster and determines which of the nodes may possibly go wrong, and then makes some copies of the intermediate data results made by the node that may go wrong soon, especially big and long tasks which may make more intermediate data results. Thus it makes the cluster run stronger to deal with node failures, and keeps its high efficiency. In section 1, several built-in scheduling algorithms in Hadoop are discussed. In section 2, the operating mode and fault tolerance are analyzed. In section 3, a task scheduling algorithm based on fault tolerance in Hadoop platform is presented. In section 4, the simulation experiment is done to show that the strategy is effective. Finally, the conclusions are drawn in section 5.

2 Hadoop Built-in Schedulers

2.1 FIFO Scheduler

It is the default scheduler in Hadoop platform, Using an FIFO queue to server those that come first. In accordance with the priority of the job and then according to the time the job submitted to select a job to execute. The advantages of this scheduler are that it is very simple to use, and the scheduling process is very fast. The disadvantages of the scheduler are that the utilization of cluster resources is very low, and small jobs are very likely to become hungry.

2.2 Fair Scheduler

The purpose of fair scheduling is to ensure all the jobs could maintain an average of the cluster shared resources as time passes. The scheduler manage tasks and assign them to each pool so that they can share the resources fairly, and multiple queue and multiple user is supportive, jobs in the same queue share all the resources of the queue fairly. The advantages of this scheduler are that it can schedule the tasks by classification. Different

kinds of jobs will have different resources assigned by the cluster, thus, it will improve the quality of the service and make the most of the cluster resources. The disadvantages of this scheduler is that it does not consider the actual loading state of each nodes[9], so the load balance in each node may not that good.

2.3 Capacity Scheduler

Capacity scheduler improved the function of HOD(Hadoop On Demand), and overcome some of its disadvantages. This scheduler uses several queues, every queue get its resources according to its computing ability, and the rest resources will be assigned to those queues that have not met its usage limits, and allocate them base on the load of its calculation, which is more reasonable. In case that all the resources of the queue are occupied by only one use, the scheduler limits the resources of every user in the queue. And this scheduler supports job priority and jobs can execute in parallel to allocate resources dynamically and thus improves the efficiency. The disadvantages of this scheduler are that it needs to know a lot of job information of the cluster, the cost of the system is very expensive. Besides, only memory resource scheduling is supported by the algorithm.

3 Analysis of Operating Mode and Fault Tolerance in Hadoop

3.1 Task Scheduling Mode in Hadoop

Hadoop use master/slave scheduling mode, all the jobs in the nodes are managed unified by the master, say NameNode. NameNode is responsible for managing the namespace file system, maintaining the file system tree and all the files and directories in the system. In general, a slave computer server is a datanode in the cluster, which keeps the block data information of every file in each datanode.

The master pings each computing datanode periodically to determine whether the node is going down or not during the running time of the cluster. If a datanode is idle and there are unexecuted jobs or tasks in the queue, the master will assign a task to the node. And when the master finds that there are faults in any of the running node or there are tasks stopped because of hardware or software going wrong, it will make a backup and re-execute the tasks that are running on the failure node in another node to ensure that the cluster is still reliable, making the cluster robust. In the Hadoop platform, jobs are running mainly use MapReduce functional programming model. Firstly, the data will be cut into splits, and then is handled by the Map function and intermediate results are made. Then, the intermediate data will be copied to the Reduce node for reducing, the output are the final results. Finally, the output are made and written into the global file system. During the computing, when a task is failed, because of other hardware failures or if there are nodes going wrong, the job will fail or the running speed of the job will decline rapidly. At the same time, the intermediate results that the Map made will be unavailable, the output of the Reduce proceed is written to the Hadoop global file system, and need not to be considered like Map results mentioned above. The topic in this paper is mainly concerned with the output of Map.

3.2 Fault Tolerance

Cloud Computing is a distributed file system, and there are many places that may cause exception to the cluster[10]. As is shown in a statistical analysis report in Google cluster, there are quite a lot of nodes that go wrong every day, and Hadoop platform is quite familiar with the Google file system. When a job cannot complete in a normal way, or because of other factors(server, network, disk failures, etc) that the server is not available, the cluster need to re-execute the task or job, using a new node to recalculate, and all the previous output by the same task cannot be used. Such manners keep the reliability of the cluster, but it is a waste of cluster resources, what is more, it will reduce the performance of the cluster.

What fault tolerance means is that the system has problems because of some certain reasons, and when the data or files are damaged or lost in the system, it can recover the data or files to the state before the accident, so that the system can work normally as usual. Cloud Computing are able to deal with failures, but with the cost of running time and performance.

The fault tolerance of Hadoop platform mainly uses redundancy to prevent data loss. Section 2 discussed the execution flow in Hadoop platform. There are multiple backups for each data in the platform typically 3 backups, that is, one data stored in local node, the second copy stored in local rack, another copy stored in a remote random rack to ensure the reliability of the cluster. When there are failures in any node, the cluster get data from another backup of the two which is close to it.

Distributed node status detection methods are mainly two kind of ways, active and passive[11]. The passive detection is that the host ping the subordinate computers periodically to see if the node is healthy, such as the master ping every datanode to determine the status of the node depending on whether there is a response from the node. The active detection method is that, a datanode sends its status messages to the host periodically, such as the datanode sends heartbeat packets to the master on a regular basis to report its own status, if the host does not receive a heartbeat packet from a datanode in a certain time, it knows that the node has gone wrong.

4 Scheduling Strategy Based on Fault Tolerance

The strategy presented in this paper discusses the re-execution of failed MapReduce tasks, mainly for the Map tasks, because the output of Reduce tasks are written into the global file system, so the effect is not that obvious compared to the Map tasks. The cluster detects the nodes that run Map tasks within the cluster periodically to see whether there is something wrong with the node or whether the running speed slows down terribly, if so, then making a copy of the intermediate results from the Map task to the cache server in case that the results are lost and can not be used again, especially big jobs or the Map task is finished but the output data has not been read. And when the cache server knows that the data is not useful any more, the data will be deleted to free memory space and other resources.

4.1 Basic Idea

Here are two definitions as follows.

Definition 1. Node about to go wrong: First detecting the running speed of the node, and if the speed slows down rapidly, such as it becomes half the speed or lower than before, we say that the node runs abnormally and is about to go wrong soon. And the speed mentioned above is that the task running speed compared to other similar tasks or the computer running speed compared to the previous computer running speed.

Definition 2. Backup interval: It means progress interval or time between two progresses, such as from 20 percent to 50 percent. Increasing the interval is that the latter progress is increased, for example, change the progress from 50 percent to 60 percent, or postpone a few milliseconds to make backups.

The general idea or progress of the strategy is as follows:

a) Add an extra computer servers whose performance is higher than the other node in the rack of the cluster as a cache node, and keep the network within the rack, so that every node within the rack can visit the cache server very quickly, and make sure that the cache node is better than the other nodes. And notice the information of the cache to the master, but the master does not control the cache server.

b) Every node server detects the running speed of the computer itself periodically, if the speed slows down rapidly(e.g., It is about 50 percent or more lower than the previous average speed, the judging standard can be changed subjectively), so we believe that the node has something wrong or it is going to go wrong very soon, and in such situations, make the output of the node written into the cache server of the rack periodically, and the data which have been copied to the cache is skipped and do not copy once more.

c) When the rack I/O load is heavy enough, that is to say when the I/O is very busy at present, the backup interval will be increased by a very short time in order to reduce the I/O pressure. But if the job is very big or the intermediate data is quite a lot or important, the copy to the cache will be make immediately.

d) If the node is broken down or because of other reasons the job or tasks are failed, and the intermediate results can not be accessed, then the data of the cache within the rack are used, and the job or tasks need not to re-executed from the beginning, only part of the job or tasks are re-executed, thus saving a lot of time, especially when the job is finished, the data have not been read and the data can be used and no job or task is done a second time.

The cache server has a larger memory or other configurations compared to other node in the rack, thus its performance is also higher. The data copied from other node are put into the memory first, and written into the disk when the data is large enough. And when the system is sure that the data is not useful any more, system will delete the data to free memory and space resources.

4.2 Flow-Process Diagram and Feasibility Analysis

Fig. 1 is the general process of the algorithm, after analyzing the algorithm above. We know that this scheduling algorithm is theoretically feasible and effective. When this

algorithm is working, first detecting whether the I/O of the rack is overload or not, if so, it will stop working for a very short time, but if the task is very big and urgent and output a lot of data, the backup should be made in time for the job will use more system resources when execute it a second time. If system I/O is very busy, the copy will be postponed for a very short interval to reduce the pressure of the current system I/O.

It is easy to add an extra cache computer server to each rack and the cost is quite little. And the cluster just inform the location status of the cache to the master, that means adding a node in each rack to the cluster, and the cache is not controlled by the master. A rack can be seen as a simple small LAN, the communication bandwidth can be guaranteed between the cache and the other nodes, if there are too many nodes within a rack, the number of cache can be increased by one or more among switches to ensure the quality of the service and its performance. As long as the intermediate data is not a one-time output and with multiple files, the cache server can play its role in the cluster, especially for the big jobs or tasks, this algorithm is more effective. And like other schedulers, this scheduler can also be an optional plug-in to use in the Hadoop platform, and the function can be closed by setting the relevant options if necessary. And the key point of this algorithm is that the cluster should detect the hardware disk failure or software failure and predict the precursor of the failure of a job or task, otherwise the performance may not be that high as expected, for causing unnecessary data communication cost.

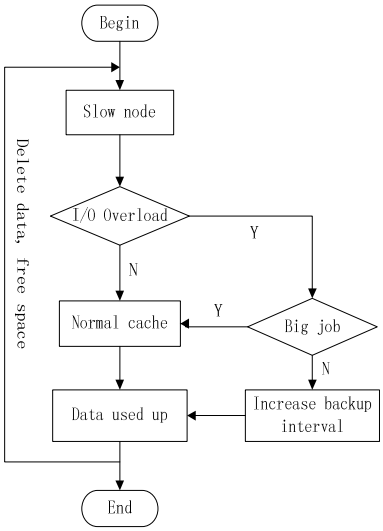


Fig. 1. Algorithm execution flow chart

5 Simulation

April 2009, the Grid Laboratory Gridbus project of the University of Melbourne, Australia announced the launch of Cloud Computing simulation software called CloudSim[12]. CloudSim is supportive of the modeling and simulation of large-scale

Cloud Computing infrastructure. It can switch neatly between time-sharing and space-sharing when assigning virtual processing cores. The components are all open sources in the CloudSim platform, it is very helpful to accelerate the development of Cloud Computing algorithms, methods and norms. First, create a project of CloudSim in eclipse software development platform, rewrite VM-Scheduler and SpaceShared scheduling algorithm and Mapping functions from jobs to virtual machines, then start your own scheduling algorithm in CloudSim, which expand the CloudSim by defining new VM (virtual machine) classes and new scheduling classes. Compile the demo program after everything is ready, and simulate the algorithm program, at last we get the data.

The simulation experiment conditions in CloudSim are configured as follows.

Table 1. Virtual machine configuration

VMs	Tasks	Memory (MB)	CPU core(s)	Speed (MIPS)	Disk (GB)	Network (Mbps)
100	500	1024	1	1000	100	1000

a) Version of the simulation software is 3.0.1, JDK 1.6.0 and Eclipse software development platform. b) One hundred of virtual machines that are similar to each other are created. c) Five hundred of similar jobs are created, whose execution time are similar but not exactly the same. d) Two data centers to work together.

Fig. 2 and Fig. 3 are the data results comparison in conditions that there are failed nodes but different failure rate using fault tolerance and without fault tolerance. It shows that the advantage is not so obvious when there are few failed nodes in the cluster, because the algorithm will cost some of the cluster resources itself. With the increase of failed node, the superiority of the algorithm begins to emerge, only in conditions that when failure rate is quite enough, the cost the algorithm can be ignored compared to re-executing the job or tasks. In such conditions, the algorithm could save some resources and performance of the cluster for the cluster need not to re-execute all the job or tasks that have finished or partly finished where there are failed job or tasks because hardware or software failures. Fig. 4 shows that when there are no failed nodes the algorithm based on fault tolerance cost a little more system resources and time because in such conditions, the cache will read and write data from the nodes which product data within the rack to make copies, but the cost is quite little compared to the whole cost of the cluster which is acceptable.

This strategy is based on fault tolerance, and the cluster still use the default scheduling algorithm in Hadoop, when there are failed job or tasks and need to be re-executed because of node failures or other reasons, the strategy can be helpful, especially for big jobs. When using the strategy presented in this paper, the cluster need to detect the health status of every node and it can be done by the node itself. What is more, an extra cache server within the rack is needed, and keep its network in good health. On the whole, this strategy presented in this paper does not cost too much compared to the whole, and can obtain good results in terms of fault tolerance to show that it is effective.

The following charts are the results comparison in different conditions, which means the jobs are in different complete progress when failure occurs.

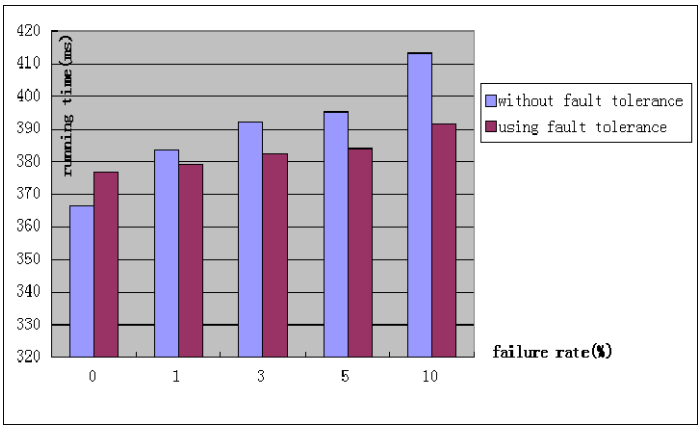


Fig. 2. Results comparison of using and without *fault tolerance*

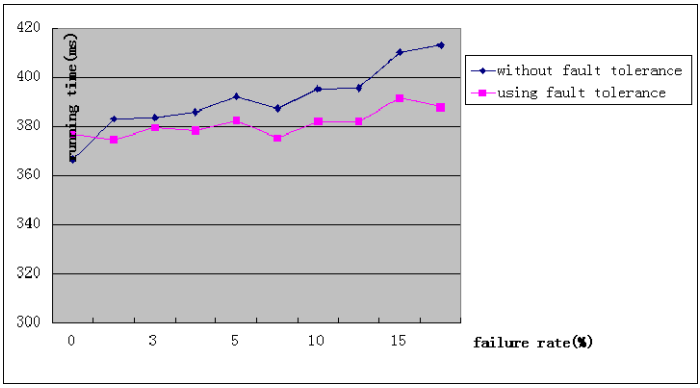


Fig. 3. Results comparison of using and without *fault tolerance*

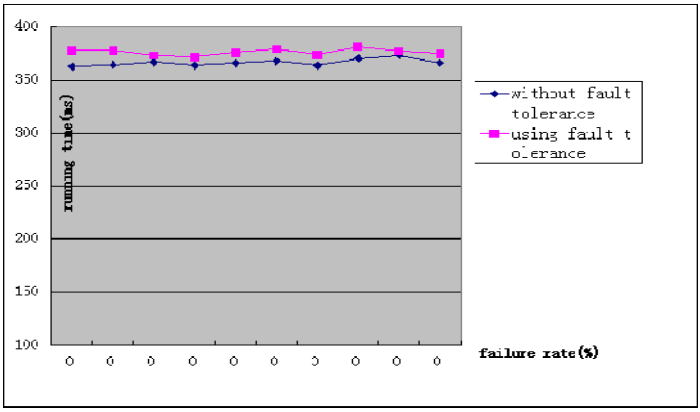


Fig. 4. Results comparison without failed nodes

6 Conclusions

The computer simulation shows that this backup algorithm could keep its high performance of the cluster when there are failed jobs or computer server nodes, it is feasible and effective. The experiment was done in simulation situations, and the failed nodes were set in the program during the execution of the job by shutting down some of the virtual machines forcefully. And the testing data is not that large like the cluster, and the experiment of this algorithm in actual platform will be done in the future, and the strategy does not solve the problem of NameNode single point of failure.

Acknowledgments. This work was partly supported by the Natural Science Foundation of Chongqing(No. cstc2012jjA40064) and the National Natural Science Foundation of China(No. 61075019).

References

1. Vaquero, L.M., Rodero-Merino, L., Caceres, J.: A break in the cloud: Towards a cloud definition. ACM SIGCOMM Computer Communication Review (2009)
2. Apache: Welcome to Apache Hadoop (2011), http://Hadoop.apache.Org
3. Dean, J., Ghemawa, S.: MapReduce: Simplified Data Processing on Large Clusters (2008)
4. Borthakur, D.: The Hadoop Distributed File System: Architecture and Design, http://Hadoop.apache.org/common/docs/r0.17.2/hdfs_design.htm
5. White, T.: Hadoop: The Definitive Guide. O'Reilly Media, Inc. (2009)
6. Thirmala Rao, B., Sridevei, N.V., Krishna Reddy, V., et al.: Performance Issues of Heterogeneous Hadoop Clusters in Cloud Computing. Global Journal Computer Science & Technology (2011)
7. http://hadoop.apache.org/docs/r0.2.0/fair_scheduler.html (2012)
8. http://hadoop.apache.org/docs/r0.20.2/capacity_scheduler.html (2012)
9. Hu, D., Yu, J., Ying, C., Zou, W.: Improved LATE scheduling algorithm on Hadoop platform. Computer Engineering and Applications (2012)
10. Zhang, Y.: Data Collection and Fault-tolerance Analysis about Cloud in the Calculation. Computer Knowledge and Technology (2011)
11. Li, C., Dai, Y., Wang, W., Wu, B.: Hadoop Framework for Disaster Technology. Computer Knowledge and Technology (2009)
12. China Cloud Computing: CloudSim, Cloud Computing simulator of University of Melbourne Australia (2012), http://blog.csdn.net
13. Liu, P.: Cloud Computing, 2nd edn. Electronic Industry Press (2011)

Efficient Geographical Diversification of Export Trade: The Case Study of China

Jin Dai[1,2]

[1] Pinghu Campus of Jiaxing University, Pinghu, Zhejiang 314200, China
earl@hz.cn
[2] Institute of Finance and Economics Research, Shanghai University of Finance and Economics, Shanghai 200433, China
earldai@gmail.com

Abstract. The issue of export diversification has received considerable attention since 1950s. However, there is very few works referred to the efficient geographical diversity of export trade. This article attempts to fill the gap in the literature. Based on the overview of measures used in empirical literatures, the paper constructs the deviation index to measure the efficient diversity. With the time series data from 1990 to 2011, this article analyzes the structure of China's export markets. The results show the structure has been optimizing constantly since China introduced the strategy of market diversity in early 1990s. Meanwhile, the deviation index further indicates there is still a huge improvement space for China's exports to the markets of Europe and Asia.

Keywords: Export Trade, Geographical Diversification, Deviation Index.

1 Introduction

Although China has implemented the strategy of market diversity for more than 20 years, it has observed a continual increase in trade frictions between China and its trade partners since the strategy employed in early 1990s (CEPR, 2012). According to figures from China's Ministry of Commerce, China's exports accounts for about 10% of the world's total amount at present, but about a third of the world anti-dumping case and half of anti-subsidy cases against China, China has been ranked at the top of the countries suffered from trade friction for 18 years (1995-2012). In 2012, for instance, China was confronted with 77 cases of trade remedy from 21 countries, the total amount involved reached to $27.7 billion. According to the ministry's recent statistics, only in the first quarter of 2013, 12 countries brought 22 cases of trade remedy against China, with more than 40 percent uprising from the same period of last year. Meanwhile, various forms of trade frictions emerged one after another, and the export markets involved in the trade frictions are gradually diversified. China's export is blocked not only in traditional markets, such as Europe and the United States, but also in some emerging economies. It seems that "Made-in-China" is confronting with an unprecedented impact. Consequently, it is of crying need to optimize China's export market structure at present.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 518–530, 2013.

The issue of export diversification in economic studies is not a newcomer. The role of export diversity has received considerable attention since 1950s. Works on export diversity developed rapidly in the followed decades, more and more researchers paid attention to the correlation between export diversity and export stability (Mac-Bean & Nguyen, 1980 [11]). In recent literature, empirical researchers provided a link between export diversity and economic growth (Al-Marhubi, 2000 [3]; Berthelemy & Chauvin, 2000 [5]; Sanjay, 2011 [14]). However, there is no work referred to the efficient geographical diversification of export market. This article attempts to fill the gap in the literature.

The main questions of this article devoted to as follows: Did China's strategy of market diversity work on well in the past years? How to measure the efficient geographical diversification of export? Where is the way out for China's export market diversification?

The rest paper is organized as follow. Section 2 reviews literature on diversification measurements used in the existing empirical researches. Section 3 constructs the deviation index and then we apply the index to analyze China's export market structure in section 4. The paper concludes in section 5.

2 Review of Literature on Diversification Measurement

This section presents an overview of the most commonly used measures in empirical studies on export diversification. These measuring tools include: the concentration ratios, CSCEEF and absolute deviation, etc.

2.1 Concentration Ratios

The concentration ratio is the most commonly used measure of diversity, which appears frequently on empirical studies related to export diversity and export instability. There are several tools developed to measure export diversity under this group, such as Hirschman index (Hirschman, 1964 [7]), the Ogive index (Ali, Alwang & Siegel, 1991 [2]), the Entropy index and the aggregate specialization index (Attaran & Zwick, 1987 [4]), etc. We only introduce the most widely used tool, the Hirschman index, since the approach of these measures are quite similar, which take a hypothetical distribution as the standard and then compare it to the actual distribution.

The Hirschman index was developed by Albert Hirschman in 1945. This index is also known as the Hirschman-Gini index, or Gini-Hirschman Index, and also refers to as the Herfindahl-Hirschman index (Adelman, M., 1969 [1]; Kelly, 1981 [8]). It may be written as:

$$H_1 = \sqrt{\sum_{i=1}^{n} (X_i / X)^2}$$

Where Xi represents the exports of a specific commodity i, X is the country's total amount of exports, and n is the number of commodity classification. A higher H1 indicates greater concentration of exports on a few commodities. For a given n,

the minimum value of H1 is n-1/2 when each commodity's share of export is identical. Meanwhile, H1 get the maximum value, 1, when there is only one export commodity group. Hirschman index also used to measure geographic concentration of exports (Kingston, 1976 [9]).

In empirical studies, the Hirschman index has been frequently used to investigate the relationship between export diversification and export instability (Kingston, 1976 [9]; Mac-Bean & Nguyen, 1980 [11]; Love, 1979 [10]).

2.2 CSCEEF

CSCEEF is the abbreviation of the Commodity-Specific Cumulative Export Experience Function. It is widely used to measure the structural changes of exports (Pineres & Ferrantino, 1997 [12]). CSCEEF has properties analogous to a cumulative distribution function, as it may take a very small value in the initial period and subsequently rise in subsequent periods. The CSCEEF for a particular commodity i can be obtained by the following formula:

$$CSCEEF_{i,\tau} = \sum_{t=t_0}^{\tau} X_{it} / \sum_{t=t_0}^{T} X_i$$

Where X_{it} is the export value of commodity i in year t; t_0, τ, T represent the initial, current and terminal periods of the sample period, respectively. The CSCEEF may take zero value in the initial period and subsequently rises to 1 in the terminal period. On the plots, and for a more traditional commodity, one would expect to see the plot shifted to the left, whereas for a non-traditional commodity, one would expect to see the plot shifted more towards the right. With the comparison of CSCEEF across different commodities, we can get the direction for export diversification. For example, if the plots of CSCEEF are shifted further to the right, the export commodities should not only be considered to be more non-traditional but also be expected to be more vertically diversified (Ben Hammouda etc., 2006 [6]).

2.3 The Absolute Deviation

A third method to measure export diversity is the absolute deviation of the country's commodity share from the world structure. This measure can be calculated as:

$$S_j = \frac{1}{2} \sum_i \left| h_{ij} \right| - \left| h_i \right|$$

Where h_{ij} is the share of industry i in total exports of country j and h_i is the share of industry i in the world exports. Again the value of this measure ranges from 0 to 1, where 1 represents total concentration and 0 represents total diversity.

The absolute deviation is widely used by UNCTAD to measure the extent of the differences between the trade structure of a particular country and the world average (Samen S., 2010 [13]). The value closer to 1 indicates a bigger difference from the world average. The index was also used by Al-Marhubi (2000 [3]) as one of the explanatory variable in the growth regression and found to be significant in explaining faster economic growth.

To sum up, the concentration ratio, such as Hirschman index, the Ogive index, the entropy index is frequently used to analyze the problem of export diversification by the above-mentioned literature, but the studies did not shed light on the efficient diversification of export market. We can attribute the gap to the limitations of the measures adopted by the existing studies on trade diversification. Despite lack of clear economic meaning and subjective selection of data, the existing measures are unilaterally designed from the eye of export side, excluding the influence of import side. In addition, some indexes ignore the capacity of import market, which made their conclusions obviously biased. Furthermore, the tools mentioned above only plot the macro-level structure of the whole export market, but it's too far away from the destination. We need to figure out the specific structure of each specific market. Without the specific structure, we cannot find the way to the efficient diversification. Based on the above considerations, this article introduced the deviation index to measure the efficient diversification. With the empirical analysis followed up in section 3, we should find ourselves in the right way to efficient diversity. In other words, deviation index is not only a good tool to measure the degree of diversification, but also a practiced guider to point out the right way to the efficient diversification of export market.

3 Methodology

According to the modern portfolio theory, diversification effectively dissolves the risk of investment. However, diversification is not always act on in practice. Only the multiple structure of diversification is reasonable, is it likely to reduce the risks. Hence, each unit of risk in the portfolio investment, namely "risk unit", must be considered when we measure the performance of diversity, so does the export market diversification. As a result, the market diversification will approach to the reasonable level when "risk units" get into harmonious coexistence.

In light of the above ideas, through the way of investigating the "risk unit" in each specific market and optimizing the market structure, we can promote the market diversity to a reasonable level and reduce trade frictions among export markets. However, there is no existing tool to evaluate "risk unit", which became an obstacle in the way. In order to moving forward, this article employs the deviation index from mathematics. Deviation refers to the absolute value of difference between actual data and target data accounted for the proportion of the target value. The formulas may be expressed as: |X - A|/A; where, A is target value, X represents actual data. According to the ideas of deviation, the "risk unit" of export market, R_{ij}, can be evaluated as:

$$R_{ij} = X_{ij} - \gamma_j X_i \tag{1}$$

Where, X_{ij} is the actual value of country i exports to county j, target value evaluated as $\gamma_j X_i$, and X_i namely the total exports of country i, γ_j is country j's import market share. Obviously, R_{ij} measures the difference between the actual exports and expected value, which including two directions of deviation: excessive and deficient.

These two deviation effects lead to the impact on the target market, and finally induce trade friction.

In order to express the degree of risk's influence in the country's total exports, we modify the mathematical expression of deviation. Hereinafter, we define the deviation ratio of export market (r_{ij}) as the difference between the actual exports and expected value accounts for the proportion of total exports. The ratio can be calculated as:

$$r_{ij} = (X_{ij} - \gamma_j X_i) / X_i \tag{2}$$

On this step, we build up the tool to measure the deviation of "risk unit" in a single export market, that is to say we find the way to figure out the specific structure of each specific market with the measure of r_{ij}. Meanwhile, with the sum of all "risk units" deviation, we get the deviation ratio for the whole export market. Since the effects of excessive and deficient will offset each other in whole "risk units" deviation, we take the sum of the absolute value as the measure for the overall export market. Now the deviation ratio may be written as:

$$d(X_i) = \sum_{j=1}^{n} \left| r_{ij} \right| = \sum_{j=1}^{n} \left| \frac{X_{ij} - \gamma_j X_i}{X_i} \right| = \sum_{j=1}^{n} \left| \gamma_{ij} - \gamma_j \right| \tag{3}$$

Where, $d(X_i)$ is country i's deviation ratio of export market, γ_{ij} is the proportion of exports to country j account for country i's total exports. In fact, $d(X_i)$ including two directions of deviation, namely excessive and deficient, and the absolute value of both directions is adopted in $d(X_i)$. Therefore, $d(X_i)$ takes maximum value, 2, when country i's exports absolutely concentrate on only one market and completely excessive in the market. Contrarily, when export market structure approach to the expected distribution, exports according to the capacity of each import market, $d(X_i)$ takes minimum value. For the purpose of the standardization, multiplied $d(X_i)$ by 0.5, namely:

$$D(X_i) = \frac{1}{2} d(X_i) = \frac{1}{2} \sum_{j=1}^{n} \left| \gamma_{ij} - \gamma_j \right| \tag{4}$$

Accordingly, the value of $D(X_i)$ is constrained between 0 and 1. Export market structure is closer to the expected distribution as $D(X_i)$ closer to the minimum value 0, which implies that export market become more harmonious. On the other hand, it suggests there are many disharmonious factors in export markets while the ratio gets close to the maximum values 1, and finally result in trade frictions. The formula expression of $D(X_i)$ happens to coincide with the explanatory variable S_j of Al-Marhubi's empirical study in 2000, but he directly introduced the absolute deviation from mathematics to study the issue of commodity diversification, without strict derivation on economics argument, nor extends his research to market diversification. As yet, we find no literature used the deviation index to study the issue of market diversification.

Obviously, $D(X_i)$ measures the deviation of the country's export market structure from the world's import market structure, which paying adequate attention to the

capacity of import market. In addition, the index has a good explanation not only in the whole export market but also in each "risk unit". With excessive or deficient deviation, the index shows positive or negative symbols to distinguish each other, which point out the direction to optimize export market structure.

4 Empirical Investigation

4.1 Data

With the deviation index, we take China's export market as an example to apply the index to analyze the export market structure. In order to evaluate the practical effects of China's market diversification strategy introduced in the early of 1990s, the paper adopts the time series data from 1990 to 2011 for both of China's export market structure and world's import market structure. The data source of China's export is from China's statistical yearbook from 1991 to 2012, world import data from the database of WTO.

We should do a subtraction for the world import data before they are used to analyze China's export market diversification. Since China's foreign trade play a significant role in both Asia and the world market, therefore, we should take out China's imports from imports of Asia and the world while discussing China's export market distribution, so as to measure the real import market capacity confronted with China. We calculate the data of China's export market structure and the world's import market structure in Table 1 and Table 2 respectively. The deviation ratio is shown in Table 3.

4.2 Results

We first analyze China's export market structure with data in Table 1 and Table 2, and then carry out market diversification analysis with data in Table 3.

China's Export Market Structure

Compared Table 1 with Table 2, we can make clear the relationship between China's export market structure and world's import market structure. Meanwhile, it's easy to tack the trend of structural changes with the time series data of both structures. From the overview on both structures, China's export market mainly concentrated in Asia, Europe and North America, which account for about 90% of China's total exports. However, this proportion has been decreased continuously since the strategy of market diversification employed in 1992. The proportion decreased from 96.6% in 1991 to 87.6% in 2011. Specifically, the share of Asia in China's exports is greater than the share of Asia's imports in the world, up to 2011, the excessive proportion still more than 17%. On the contrary, the share of Europe in China's exports is less than the imports share of Europe in the world, and the deficient proportion still approach to 20% in 2011. The share of North America in China's exports is initially less than its imports share, but this deficient deviation improved with China's strategy

of export diversity. The share of North America in China's exports stands on a reasonable level in the recent years.

Although the share of other continents is very small, amounting to about 10%, the proportion keeps rising year by year. With effects of diversification strategy, China's exports in Africa and Oceania market grow along with their imports and gradually close to a reasonable level. However, the growth rate of China's exports to Latin America is more than the growth rate of its import market capacity, and the deviation trend was increased in recent five years.

Geographical Diversification of China's Export Trade

We analyze the geographical diversification with the data of deviation ratio shown in Table 3. In order to make the boring data lively, we transform the data set into line chart, as shown in Fig. 1.

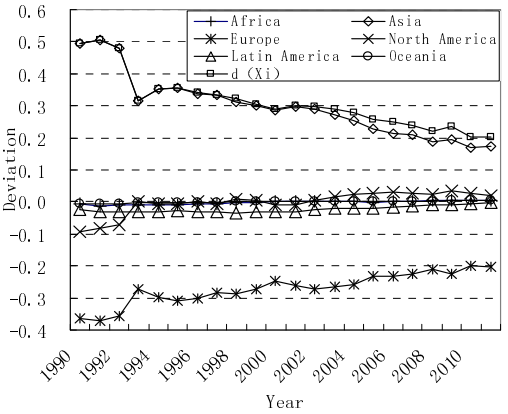


Fig. 1. The deviation ratio of China's export markets

In Fig. 1, the curve of D(Xi) keeps a declining trend from 1992, and fallen sharply in 1993, dropped from 47.8% in 1992 to 31.4% in 1993. Undoubtedly, China's strategy of market diversification has worked on well since employed in 1992, and the deviation ratio dropped to 20.3% in 2011. Specifically, the deviation curves behaved well in four continents in Fig. 1, including Latin America, North America, Africa and Oceania. The deviation curves of those four regions show a convergence trend to zero value. Nevertheless, Asia and Europe go their separate ways, and the curves deviated from the zero value in excessive and deficient direction respectively. Although both of the curves have a trend of convergence to zero value, the value of deviation ratio still stand on a high level, about 20%. With that, we follow up with the discussion in two dimensions, excessive and deficient deviation in exports markets.

a) Markets with deficient deviation: performed in two shows, always deficient and initially deficient.

As shown in Fig. 2, China's exports to Europe displays a long-term deficient deviation, namely always deficient. Specially, the deviation ratio of Europe is still as

high as 20.3% in 2011, which implies that there is huge development space for China's exports to Europe market.

The curves of Africa and North America displayed deficient initially and stable lastly. The deviation ratio of North American sharply improved from 8.8% in 1992 to 1.7% in 1993 and the ratio was significantly close to the ideal value after 2002, which indicates that along with the market diversification strategy launched in 1992 and accession to the WTO in 2001, China's export market structure is optimized gradually.

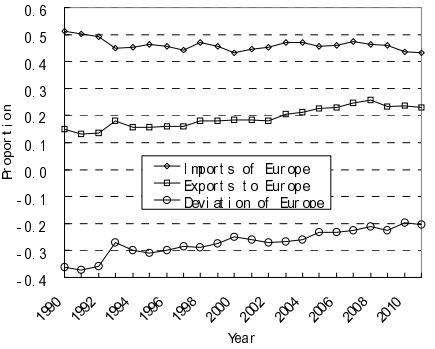


Fig. 2. The deviation in Europe market

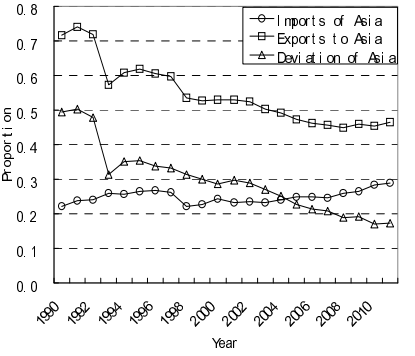


Fig. 3. The deviation in Asia market

b) Markets with excessive exports: also divided into two cases, always excessive and excessive lastly.

In Fig. 3, Asia's curve displayed a long-term positive deviation, namely always excessive. The deviation ratio of Asia is still more than 17% in 2011, which implies that China's exports to Asia should be transferred substantially.

In Latin America, the curve presents deficient initially and excessive lastly. Compared with Table 1 and Table 2, we can find that the growth rate of Latin America's share in China's exports is faster than the growth rate of its import market capacity, as the reason for excessive trend continuously in recent years. The deviation ratio is up from 1 ‰ in 2002 to more than 2% in 2011, which indicates that China's exports to Latin America expanded too fast in recent years. It should be noted such recent excessive deviation trend only appeared in Latin America.

To sum up, except Latin America, the deviation curve of the continents displayed a trend close to ideal value in recent years. Compared with Latin America, the deviation ratio of Asia and Europe is significantly higher, and there is huge improvement space for both markets. In addition, the deviation curves of Africa and North America behave well and stand on the reasonable level stably in recent years.

5 Conclusions and Suggestions

With the results of empirical analysis, we can safely draw the conclusion that China's strategy of market diversity has worked on well since it was employed in 1992, but there is still huge space to improve. Specific conclusions as follows:

- China's export market structure was optimized in recent years, and the exports share of emerging markets keep improving.
- China's Exports to Africa, North America and Oceania are relatively stable in recent years, but the exports to Latin America expanded too fast recently.
- Asia displayed a long-termed excessive deviation, conversely, always deficient in Europe, and there is huge improvement space, about 20%, for both markets.

According to the above conclusions, China's export markets would approach to the efficient diversification finally with the guide of deviation index. To optimize China's export market structure, this paper argues that it should begin with the following steps:

- Take transferring and expanding export strategy in Asia and Europe respectively.
- Optimize the internal market structure in Africa, North America, Latin America and Oceania with the guide of deviation index. In view of the excessive deviation trend of Latin America in recent years, keep the export step with its import step and do not move forward too fast.

In addition, the market structure optimization must be combined with commodity optimization to attain comparative advantage by the self-discovery mechanism of diversification (Samen S., 2010 [13]), which, in turn, promotes China's export diversity strategy forward.

Acknowledgments. This paper is supported by the Graduate Innovation Foundation of Shanghai University of Finance and Economics (NO. CXJJ-2011-354).

References

1. Adelman, M.: Comment on the "H" concentration Measure as a numbers-equivalent. J. Review of Economics and Statistics 51(1), 99–101 (1969)
2. Ali, R., Alwang, J., Siegel, P.: Is Export Diversification the Best Way to Achieve Export Growth and Stability? A Look at Three African Countries. World Bank Policy Working Papers, No. 729 (1991)
3. Al-Marhubi, F.: Export diversification and growth: an empirical investigation. J. Applied Economics Letters 7(9), 559–562 (2000)
4. Attaran, Zwick: Entropy and Other measures of Industrial diversity. Quarterly Journal of Business and Economics 26, 17–34 (1987)
5. Berthelemy, Chauvin: Structural Changes in Asia and Growth prospects after the crisis. CEPII Working papers, No. 00-09 (2000)
6. Ben Hammouda, H., Stephen, K., Angelica, N., Sadni Jallab, M.: Diversification: towards a new paradigm for Africa's development. ATPC Work in Progress, No. 35 (2006)

7. Hirschman, A.: Paternity of an Index. J. American Economic Review 54(5), 761 (1964)
8. Kelly, W.: A generalized interpretation of the Herfindahl index. Southern Economic Journal 48(1), 50–57 (1981)
9. Kingston, J.: Export concentration and export performance in developing countries, 1954-67. Journal of Development Studies 12, 311–319 (1976)
10. Love, J.: Model of Trade Diversification Based on the Markowitz Model of Portfolio Analysis. Journal of Development Studies 15(2), 233–241 (1979)
11. Mac-Bean, A., Nguyen, D.: Commodity concentration and export earnings instability: A mathematical analysis. The Economic Journal 90, 354–362 (1980)
12. Pineres, S., Ferrantino, M.: Export Diversification and Structural Dynamics in the Growth Process: the case of Chile. Journal of Development Economics 51, 375–391 (1997)
13. Samen, S.: A Primer on Export Diversification.Key Concepts, Theoretical Underpinnings and Empirical Evidence. WBIGC (2010), http://blogs.worldbank.org/files/growth/EXPORT_DIVERSIFICATION_A_PRIMER_May2010(1).pdf
14. Matadeen, S.: Export diversification and economic growth: Case Study of a Developing Country Mauritius. In: ICITI 2011 (2011)

Appendix: Data Sets

Table 1. China's export market structure

Year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Africa	0.0209	0.0139	0.0153	0.0166	0.0145	0.0168	0.0170	0.0176	0.0221	0.0211	0.0202	0.0226	0.0214	0.0232	0.0233	0.0245	0.0275	0.0306	0.0358	0.0397	0.0380	0.0385
Asia	0.7175	0.7416	0.7191	0.5736	0.6070	0.6184	0.6041	0.5961	0.5343	0.5262	0.5309	0.5296	0.5261	0.5079	0.4980	0.4809	0.4703	0.4663	0.4642	0.4732	0.4639	0.4736
Europe	0.1501	0.1308	0.1337	0.1790	0.1551	0.1545	0.1580	0.1586	0.1819	0.1820	0.1825	0.1850	0.1790	0.2012	0.2063	0.2174	0.2223	0.2364	0.2400	0.2202	0.2251	0.2179
North America	0.0903	0.0935	0.1088	0.1979	0.1889	0.1764	0.1874	0.1894	0.2182	0.2277	0.2218	0.2166	0.2281	0.2239	0.2245	0.2292	0.2261	0.2070	0.1917	0.1985	0.1938	0.1844
Latin America	0.0126	0.0111	0.0127	0.0194	0.0203	0.0212	0.0206	0.0252	0.0290	0.0270	0.0288	0.0309	0.0291	0.0271	0.0307	0.0311	0.0372	0.0423	0.0502	0.0475	0.0582	0.0641
Oceania	0.0086	0.0090	0.0094	0.0134	0.0142	0.0128	0.0130	0.0131	0.0145	0.0160	0.0157	0.0153	0.0162	0.0166	0.0171	0.0169	0.0165	0.0173	0.0181	0.0207	0.0209	0.0215

Table 2. World's import market structure

Year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Africa	0.0285	0.0266	0.0265	0.0262	0.0247	0.0246	0.0232	0.0237	0.0240	0.0222	0.0200	0.0216	0.0211	0.0222	0.0236	0.0251	0.0260	0.0280	0.0312	0.0349	0.0337	0.0336
Asia	0.2226	0.2383	0.2413	0.2597	0.2566	0.2646	0.2671	0.2626	0.2209	0.2263	0.2445	0.2329	0.2344	0.2332	0.2410	0.2477	0.2477	0.2472	0.2599	0.2660	0.2833	0.2900
Europe	0.5126	0.5012	0.4913	0.4508	0.4546	0.4638	0.4582	0.4434	0.4701	0.4557	0.4305	0.4460	0.4535	0.4722	0.4698	0.4577	0.4618	0.4724	0.4653	0.4586	0.4339	0.4305
North America	0.1958	0.1921	0.1968	0.2150	0.2149	0.1969	0.2016	0.2169	0.2311	0.2479	0.2592	0.2533	0.2484	0.2313	0.2231	0.2238	0.2179	0.2021	0.1884	0.1851	0.1902	0.1846
Latin America	0.0246	0.0265	0.0288	0.0326	0.0327	0.0344	0.0340	0.0381	0.0390	0.0325	0.0319	0.0329	0.0281	0.0257	0.0270	0.0299	0.0317	0.0347	0.0391	0.0383	0.0415	0.0436
Oceania	0.0160	0.0154	0.0152	0.0159	0.0164	0.0157	0.0159	0.0154	0.0148	0.0154	0.0139	0.0132	0.0145	0.0154	0.0156	0.0157	0.0150	0.0155	0.0161	0.0172	0.0174	0.0177

Table 3. The deviation ratio of China's export markets

Year	Africa	Asia	Europe	North America	Latin America	Oceania	D(Xi)
1990	-0.0076	0.4949	-0.3625	-0.0929	-0.0244	-0.0075	0.4949
1991	-0.0127	0.5033	-0.3703	-0.0842	-0.0300	-0.0064	0.5034
1992	-0.0112	0.4778	-0.3576	-0.0711	-0.0331	-0.0059	0.4783
1993	-0.0095	0.3139	-0.2717	0.0009	-0.0312	-0.0025	0.3149
1994	-0.0103	0.3504	-0.2994	-0.0070	-0.0315	-0.0021	0.3503
1995	-0.0078	0.3538	-0.3093	-0.0061	-0.0277	-0.0029	0.3538
1996	-0.0062	0.3369	-0.3002	0.0028	-0.0304	-0.0029	0.3397
1997	-0.0061	0.3335	-0.2848	-0.0073	-0.0331	-0.0023	0.3336
1998	-0.0019	0.3134	-0.2882	0.0105	-0.0334	-0.0004	0.3238
1999	-0.0011	0.2998	-0.2737	0.0052	-0.0308	0.0006	0.3056
2000	0.0003	0.2865	-0.2480	-0.0098	-0.0307	0.0018	0.2885
2001	0.0009	0.2967	-0.2610	-0.0090	-0.0297	0.0021	0.2997
2002	0.0003	0.2889	-0.2716	0.0065	-0.0258	0.0018	0.2974
2003	0.0010	0.2700	-0.2663	0.0161	-0.0221	0.0012	0.2884
2004	-0.0003	0.2520	-0.2584	0.0240	-0.0188	0.0015	0.2775
2005	-0.0006	0.2263	-0.2335	0.0278	-0.0212	0.0012	0.2553
2006	0.0016	0.2147	-0.2315	0.0309	-0.0171	0.0015	0.2486
2007	0.0026	0.2087	-0.2256	0.0266	-0.0141	0.0018	0.2397
2008	0.0046	0.1885	-0.2095	0.0239	-0.0096	0.0020	0.2191
2009	0.0048	0.1932	-0.2244	0.0340	-0.0113	0.0036	0.2356
2010	0.0043	0.1702	-0.1983	0.0257	-0.0053	0.0035	0.2036
2011	0.0050	0.1737	-0.2029	0.0213	-0.0011	0.0039	0.2039

Design and Realization of Ecological Tourism Information System Based on Tianditu Web APIs

Yating Chen and Xiaoliang Meng

International School of Software, Wuhan University,
No. 37, Luoyu Road, Wuhan, Hubei Province 430079, China
xmeng@whu.edu.cn

Abstract. The issues of tourism development and ecological environment protection restrict regional sustainable development, and arouse government's and citizen's concerns. An ecological tourism information system for solving the dilemma is described in this paper. The study focuses primarily on the area of East Lake Scenic Spot of Wuhan City, China. In the Web GIS system, a custom map server with volunteered geographic information is created by using Tianditu Web APIs which provide basic GIS functions and tools. Through the Ecological Tourism Information System, public could search and view the information of the scenic spots, upload their own knowledge about the interests. The system helps citizens to concern the ecological environment problem, monitor the government to take appropriate strategies to develop tourism resource and take practical actions to resolve the sustainable development of ecological environment.

Keywords: Volunteered Geographic Information, Web GIS, Tianditu Web APIs, Ecological Tourism Information System.

1 Introduction

Traditional geographic information systems (GIS) applied to assessment of environment and resources, relied on professional data such as land use/cover, geology, soils, topography, hydrology, and infrastructure. These data are usually collected through field surveys, aerial photography or remote sensing, analyzed by specialists or trained staffs, and disseminated through government agencies. Global positioning system (GPS) techniques are also applied to contribute conventional system of providing guiding information services at popular tourism destination. Professional GIS and GPS means of data updating and acquiring are expensive, slow, and lack of public participation.

Recent years, the issues of tourism development and ecological environment protection restrict the sustainable development of regional and arouse government's and citizen's concerns. On one hand, the traditional methods of providing guiding information services are no longer satisfied all tourist needs, the increasing demand for detailed and individual information of destination has driven conventional tourism system to update to a higher version that could find better solutions. On the other

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 531–540, 2013.

hand, with the fast development of tourism, the ecological environment and tourism resources are likely to be damaged and destructed because of increasing tourists. So the scenic zone management, landscape conservation and environment monitoring, all these factors need to be reasonable assessed and improved. In a third aspect, since the technological advances, the price of equipment deployed global positioning system unit decreases and the approach for web2.0 application varied, the general public who have previously only been on the receiving end of that scientific information now have opportunities to learning GIS and map making tools, and they can create free digital map of the world like OpenStreetMap. The public participate in GIS is defined as volunteered geographic information (VGI) [1].

In this paper, a web GIS framework including the VGI system and social networking is developed to solve the dilemma between tourism development and environment protection. In the system, public play the role of 'senor', they upload the information of their immediate surroundings - the area and environment they are the most familiar with - and thus often make fewer mistakes and can derive higher quality and more frequently updated results than professionally mapped public and private map data layers [2]. This system has the potential to increase their participation in tourism development and ecological environment monitoring, promote their environmental awareness, and provide timely data.

2 VGI in Ecological Tourism Information System

Mobile phones and laptops have already been widely and increasingly distributed in present years. These devices equipped with the GPS receiver, compass, internet access, camera and etc are considered as sensors. And with the help of broadband connections and Web 2.0 tools, a large number of individuals became able to create and share VGI [3]. VGI is known as the geospatial case of user-generated content phenomenon where web users create contents which are later integrated and made available through web sites. We develop the Ecological Tourism Information System through which users can access to geospatial data, as well as advanced mapping and spatial analysis over the internet, so that the VGI functionality can be efficiently utilized [4].

The study focused primarily on the area of East Lake Scenic Spot of Wuhan, China. The East Lake Scenic Area is located in east of Wuhan, it extends over 82 sq.km within which the lake covers 33 sq.km. The broad area of the water makes the East Lake the biggest "city lake" in China. East Lake is one of key national scenic areas. And it was awarded the ISO14001 Environment Quality Certificate. The magnificent scenic area attracts large numbers of tourists. So, taking advantage of the widespread use of Web GIS, when users browse this website as tourists, they not only can view, query and analysis geographical data, but also can remark own information on maps and share the data with others. This method for tourists tracking can effectively increase accuracy and reduce the cost of investigation while gathering information from the geospatial behavior of tourists.

2.1 Basic Geographic Information

The Ecological Tourism Information System should provide basic geographic position information about scenic spot on the map. This basic information includes the scope of the scenic spot, the iconic attractions, the position and name of these attractions, etc.

2.2 Attribute Information of Scenic Spots

In order to retrieve tourism information and environment condition, the system needs the data provided by visitors who are familiar with the scenic area. The data are attribute information of the interests, including the following two aspects.

(1) From the scenic construction terms: a. Scenic type: sightseeing, vacation, eco-tourism, outdoor sports tourism and cultural landscape type; b. Scenic tourism service facilities condition; c. Abundance of scenic tourism resources including natural landscape, cultural landscape resources, customs resources, traditional food resources, cultural resources and handicrafts resources.

(2) From the scenic natural environment: a. Scenic air quality; b. Scenic vegetation coverage; c. Quality of water; d. Quality of white pollution; e. Scenic comfort rating. The information listed is stored by the grading score.

2.3 Different Data between Environment Assessment and Tourism Recommendation

For the environmental assessment and monitoring, the study chooses four attributes of environment conditions - scenic air quality, vegetation cover, water quality and white pollution levels.

Then, considering the type of scenic, scenic tourist comfortable ratings and the environmental evaluation which calculated in front part, the system can be used to recommend tourist hot-spots.

Finally, comparing the results between environment assessment and scenic-spot recommendation, the system provides a comprehensive analysis of the data. As for the places which are rich in tourism resources, high in the number of visitors and good in scenic facilities, we need strengthen environmental monitoring efforts, protect the environment from destruction. As for the places which are good areas of environmental assessment, while their suitable for tourism scores are low, it should strengthen scenic tourist service facilities, heat and vigorously promote tourism here.

2.4 Data Validity

Using VGI to retrieve tourism information and environment condition, visitors need to be able to add a scenic spot on the map and add the attribute information of interest. Therefore, there should be a legal judgment of adding a scenic spot. Since, duplicate of tourists causes data redundancy added, and the legal judgment has three ways:

(1) As visitors select a point on the map and add attractions, the system judges whether the point has been marked. If the spot has been marked, the system will remind user not to repeat to add. If not marked, the new point will be set to an attraction, and its location and name will be stored in the database.

(2) Users can have a marked interest reporting option or a false report on the spot.

(3) Users can resort administrator privileges to supervise whether the additive data is true and reliable.

3 Technical Infrastructuree

Primarily, the website was developed by using Java platform, B/S (browser/ server) mode [5]. It means the system runs on a web server, and tourists can browse and use the system through browser of computer or mobile phone. In particular, 'mash-up' is one feature of this website. The word 'mash-up' in this paper refers to applications that combine data from different sources into new web services in order to produce new meaningful information. And the availability of APIs has fostered the concept of a "mash-up" as the ideal presentation vehicle for VGI by providing a geographical backdrop [6]. The architecture of the system is shown in figure 1.

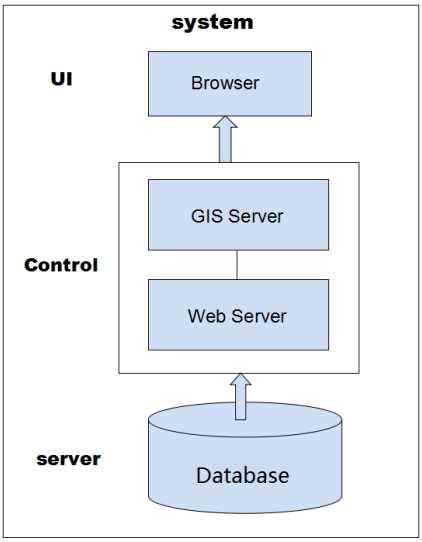


Fig. 1. System Architecture

3.1 Database Server

The internal source of content is represented by the data stored in the system DBMS: MySQL [7]. MySQL is a powerful open source DBMS. With its spatial extension, that adds support for geographic objects to the MySQL database. MySQL Spatial Extensions supports a subset of the openGIS standard, includes a few spatial data types (such as Point, LineString, Polygon, etc.) and some functions. The installation

and configuration of MySQL Spatial are very simple. In fact, it doesn't need to install. The default MySQL configuration will be able to use these spatial data types. MySQL Spatial allowed archival, retrieval and management of all the data received, processed and generated in the Web GIS system.

3.2 Control Layer

In this case, a custom map server with VGI is created by using the Tianditu Map Web APIs [8]. The Tianditu Map Web APIs allow embedding maps in the website, and provides some practical tools to deal with maps in order to realize the system functions such as spatial data display, querying and manipulation.

There are API calls to retrieve map data by bounding box, to create/retrieve change-set, to add, delete and update the three basic elements: node, way and relation. Each of them returns or expects the data for the elements in Map server. The communication between the client application and the on site web server is achieved through asynchronous communication and the JOSN data format [9]. JSON (JavaScript Object Notation) based on a subset of the JavaScript Programming Language is a lightweight data interchange format. JSON is a text format that is completely language independent but uses conventions. It is easy for humans to read and write and for machines to parse and generate. These properties make JSON format an ideal data transmission structure over a network connection. Its main application is in Ajax web application programming. In our system, JSON is used to describe the Geographical coordinates-latitude and longitude, name, ID values and other attributes information, and then these information is packaged via XMLHttpRequest for asynchronous transmission in order to interact with the server data.

The Tomcat is used as web server [10]. Tomcat provides an environment for Java-based dynamic web content. As a container, it provides the implementation of Java Server Pages (JSP) and Java Servlets technology. When you write some web pages in HTML, you reference the JSP code or servlets, then compile them and deploy it to Tomcat. Final, through the browser, web pages cater the request of clients using the Hypertext Transfer Protocol (HTTP).

3.3 User Interface

For data transmission and web page design in the system, Ajax (Asynchronous JavaScript and XML) is the critical technology. It is a way of programming for the communication between the client application and the web server. Data, content, and design are merged together into a seamless whole. In view of WebGIS in the map data transmission and structure of web page design, the use of Ajax is provided with a better user experience. When user select a point or a polygon and choose the “edit”, this action trigger an AJAX driven function, and calling a form in a child window, waiting for a little lag time, the page will displays what they’re asking for. While the request submitted by the user controls, only part of a page need to be reload, that reduces the network flow between client and server-side and the burden on the server. Figure 2 describes the communication process between browser and server.

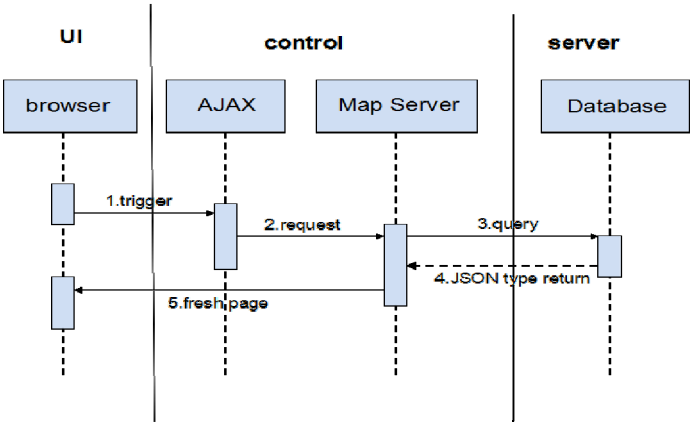


Fig. 2. AJAX Communication between Browser and Server

4 Implementation

The objectives of this study are: (1) Using current technologies to develop a map service website through which public could search and view any points they interested and upload and share their own knowledge such as photos, description and attributes about the circumstance they are familiar with, identified or explored. (2) Analyzing data such as the air quality, water quality and greenery degree, etc, and assessing the development mechanism of ecological environment and local tour route, and studying the sustainable development strategies.

So, the web GIS system realizes the following four functions: 1) In the browser of client, the application can load base maps from Tianditu server. Several key scenic locations were noted and users can view the information about the scenic spots. 2) Users can directly search a place or find a place by zooming and moving, at the same time, users can query the surrounding information about this location. 3) If users are interested in some scenic spots, they can add their own evaluation about the environment condition according to your own common sense and their local knowledge. 4) The system could recommend tourism spots and monitor the ecology environment by calculating the VGI data and analyzing them.

The functions are illustrated in figure 3. The section VGI realizes the function of viewing and adding information. Then the system shows the result of calculating and analyzing through the browser.

In the figure 4, the values of the attribute information of different scenic areas are calculated according to the volunteered geographic data which users upload. And, the selection of the valid data should be in a suitable geographic range. For example, when we need the assessment result of East Lake Scenic Spot, the calculation data should include the marks around the exact location within one kilometer scope.

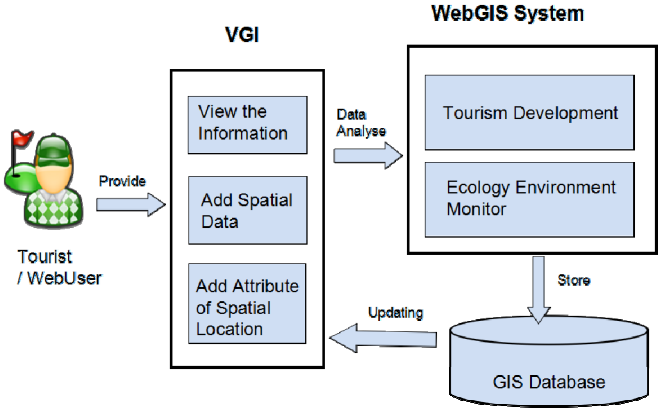


Fig. 3. Web GIS System Implementation

Clicking the "user evaluation" button in the Figure 4, the child window will open as Figure 5. In this figure, you need to complicate these data. This is the most important part of the VGI system. The information includes both aspects of tourism and environment. And except the option of scenic type is multiple-choice, the rest of the clauses are all scoring. The score ranges from 1 to 5, and the score 1 means the condition is awful and 5 means perfect. Then these data should be stored for system to calculate and analyze. When store the attribute information to the database, it is important for the system to store the corresponding location. The location information should include the longitude and latitude.

Fig. 4. The Function of Browsing the Whole Scenic Area

Fig. 5. The Function of User Evaluating

After clicking the submit button, the system transmits the data to server and database update. After a period of time, compare the attribute data in Figure 6. The figure reveals the changing of each attribute. On one hand, three of the most obvious changes are water quality, air quality and the degree of white pollution. The value of the three clauses all declines, which means the environment condition is deteriorating. On the other hand, the value of service facilities also declines, this condition should be report to the office of tourism management, and the service should be improved.

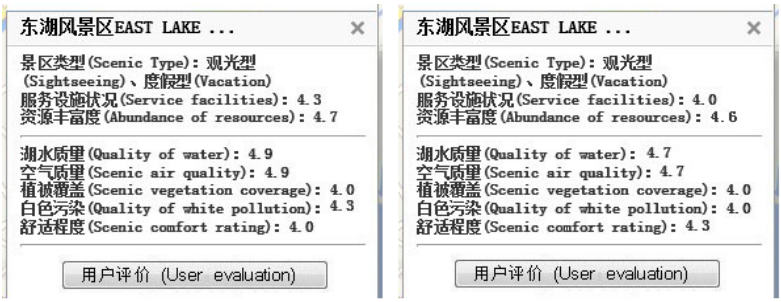


Fig. 6. Compares of Attribute Values

At last, the study tries to analyze the reason of the change. A possible reason is the tourist flow rate. Figure 7 is a statistical chart about the count of searching the scenic spots. The large search count in the second half of March means it is possible that the number of tourists is increasing, and it leads the environment condition to deteriorate. The office of tourism management should pay attention to this.

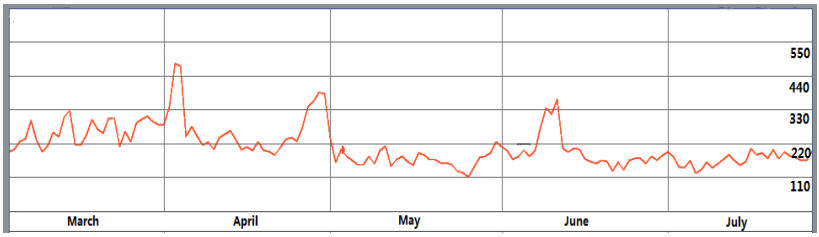


Fig. 7. The Count of Searching the Scenic Spots

5 Conclusion

In this study, the Ecological Tourism Information System integrates traditional GIS, public participation GIS and social networking technology. The objective of the system is efficiently to provide scenic spot information for touring guiding and assess environment condition for environment monitoring. The system utilizes the user's geographic knowledge about the location where they had ever been to provide relevant information about services, facilities or environment. The knowledge of the location of individual can effectively increase accuracy and reduce the cost of investigation while gathering information in traditional GIS system.

The study utilizes Tianditu Map Web APIs and Web Service techniques to develop a useful VGI system designed to meet tourists' demands and to improve landscape conservation, environmental assessment and scenic zone management in the East Lake Scenic Spot of Wuhan, China. Firstly, for the citizens who want to tourism in the beauty spot, they could search their interesting point in the web map by selecting the scenic spot type or considering the score through this VGI system. And then, after they have own experience they could take other information in this map for other reference. This means a virtuous cycle of information. Secondly, this VGI system provides reality and timely data about local environment, analyzes these data, and the result will help citizens to concern the ecological environment problem. And the public influence can monitor government program appropriate strategies to develop tourism resource and take practice action to resolve the sustainable development of ecological environment.

In the follow-up works, this eco-tourism system still needs further improvement: 1) Classification for registered users. The system should set different rights for users, in order to make the data of the system more accurate and reliable. 2) Recalculation of the data for each attribute. Based on user classification mechanism, the system calculates each attribute score by granting them a certain proportion of the weight to arrive at a more reasonable result. The weight is decided by users' level, the high level users' data are more reliable. 3) Filtering data of interests. The system supports the function for 'simple filtering of the data depend on certain geographical scope,' but to face the future more massive amount of data, the system needs to be more reasonable data filtering mechanisms. 4) The system can also add multimedia data, such as dynamic images, audios and videos to achieve more functions.

References

1. Goodchild, M.F.: NeoGeography and the nature of geographic expertise. Journal of Location Based Services 3, 82–96 (2009)
2. Ardagna, C.A., Cremonini, M., Gianini, G.: Landscape-aware location-privacy protection in location-based services. Journal of Systems Architecture 55(4), 243–254 (2009)
3. Goodchild, M.F.: Citizens as sensors: The world of volunteered geography. GeoJournal 69, 211–221 (2007)
4. Zhong, Z.N., Jing, N., Chen, L., Wu, Q.Y.: Spatial data management for WebGIS. Journal of Systems Engineering and Electronics 15, 760–765 (2004)
5. Li, H., Zhang, H., Jin, T.F., Wang, Y.: WebGIS developing architecture and working mechanism based on J2EE platform. In: International Conference on Wireless Mobile and Multimedia Networks Proceedings, ICWMMN 2006, p. 425 (2006)
6. Ho, S., Rajabifard, A.: Learning from the Crowd: The Role of Volunteered Geographic Information in Realising a Spatially Enabled Society. In: Proceedings of the 12th GSDI World Conference, Realising Spatially Enabled Societies, Singapore, October 19-22 (2010)
7. http://www.mysql.com/
8. http://api.tianditu.com/api-new/home.html
9. Luo, Y.W., Wang, X.L., Xu, Z.Q.: Design and implementation of XML-based communication protocols for WebGIS. Ruan Jian XueBao/Journal of Software 15, 899–907 (2004)
10. http://tomcat.apache.org/

A Study of the Structure of China's Mainstream Online Tourism Information Network Based on SNA

Na Feng, Junyi Li, and Gaojun Zhang

College of Tourism and Environment, Shaanxi Normal University, Xi'an 710062 Shaanxi, China
lijunyi9@snnu.edu.cn

Abstract. Taking China's mainstream tourism websites as research objects, the paper introduced Social Network Analysis into studying tourism information network in order to construct the network structure of tourism information websites based on the friendly links among websites. Social Network Analysis method was used to analyze the structure of mainstream tourism information network in China. The study confirms that: firstly, China's online tourism information network is a typical complex network with scale-free. The density of the network is 0.0584, which indicates that the overall network structure is relatively loose; secondly, a few websites play key roles in the tourism information and resources flow and occupy central position in the whole network; thirdly, there is close connection within a group, while the connection between groups is weak. In addition, this paper analyzed the relationship between the characteristics of online tourism information network and the space distribution of tourism websites. The research results provide some theoretical basis for establishing a more reasonable tourism information network structure in China.

Keywords: Tourism informatization, Social Network Analysis, Tourism Website, online tourism information network.

1 Introduction

As an information intensive industry, the tourism industry survives and operates with information as the fundamental basis, which affects consumers' decision-making on destination and their satisfaction of travel experience [1]. Using network and information technology in the tourism industry has unique advantages because of the invisibility and immovability of tourism products [2]. Therefore, the development of tourism industry and tourism informatization is closely connected [3]. With the rapid development of information and communication technologies, the popularity of computer reservation system (CRS), global distribution system(GDS) and the Internet is gradually growing in the tourism industry[4], which has had inevitable impact on the tourism industry and has been touched upon in tourism researches[5]. In practice, every country has been paying great attention to tourism informatization. For example, China

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 541–552, 2013.

began tourism informatization work in the 1990s and launched "golden tour project" in 2003. At present, great progress has been made in tourism informatization.

Tourism informatization, being the basic stage of digital tourism, is of great significance to the development of the tourism industry. It changes the traditional tourism production, distribution, and consumption mechanism by using information and communication technologies, and optimizes the operation of the tourism economy via informatization in order to achieve rapid growth in the tourism economy [6]. Tourism websites, calling system, digital management and infrastructure construction supporting informatization are major manifestations of tourism informatization, while tourism websites are the most widely used, most direct and most effective means of tourism informatization. Tourism informatization mainly includes tourism enterprise informatization, electronic commerce informatization, and e-government informatization which manifest themselves in the form of tourism websites [7]. Many tourism scholars have done a lot of researches into tourism informatization. Buhalisa etc. found that the application of electronic media has huge potential in tourism industry [8]; Du Xiaohui etc. [9] proposed three applicable modes of mobile e-commerce development in tourism and demonstrated the application values of those modes in tourism by analyzing the connotation of mobile e-commerce. With the rapid development of Internet technology, the number of tourism website has soared. At the same time, scholars studied tourism websites and their evaluation model from different perspectives. For example, Zhang Jie [10] discussed the distribution types and regional differences of tourism websites in China from the perspective of space; Lu Zi etc. [11] considered the practicality of tourism websites in China from the perspective of their functions; Li Kai etc. [12] studied Chinese tourism websites from the perspective of the website users, and put forward important suggestions for personalized service provided by tourism websites in China. Cheng Xia and Shan Fang discussed the influence of tourism website on the potential tourists' behavioral intention from the perspective of environmental psychology [13].The development of tourism informatization depends hugely on the transmission of network information. Nowadays, people rely greatly on the network information when travelling, so more and more importance is being attached to the marketing of tourist websites which serve as important for tourism information suppliers [2]. Tourism marketers should make use of the advantage of social media to make up for the functional deficiency of travel industry website marketing [14].

As discussion attests above, the current studies of tourism informatization mostly took a single website as research object, and the researches mostly focused on tourism e-commerce, tourism website and evaluation, and tourism network marketing, but there isn't sufficient research about tourism information network structure from the view of online tourism. Therefore, based on the relationship pattern between websites, this paper introduces the Social Network Analysis method into tourism informatization study, which provides a new perspective to understand and analyze the construction of China's tourism information network structure at present and offers theoretical basis for constructing effective tourism information network structure in China.

2 Data Sources and Research Methods

2.1 Data Sources

This paper obtained TOP 120 tourism websites from ChinaZ.com, and eliminated those unavailable ones. Finally, we selected 112 tourism websites as samples. At the same time, we confirmed that these 112 research samples are representative of tourism websites by searching for "China tourism website", "tourism website" and "tourism website ranking" as keywords in Baidu.com and analyzing the related websites of the first ten pages. The research objects of this paper are the 112 tourism websites and the network structure formed by the linking relations of these websites. For the convenience of analysis, we numbered these tourism websites according to their relative rankings (partly detailed in table 1). We collected data from November 1 to November 7 in 2012 using SocSciBot4 and stored the selected data in excel so that they can be imported into UCINET for analysis.

Table 1. Number the Study Sample (partly)

NO.	Website	NO.	Website	NO.	Website
1	www.ctrip.com	6	www.cncn.com	11	www.daodao.com
2	www.qunar.com	7	www.lvmama.com	12	www.8264.com
3	www.kuxun.cn	8	www.tuniu.com	13	www.uzai.com
4	www.lotour.com	9	www.17u.com	14	www.mafengwo.cn
5	www.elong.com	10	www.17u.cn	15	travel.sohu.com

2.2 Research Methods

Social Network Analysis (SNA), beginning in the 1930s as an important branch of western sociology, is a method and theory of studying social structure as well as a new social science research paradigm [15]. Social Network Analysis method developed in the fields of psychology, sociology, anthropology and mathematics. It aims to show the complicated relationships in the form of certain network configurations, and then explain their significance to individual action and social structure based on these configurations and their changes [16]. The measurement indicators of SNA include centrality, condensed-subgroup and structural holes etc. SNA has been widely applied not only in the field of sociology, but also in management and communication etc. [17]. At the same time, with the breakthrough of multivariate statistical technology and the rapid development of computer technology (the development and utilization of SNA software such as UCINET, PAJEK), SNA has becoming an important and convenient research paradigm. Online tourism information network is a complex network in which the linking relations (ties) between tourism websites (nodes) are mingled with each other, and it can be fully explained by using SNA method.

3 Characteristics of Online Tourism Information Network

The matrix diagram is imported into the Social Network Analysis software UCINET, and the structure diagram of online tourism information network is drawn via the visualization tool NetDraw (figure 1). In figure1, each node is a tourism website, the lines between nodes represent the links of the websites, and the directions of the lines represent the inbound link and outbound link of the websites. The figure shows that the structure of online tourism information network composed of mainstream tourism websites is loose. Among the 112 tourism websites, there are 9 isolated points which account for 8% of the total websites, and 12 points with only one node which account for 10.7%. The network connections between center nodes are intensive, while the network connections between the periphery nodes are thin.

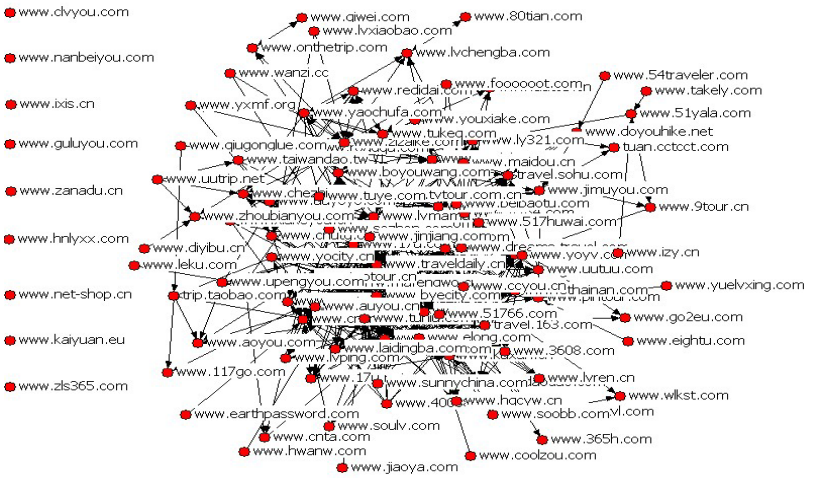


Fig. 1. The Structure of Online Tourism Information Network

3.1 Characteristics of the Whole Structure of Online Tourism Information Network

The whole network is formed by the connection among all members of groups. And its measurement index includes density and average distance etc. The whole structure of online tourism information network can measure the affinity-disaffinity relationship between mainstream tourism websites and the characteristics of community structure etc. Compared with the existing researches on recruitment website density (0.0907)[18], agricultural site density (0.0907)[19], and study abroad consulting website density(0.0002972)[20], the connection between tourism websites is not quite close with the density of online tourism information network being 0.0584 only, and the mutual information resources exchange is not frequent enough. In a word, the network structure formed by the mainstream tourism websites is relatively loose. In this distance matrix, the average distance between nodes is 2.679, which is in line with the

Small-World effect. Citing the six degrees separation principle, we can explain that in the tourism information network, any tourism websites can be accessed through no more than three websites. However, the cohesion index based on the distance is 0.282, which confirms that the network structure is not compact enough and the links between websites should be strengthened.

3.2 Characteristics of the Node Structure of Online Tourism Information Network

One of the main functions of SNA is to identify important nodes in the network and the most important or prominent node always holds a strategic position [21]. Centrality is a major indicator to identify the important nodes in SNA. By analyzing the centrality of online tourism information network, we can decide the position of a tourism website in the whole network information system, and further measure its role in the network structure. The measure indicators of centrality include degree centrality and betweenness centrality betweenness centrality.

3.2.1 Degree Centrality

Degree centrality is used to measure the position and importance of tourism websites in online tourism information network structure, while degree centralization depicts the whole centrality of the network diagram. If "the degree centrality of the nodes differs greatly" in a network, the centripetal tendency is obvious, which illustrates that the centralization of the network structure is quite large and there exist key nodes which have frequent exchange with other nodes [22]. Degree centrality can be divided into absolute degree centrality and relative degree centrality. Absolute degree centrality refers to the number of websites which are connected directly with the tourism website in this paper. For the directed graph, absolute degree centrality includes out-centrality and in-centrality. Table 2 shows the degree centrality value of node in online tourism information network.

Table 2. The Degree Centrality of Online Tourism Information Network Node

No.	Website	OutDegree	InDegree	NrmOutDeg	NrmInDeg
40	www.traveldaily.cn	24.000	20.000	21.622	18.018
13	www.uzai.com	22.000	19.000	19.820	17.117
68	www.ziyou.com	22.000	9.000	19.820	8.108
7	www.lvmama.com	21.000	28.000	18.919	25.225
8	www.tuniu.com	20.000	24.000	18.018	21.622
14	www.mafengwo.cn	20.000	20.000	18.018	18.018
48	www.byecity.com	19.000	16.000	17.117	14.414
56	www.517huwai.com	19.000	8.000	17.117	7.207
19	www.yododo.cn	19.000	21.000	17.117	18.919
6	www.cncn.com	18.000	21.000	16.216	18.919
9	www.17u.com	17.000	19.000	15.315	17.117
10	www.17u.cn	17.000	38.000	15.315	34.234

Network Centralization (Outdegree) = 15.924%

Network Centralization (Indegree) = 28.650%

(1) Degree Centralization

The out-centralization of the tourism information network is 15.924% and the in-centralization is 28.650%, which indicates the degree centralization of the network differs greatly, the centripetal tendency is obvious, and there exist key nodes. However, the in-centralization is bigger than the out-centralization, which reveals that the mainstream tourism websites have the awareness of promotion and pay attention to other websites unconsciously now. But due to the differences between websites, there exists great diversity in website being linked.

(2) Out-centrality

Out-centrality is the number of website outlinks in this paper, which measures the position and the role of a website in the whole network. By analyzing the out-centrality of tourism websites, we can conclude that those websites such as traveldaily.cn, uzai.com, ziyou.com, lvmama.com, tuniu.com, mafengwo.cn, byecity.com, 517huwai.com, yododo.cn, cncn.com, 17U.com, 17u.com, 51766.com have much more outlinks, and have stronger power in controlling tourism information. They play the role of "gatekeeper" in the whole tourism information network, namely, tourism information can be obtained quickly through these links.

(3) In-centrality

In-centrality is the number of website inlinks, which reflects the quality of tourism information and services provided by the website or its prestige approved by other websites. We discovered that 17u.cn was the most authoritative website in the whole tourism information network, and 38 websites among the rest 111 tourism websites have direct link to it. Its standard in-centrality was 34.234%, which showed that 17u.cn has great influence in the whole tourism information network. Being recognized as the leader of online travel service providers by other tourism websites in China, 17u.cn has gained certain popularity in the whole tourism information network. In other words, 17u.cn occupies a central position in tourism information network and is believed to be able to provide tourism information and service of higher quality. The standard in-centrality of lvmama.com, tuniu.com, elong.com, cncn.com, yododo.cn, 51766.com, mafengwo.cn, traveldaily.cn, qunar.com, 17u.com are all above 17%, which means these are the authoritative websites and are recognized by other tourism websites in the whole network. These websites play important roles in posting information and sharing resources. In addition, the in-centrality of other 26 tourism websites is 0, accounting for 5% of the total, which further illustrates the centripetal tendency of in-centrality is obvious.

3.2.2 Betweenness Centrality

By analyzing the betweenness centrality of online tourism information network, we can find out which tourism websites are in the position of "bridge", which means the position of tourism websites at the key path of the tourism information and resources flow. The higher betweenness centrality a tourism website shows, the greater control it has over the tourism information and resources. Betweenness centralization displays the gap between the tourist websites with the highest betweenness centrality and other websites. The greater the gap is, the higher the betweenness centralization is. The analysis results of betweenness centralization are shown in table 3. That the network's

betweenness centralization is 8.41% indicates that the network's connectivity is relatively good and the tourism resources and information flow smoothly relatively in the network. Traveldaily.cn, with the highest betweenness centrality, has the strongest power in the whole information network. And being the most important "portal website" in information network, it acts as an intermediary, which brings it considerable profit. Tourism information providers may put much more information onto this website, making it grasping more information flow and business opportunities. Among the tourism websites with high betweenness centrality, the in-centrality (7.207%) and the out-centrality (5.405%) of maidou.cn are relatively low, which indicates that those tourism websites linked to maidou.cn are mostly authoritative ones, while Maidou.com simply acts as a "pure" intermediary in the network.

Table 3. The Betweenness Centrality Index of Online Tourism Information Network Node

NO.	Website	Bet	nBet	NO.	Website	Bet	nBet
40	www.traveldaily.cn	1137.205	9.314	6	www.cncn.com	458.366	3.754
10	www.17u.cn	818.372	6.702	82	www.maidou.cn	436.598	3.576
78	www.zizaike.com	795.645	6.516	8	www.tuniu.com	422.882	3.463
110	www.boyow.com	618.471	5.065	9	www.17u.com	365.468	2.993
7	www.lvmama.com	568.226	4.654	48	www.byecity.com	312.716	2.561
14	www.mafengwo.cn	528.496	4.328	68	www.ziyou.com	303.889	2.489

Network Centralization Index = 8.41%

All in all, a few websites being in the core of the whole network play key roles in the tourism information and resources flow in the online tourism information network. as a "portal" website in the online tourism information network, traveldaily.cn acts as an "intermediary " and helps tourists seek information quickly. 17u.cn is recognized by other tourism websites as the most authoritative website in information network. At the same time, 18 tourism websites with relatively high centrality index are in the central area of the whole network and dominate the whole information flow. In order to verify that the 18 tourism websites have great influence on the online tourism information network, we removed the 18 tourism websites and the average distance obtained in the new network was 3.497. Whereas the average distance in the network is 2.679 when the 18 websites are present. That is to say, the absence of these sites increases the distance of the communication in the whole network by almost one website.

3.3 Small Groups

In the online tourism information network formed by tourism website link relations, a small group refers to a secondary group formed by some tourism websites with close relationship. From the previous analysis, we can found that there existed small world effect in the online tourism information network. Then are there small groups in it?

In order to eliminate the influence of isolated nodes on small groups, the isolated nodes are removed in the process of analyzing small groups in this paper. Based on the

UCINET analysis, the whole network can be divided into 8 groups. The density matrix is in table 4.

The members of the small group A: 1、2、3、4、5、6、7、8、9、10、67、40、13、14、29、44、17、74、19、20、21、50、23、52、28、61、48、45、64、104、70、38.

The members of the small group B: 30、26、11、85、51、46、18、12、41、84、92、16、73、42、80、34、24、62.

The members of the small group C: 103、93、15、54、33、35、57、55、91、31。

The members of the small group D: 81、58、39、22、86、27、56、68、71、76、43.

The members of the small group E: 25、87、49、37、79、66、47、82、69、78、72、110、36.

The members of the small group F: 77、59、53、90、105、89、108、106、75.

The members of the small group G: 32、95、88、102、101、65、111、94、100、99.

The members of the small group H: 98、97、60、107、63、109、96、83、112.

The group A is the largest one with 32 members, accounting for 28.5% of the whole tourism information network members. The density table shows that the densities of group A and group E are 0.289 and 0.179 respectively, which are far greater than the density of the whole online tourism information network, indicating that the link relations between the members of the two groups are quite close. Furthermore, key websites No.5, 6, 7, 8, 9, 10, 13, 14, 19 and 40 are all in group A, which brings close connection to this group. In group E, websites No.87(tukeq.com), No.47(roadqu.com), No.66(autob.cn), No.79 (fooooot.com), and No.75 (redidai.com) are all community websites, No.49 (youbibi.com) is the best tourism search engine in China, No.69 (ly321.com) is a tourism information site, and No.72 (yaochufa.com) is a professional website providing limited-time sales of high-end hotels and vacation products. These 8 websites form a closely connected small group, which suggests that China's tourism websites tend to get linked to different tourism websites based on their different purposes.

Table 4. The Density Matrix of Small Group

NO. of Small Group	A	B	C	D	E	F	G	H
A	0.289	0.082	0.047	0.045	0.017	0.000	0.003	0.000
B	0.115	0.046	0.039	0.051	0.056	0.000	0.006	0.000
C	0.031	0.039	0.011	0.027	0.008	0.000	0.000	0.000
D	0.151	0.096	0.109	0.082	0.042	0.010	0.000	0.000
E	0.043	0.043	0.015	0.035	0.179	0.111	0.008	0.000
F	0.024	0.006	0.022	0.020	0.154	0.000	0.011	0.000
G	0.006	0.006	0.000	0.009	0.023	0.022	0.033	0.000
H	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Among the 8 small groups, group F and group H have the smallest density and there is little linking relation between them. Compared with the whole network, the interactive connection between them can almost be ignored. By comparing the densities between different small groups and those within each small group, we can find that most information and resources flow occurs within groups in tourism information network, while the connection between groups is weak.

4 The Relationship between Online Tourism Network and the Website Features

The data used in this paper is the link connections among tourism websites, so the website rankings determined by the academic indicators (degree centrality, betweenness centrality) can basically reflect the recognition between websites. At the same time, the comprehensive rankings of tourism websites provided by ChinaZ.com are determined by the alexa rankings, Baidu weights and PR index, therefore, they can basically reflect which websites are recognized by users.

4.1 The Correlation Analysis of Website Rankings

In order to analyze the correlation of tourism website rankings between website acceptance and user acceptance, this paper makes an analysis of the tourism comprehensive rankings and centrality rankings of tourism websites. The results show that the related spearman coefficients of comprehensive ranking and in-centrality, out-centrality, betweenness centrality are 0.614, 0.743, 0.602 respectively, which have significant positive correlation. But there also exist an obvious difference ($P<0.01$). The correlation between the comprehensive ranking and the in-centrality ranking of websites is the most positive, which shows that those websites with high popularity are more easily recognized by other websites. However, the authoritative websites recognized by users are somewhat different from those recognized by websites. For instance, ctrip.com enjoys the highest user's recognition, but it doesn't have an outstanding performance in this study with its in-centrality, out-centrality, and betweenness centrality being in the upper-middle position only.

4.2 The Online Tourism Network Structure and the Spatial Distribution of Tourism Websites

Studies have explored the regional differences of the spatial distribution of tourism websites in China as well as the characteristics of spatial scale structure. It is shown that tourism websites are mainly developed in capital cities and tourism developed cities, and the number of tourism websites decreases from coastal areas of East China to Central and West China [10]. In this paper, a new perspective is adopted to analyze the relationship between the characteristics of online tourism information network and the spatial distribution of tourism websites. A comparative analysis has been conducted by searching online for the cities where mainstream tourism websites are located.

The results are as follows: (1) Among the 18 tourism websites with higher centrality, Beijing boasts 7, Shanghai 5, Guangzhou 1, Xiamen 2, Suzhou 2, and Nanjing 1. Due to the differences of economic and technological development, China's tourism websites with higher quality are located mostly in the eastern coastal cities; (2) There doesn't appear any small group formed on the basis of geographical relations. As the cyberspace is not under the constraint of time and space, the links between mainstream tourism websites are not limited by space also. No regional exclusive occurs in the small groups based on the link relations, which are formed by various types of websites for a certain purpose.

5 Conclusion

The development of tourism is inseparable from informatization and the development of tourism informatization can greatly promote the development of tourism. This paper used a new study paradigm—Social Network Analysis to measure and analyze the online tourism information network formed by link relations between tourism websites and the conclusions are as follows:

(1)The whole structure of online tourism information network is relatively loose and the website density is low, which indicates that there is a little connection between tourism websites. And the average distance between websites is short, which is in line with the small world effect. As little obstacle exists between tourism websites, the users can easily access a tourism website from another through less websites in between. In addition, most tourism websites already have developed the awareness of promotion, but not enough practical action has been taken. Therefore, in order to enhance the information and resources flow between websites, all the websites should strengthen their mutual relations.

(2) In the online tourism information network, a few websites being in the core of the whole network play key roles in the tourism information and resources flow. Traveldaily.cn is a "portal" and an "intermediary" in the tourism information network, and through it, the information seekers can quickly find the information they need. Ctrip.com occupies the most important position in the whole network. It has gained certain popularity in the whole tourism information network, and can provide tourism information and service of high quality.

(3)The whole network can be divided into 8 small groups. There is close connection within each group, while the connection between groups is weak though there is some information flow between groups. In addition, no regional exclusive region exclusion occurs in the small groups

(4)The websites with high popularity are more easily recognized by other websites, but the key websites recognized by users are somewhat different from those recognized by websites. Meanwhile, China's tourism websites with higher quality are located mostly in the east coastal cities

In this study, we have had a basic understanding of the characteristics of the current online tourism information network structure. And the findings of this paper will be of some help in actual practice. In order to enhance the influence of their tourism websites,

webmasters should pay attention to establishing links with those key websites; tourism marketers should put their tourism marketing information onto appropriate tourism websites to attract more tourism consumers; the government should amalgamate some similar tourism websites so that the jumbled tourism information and resources can be reasonably integrated to help tourist information users obtain information efficiently.

Acknowledgement. We gratefully acknowledge National Natural Science Foundation of China, No. 41001077,China Postdoctoral Science Foundation (to LI Junyi) No.2012T50794, No.2011M501442 and Postgraduate Training Innovation Fund of SNNU (to FENG Na) , No. 2013CX015 contributions to this paper.

References:

1. Wang, Z.: The development Informatization and tourism industry. Commercial Research, 114–120 (2011)
2. Zhang, B., Yan, Y.: A Review on the domestic and Overseas Tourism Information Research. Geography and Geo-Information Science 28, 95–99 (2012)
3. Yuan, J., Chen, Z.: The Development status, problems and countermeasures of tourism informatization in China. Journal of Changsha Railway Institute (Social Science Edition) 10, 166–168 (2009)
4. Liang, Y., Luo, Z.: The research of International tourism informatization. Tourism Tribune 27, 4–5 (2012)
5. Li, J., Zhang, L., Sun, J., Yang, M.: Tourism Information Science: A Research Framework. Tourism Tribune 26, 72–79 (2016)
6. Liao, K.: A Review on the Tourism Information Research in China. Guide to Business, 47–48 (2009)
7. Fu, X.: A literature review on Tourism informatization research. Estate Observation (2), 95–96 (2009)
8. Buhalisa, D., Licata, M.C.: The future e-tourism intermediaries. Tourism Management 23, 20–22 (2002)
9. Du, X., Zhou, L., Duan, J.: The Application Mode and Market Innovation of Mobile E-commerce in Tourism. Business Economics and Administration, 49–52 (2006)
10. Cheng, S., Zhang, J., Liang, Y., et al.: Study on the Spatial Distribution and Dynamic Mechanism of China' s Tourism Websites. Tourism Tribune 24, 75–80 (2009)
11. Lu, Z., Fan, L.: An Evaluation in Service Functions and Development Strategy for Small-Medium Size Tourism Website——the Case of Leyou Outdoors Club. Human Geography, 103–106 (2005)
12. Li, K., Wang, X.: A Mechanism Study on the Impact of Privacy Concern on the Personalized Service of Travel Websites. Tourism Tribune 26, 80–86 (2011)
13. Cheng, X., Shan, F.: Impact of Featured Tourism Websites on Behaviors of Potential Tourists. Journal of Sichuan Normal University (Social Sciences Edition) 38, 63–69 (2011)
14. Zheng, X., Ulrike, G.: Role of Social media in online travel information search. Tourism Management 31, 179–188 (2010)
15. Wang, S., Hu, R., Cheng, W.: A Literature Review on Overseas Tourism Research under the Perspective of Social Network. Tourism Tribune 24, 90–95 (2009)
16. Shao, Y., Ou Yang, Q., Sun, L.: Social Network Analysis and Its Application to Innovative Research. Chinese Journal of Management 6, 1188–1203 (2009)

17. Xu, Y., Qian, J., Yuan, Q., et al.: Communication of Negative Word - of - Mouth in Micro - Blogging. Journal of Intelligence 31, 6–10 (2012)
18. Zhang, X.: Social Network Analysis of hyperlinks of Chinese recruitment website. Journal of Intelligence Research, 95–98 (2011)
19. Jia, J., Yan, X.: The Study of Links of Agriculture website. Information Science 29, 1882–1888 (2011)
20. Hu, Y., Shao, B.: Social Network Analysis of Link Structure of Study Abroad Consulting Website. New Century Library, 52–56 (2010)
21. Shih, H.-Y.: Network characteristics of drive tourism destinations: An application of network analysis in tourism. Tourism Management, 1029–1039 (2006)
22. Kang, W.: Analysis of the Key Nodes in Public Opinion Spread during Emergencies Based on Social Network Theory ———A Case Study of the 7·23 Wenzhou High - speed Train Collision. Journal of Public Management 9, 101–111 (2012)

Study on GIS's Application in Driving in the Unstructured Environment for UGV

Qiangrong Yang[1,2] and Meiling Wang[1,2]

[1] School of Automation, Beijing Institute of Technology, Beijing, 100081, China
wangml@bit.edu.cn
[2] Key Laboratory of Intelligent Control and Decision of Complex System, Beijing, 100081, China

Abstract. For UGV, driving in the unstructured environment safely and quickly without human intervention becomes increasingly important. While many scholars have conducted researches in driving in this case, the results seem quite unsatisfied. In this paper, how to use GIS in aiding the UGV to drive in the unstructured environment is studied. First, GIS is used to process the raw data received from the 3D laser scanner to get the high-resolution gradient and elevation grid, which is used to determine the traversability of the environment. Then the relatively low-resolution gradient grid is used as the cost field in the path planning algorithm to generate the optimal path from the start point to the goal point. At last, an algorithm is introduced to get the reference speed limit from the given optimal path according to the geometry features of the path and the gradient grid. Simulation and experiment results are given to show the effectiveness of the algorithms.

Keywords: Traversability, UGV, Cost Field, Path Planning, Speed Limit.

1 Introduction

UGV is the main study target for the future transportation system and the future military application. Although the performances of UGV become much better in many areas, it is still a challenging problem for UGV to drive in an unstructured outdoor environment. In this paper we consider the following navigation task for UGV: given the start point and the goal point, a UGV should be able to drive through the unstructured environment and reach the goal point safely and quickly without human intervention. The key to this task is to determine the traversability of the outdoor environment and choose the optimal path from the start point to the goal point.

A standard approach to determine the traversability of the outdoor environment used ranging sensors such as stereo vision or 3D laser scanner [1]. Features of the terrain such as slope, roughness, or discontinuities were then analyzed to determine the traversable regions [2]-[4]. Additional visual cues such as color, shape, and height above the ground were employed in [5] - [7]. Although these methods can be effective in some tasks, UGV's capabilities are not taken into consideration, which results in that these methods are not general to all UGVs.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 553–564, 2013.

The orographic factor, such as the obstacle, gradient and elevation, is the main factor that needs to be considered when UGV is driving in the unstructured road. And the orographic factor can be presented well by the grid model in GIS, which can be used as the cost field [8]. By using the cost field grid model, the optimal path algorithm is searching for the path that has the minimum cost from the start point to the goal point. The gradient grid is chosen as the cost field model in this paper, which means that the optimal path is the flattest path.

The decision system of UGV can only get the global information from GIS, as other systems just provide local data. So it's very important for GIS to give a speed limit for UGV by utilizing the global information stored in GIS. In this paper, an algorithm to calculate the speed limit for UGV is introduced. Three parameters are mentioned and calculated to give a reference speed for the decision system.

The organization of this paper is presented as follows: the traversability analysis based on point cloud data are studied in section 2. In section 3, a path planning algorithm based on the cost field is introduced. After getting the optimal path in section 3, some parameters are given and calculated to give a reference speed limit by analyzing the optimal path in section 4. In section 5, simulation and experiment results are given and discussed, followed by the conclusions presented in section 6.

2 Traversability Analysis

2.1 Rasterizing of the Point Cloud Data

In order to get the isosurface and the contour, we need to get the high resolution elevation and gradient grid of the environment by rasterizing the point cloud data received from the 3D laser scanner. In this paper, SuperMap GIS is used to generate the high resolution elevation and gradient grid. The generation process includes the following 3 steps[9]:

Step 1: Set the field of the 3D point (x, y, z): the field of x is SmX; the field of y is SmY; the field of z is Top.

Step 2: Transform the 3D points to the elevation grid. In this step, we can get the high-resolution elevation grid and the cost field grid (a relatively low-resolution gradient grid), which will be used in Section 3.

Step 3: The gradient is the angle between the tangent plane of a point and the horizontal plane. The gradient grid is abstracted from the elevation grid in step 2. In this step, we can get the high-resolution gradient grid.

2.2 Contour and Isosurface

Generation of the Contour. The contour is the line that connects the points which have the same values (such as altitude, temperature, air pressure). From these contours, a general sense of the terrain can be determined: the denser the contours are, the more dramatically the value of the grid changes. The generating algorithm is given below:

Step 1: Get the DEM grid. The method of getting the elevation grid from the point cloud data is given in section 2.1. Based on the UGV's capacities, the width of a single grid is set to be 0.5m. The DEM elevation grid is shown in Fig. 1.

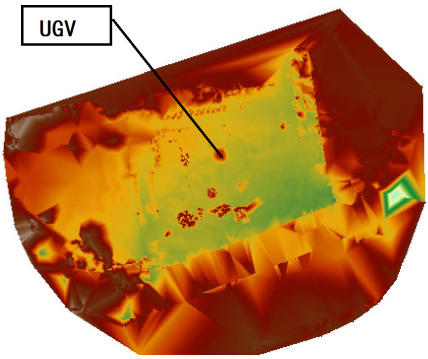


Fig. 1. The DEM elevation grid generated from the point cloud

Step 2: Get the contours from the DEM elevation grid. Set the reference value and the interval to get the contours. In this paper, the reference value is set to be the altitude of the UGV (z_{gps}) and the interval is set to be 0.5m,which means that the elevation difference between the two adjacent contours is 0.5m. The contour map is shown in Fig. 2.

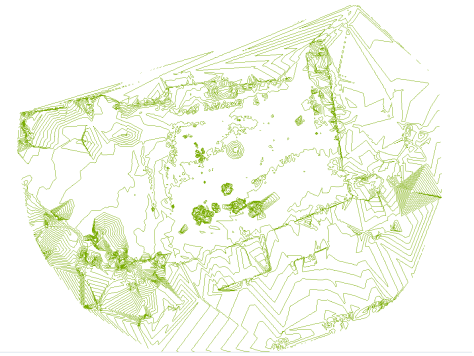


Fig. 2. The contour map

Generation of the Isosurface. Isosurface is the surface that is enclosed by two adjacent contours. It can directly represent the changing rate of the value as contours: the narrower the Isosurface is, the more dramatically the value changes. The Isosurface map is shown in Fig. 3.

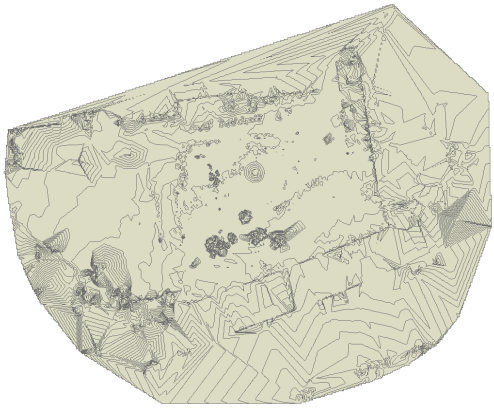


Fig. 3. The Isosurface map

2.3 Traversability Analysis

The isosurface map can directly present the change of the elevation. But for UGV, many isosurfaces are useless when analyzing the traversability of the environment. In this section, the traversable part of the environment is extracted by considering the elevation difference and the gradient. And the results is shown in

Fig. 4, the blank part is the traversable area of the environment.

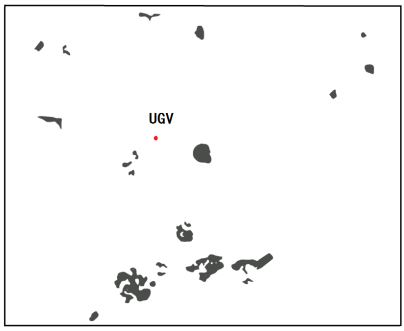


Fig. 4. The traversable area of the environment

3 Optimal Path Planning Algorithm Based on the Cost Field

3.1 Cost Field Model

The grid cost model of a point p is expressed in the following formula:

$$c=f(x, y) \quad . \tag{1}$$

where x and y are the horizontal coordinates of the point p. c is the cost value of the point p. And in the cost field model, the cost of a path from point A to point B is expressed in the following discrete system:

$$c = \sum_{i=1}^{n} f(x_i, y_i) \quad . \tag{2}$$

where c is the cost sum of these discrete points p_i (i=1, 2...n) from A to B and the optimal path is the path that has the minimum cost of all the paths from A to B.

3.2 Optimal Path Planning Algorithm

The Correction of the Gradient Cost Field Model. In the optimal path planning algorithm based on the cost field, the width of road is an important factor [10]. Different width of road brings different cost. In order to calculate conveniently, we assume that the width of the road (W_r) equals to the width of UGV (W_u). The grid cell is square and the length of the square is L. Fig. 5 shows the condition $W_r > L$. In this condition, the road covers several cells and since the complexity of the algorithm increases. So in this paper we consider the conditions as shown in Fig.6 ($W_r < L$).

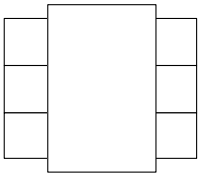


Fig. 5. The condition: $W_r > L$

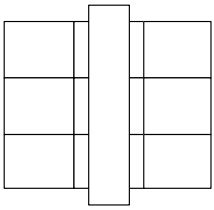


Fig. 6. The condition: $W_r < L$

In this paper the low resolution gradient grid map is chosen as the cost field and the traversability analysis results mentioned in section 2.3 are used to correct the cost field. Fig. 7 shows the gradient cost field grid. As the color is deeper, the value of the grid cell increases.

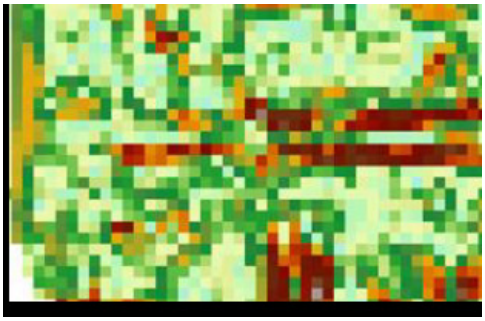


Fig. 7. The gradient cost field grid

Path Planning Algorithm. It is known that the popular path planning algorithm Dijkstra algorithm [11] utilizes the network dataset to get the optimal path. In order to use the Dijkstra algorithm, we need to transform the cost field grid into network dataset to get the path. The transforming method is described in the following:

As shown in Fig. 8, there are 8 directions of a cell. Every center of a cell is a node, and the line which connects two nodes is a link. The value of a link is determined by the cost of the two nodes that the link connects.

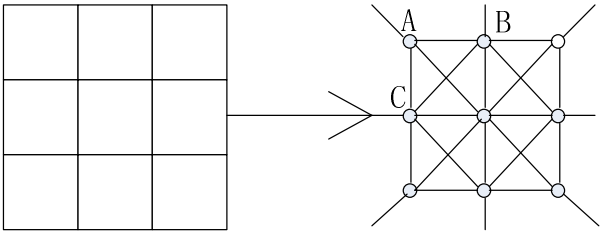


Fig. 8. Transform grid to network dataset

There are two conditions [12] to calculate the value of a link. One condition is that two nodes are close, like link AB and link AC. The cost of this condition can be calculated as equation (3). The other condition is that two nodes are located in the diagonal, like link BC. The cost of this condition can be calculated as equation (4).

$$c=\frac{c_1+c_2}{2} \quad . \tag{3}$$

$$c=\frac{\sqrt{2}(c_1+c_2)}{2} \quad . \tag{4}$$

where c_1 and c_2 is the cost value of the two nodes respectively.

4 Speed Limit for UGV

The decision system of UGV can only get the global information from GIS, while other systems just provide the local data. So it's very important for GIS to give a speed limit by utilizing the global information stored in GIS. In this section an algorithm to calculate the speed limit for UGV is introduced.

In order to get the speed limit, preprocessing of the optimal path given in section 3 is needed. The parameters extracted from the path data include the angle error and the distance to the bend [13]. Some corresponding key points need to be found on the path to calculate the above parameters. These key points are the angle error point and the bend point.

The angle error point $P_{ang}(x_{ang}, y_{ang})$ is defined as the point on the reference path with some distances from the present point, which can be expressed as equation (5):

$$\sum_{i=now}^{ang} \| x_{i+1} - x_i, y_{i+1} - y_i \| \geq d_{ang} \quad . \tag{5}$$

The bend point refers to the first point ahead that the direction changes greatly, which can be defined by equation (6):

$$| \arctan(\frac{x_{bend+10} - x_{bend}}{y_{bend+10} - y_{bend}}) - \arctan(\frac{x_{bend} - x_{bend-10}}{y_{bend} - y_{bend-10}}) | > T_{angle} \quad . \tag{6}$$

After getting the above key points, the parameters for the speed limit can be calculated. These parameters are angle error, the distance to bend and the gradient.

Angle error E_{ang} is the angle between the vehicle's heading Φ and the path's heading, which can be specified as equation (7):

$$E_{ang} = \sum_{i=now+1}^{ang} f(i) \arctan(\frac{x_i - x_{now}}{y_i - y_{now}}) \text{-} \Phi \quad . \tag{7}$$

where

$$f(i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(i-ang)^2}{2\sigma^2}}, \sigma^2 = 0.25 \quad . \tag{8}$$

Distance to bend D_{bend} is defined as the distance from the present point to the bend point, which can be used to help the vehicle to slow down in advance before sharp corners. D_{bend} is defined as the equation (9):

$$D_{bend} = \sum_{i=now}^{bend} \| x_{i+1} - x_i, y_{i+1} - y_i \| \quad . \tag{9}$$

Gradient g is the maximum gradient before the vehicle, which can warn the vehicle to slow down in advance when there is a steep slope before. And the gradient value can be gotten from the high-resolution gradient grid mentioned in section 2.

After calculating all these parameters, we can give the speed limit to help control UGV. The speed limit is given in equation (10):

$$V = V_1 - \frac{V_2}{2}[\theta * \sin g + (1-\theta)\sin(E_{ang})] - V_3 \frac{(100 - D_{bend})}{100} \quad . \tag{10}$$

V_2, V_3, θ are parameters that change the relative weight of the three parameters for the speed limit according to the driving condition and the control need. In this paper, $V_1 = V_2 = V_3 = 40\text{m/s}, \theta = 0.5$, which means the three parameters have the same weight for determining the speed limit.

5 The Experimental Results

5.1 Path Planning Algorithm

UGV used in this experiment is shown in Fig. 9 and the start point is S (3, 3), the goal point is E (39, 24) as shown in Fig. 10 .

Fig. 9. UGV used in this experiment

Using the Dijkstra algorithm for the gradient cost field grid without the correcting data from the traversability analysis, we can get the result which is expressed in a black line in Fig. 10. However, the cells within the circle 1and 2 are not traversable.

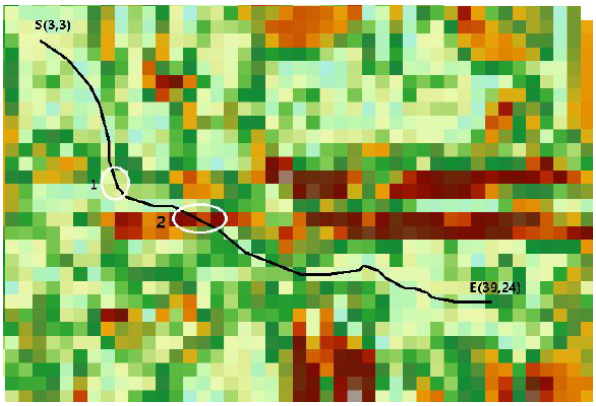


Fig. 10. The result without correcting data

As shown in Fig.11, all the cells that are not traversable are labeled blank. After considering the correcting data, a better result is shown in Fig.11.

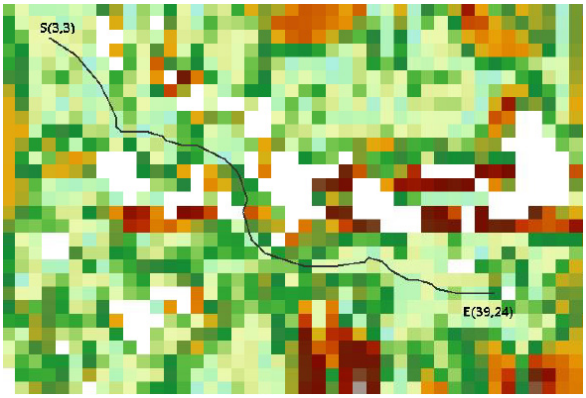


Fig. 11. The result with the correcting data

5.2 The Speed Limit

In order to show the results vividly, we choose the following two paths as reference paths to calculate the speed limit:

By using the algorithm mentioned in section 4, we can get the result which is shown in Fig. 13 and Fig.14. As is shown in Fig. 13, the path is very smooth and the changing angle is relatively small, since the speed limit is nearly the same along the path except for the points near the goal point C, because we need to slow down when getting to the goal point. But in Fig.14, there are 4 obvious road sections that are

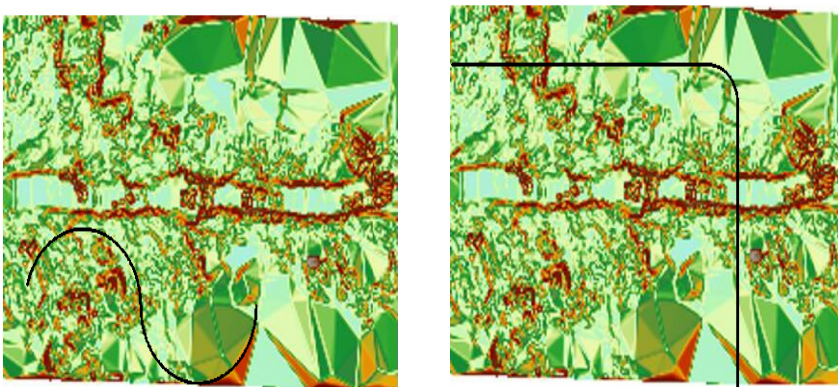


Fig. 12. The reference paths

suggested to slow down in advance. Circle 1 and circle 3 are suggested to slow down because the slope is steep as shown in Fig.12. Circle 2 is suggested to slow down because it's a very sharp (nearly 90 degree) corner. Circle 4 is the goal point, so we need to slow down in advance.

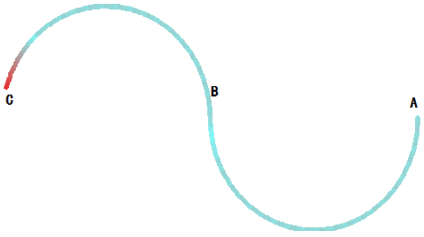


Fig. 13. The result of the arc path

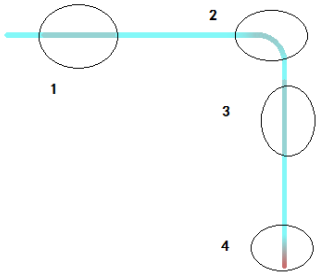


Fig. 14. The result of a right angle path

6 Conclusions

In this paper, applications of GIS in aiding UGV to drive in the unstructured environment are discussed. First, GIS is used to process the point cloud data received from the 3D laser scanner to get the high resolution elevation and gradient grid, contour and isosurfaces, which are used for the traversability analysis of the environment. Then a path planning algorithm based on the cost field which is generated from the elevation grid is introduced. The algorithm also utilizes the editing ability and the existing function of the SuperMap GIS. At last we propose an algorithm to give the speed limit for UGV by analyzing the reference path and the gradient grid. The algorithm can give an appropriate result quickly. The future work includes that the results of the vegetation classification should be taken into consideration in the traversability analysis and the parameters V_2, V_3, θ should be able to get changed dynamically based on the control need and the driving environment.

Acknowledgement. The work was supported by the Major Research Plan of the National Science Foundation of China (91120003) and Surface Project of the National Natural Science Foundation (61173076), the National Natural Science Youth Foundation of China (61105092), and Program for New Century Excellent Talents in University of China (NCET-10-0046). The Authors would like to thank the referees for their valuable and helpful comments which have improved the presentation.

References

1. Kim, D., Sun, J., Sang, M.: Traversability Classification using Unsupervised on Line Visual Learning. In: Proceedings of the 2006 IEEE International Conference on Robotics and Automation, pp. 518–525 (2006)
2. Pagnot, R., Grandjea, P.: Fast Cross-country Navigation on Fair Terrains. In: IEEE International Conference on Robotics and Automation, pp. 2593–2598 (1998)
3. Rieder, A., Southhall, B., Salgian, G., Mandelbaum, R., Herman, H., Rander, P.: Stero Perception on An Off-road Vehicle. In: Proceeding of the Intelligent Vehicles, San Jose, CA (2002)
4. Singh, S., Simmons, R., Smith, T., Stenz, A., Verma, V., Yahja, A.: Recent Progress in Local and Global Traversability for Planetary Covers. In: IEEE International Conference on Robotics and Automation, pp. 1194–1200 (2000)
5. DeSouza, G.N., Kak, A.C.: Vision for Mobile Robot Navigation. IEEE Transaction on Pattern Analysis and Machine Intelligent 24(2), 237–267 (2002)
6. Huertas, A., Matthies, L., Rankin, A.: Stereo-based Tree Traversability Analysis for Autonomous Off-road Navigation. In: IEEE Workshop on Applications of Computer Vision (2005)
7. Ulrich, I., Nourbakhsh, I.: Appearance-based Obstacle Detection with Monocular Color Vision. In: AAAI National Conference on Artificial Intelligence, pp. 866–871 (2000)

8. Liu, Y., Gao, Y., Zhang, Y.: Optimum Path Search Algorithm Based on Cost Field. In: Geography and Geo-Information Science, vol. 20, pp. 28–30 (January 2004)
9. Book Editorial Board of SuperMap, The Component-based Development of SuperMap Object (2011)
10. Yu, Y., Wang, M.: The Path Planning Algorithm Research Based on Cost Field for Autonomous Vehicles. In: 2012 4th Intelligent Conference on Intelligent Human-Machine Systems and Cybernetics, pp. 38–42 (2012)
11. Shen, Z.J.: Analysis of the Shortest Route in Network on Dijkstra Algorithm. Geomatics and Spatial Information Technology 30, 36–39 (2007)
12. Qin, K., Guan, Z.Q., Li, D.R., Zhou, J.Q.: The Best Route Analysis Based on Raster Data. Remote Sensing for Land and Resources 52, 38–41 (2002)
13. Matsushita, K., Murakami, T.: Nonholonomic Equivalent Disturbance based Backward Motion Control of Tractor-trailer with Virtual Steering. IEEE Transactions on Industrial Electronics 55(1), 280–287 (2008)

A Comparison of Spatial and Temporal Dynamics of Landscape Pattern in the Cities of Pearl River Delta

ChoNam Ng[1], Xijun Yu[2,*], Yujing Xie[1], and Jian Yang[2]

[1] Department of Geography, the University of Hong Kong, Hong Kong, China
[2] South China Institute of Environmental Sciences,
Ministry of Environmental Protection, Guangzhou, P.R. China
yuxijun@scies.org

Abstract. Landscape dynamics along urban-rural transect has the potential for understanding and assessing ecological consequences due to urbanization. Comparative study on landscape pattern among different cities is still rare. By combining five sets of remote sensing data, landscape metrics, this paper investigates the process of urban rural interactions in six cities of the Pearl River Delta (PRD), namely, Guangzhou, Dongguan, Shenzhen, Foshan, Zhongshan, and Zhuhai. The results show that urbanization in the cities was an obvious phenomenon during the study period. But different cities have different trends in their process of "diffusion and coalescence". Landscape metrics can generally capture the similarity and difference of urbanization process. The findings can help to better improve the establishment of ecological network, especially in a city region context.

Keywords: Landscape dynamics, Comparative Study, Landscape Metrics, Pearl River Delta.

1 Introduction

Studies have demonstrated that urbanization can lead to conversion of natural and agricultural land[1, 2], environmental pollutions[3], urban heat islands effects[4], wildlife habitat loss and biodiversity changes[5, 6], and landscape diversity alternation[7]. In the long term, land-use change due to urbanization will influence regional and global climate change and other kinds of global consequences[8, 9]. Thus, analyzing spatial and temporal dynamics and evolution of urban landscape pattern has become a hot topic in the study of urbanization. Gradient paradigm has been widely used to investigate urban-rural interactions and their ecological consequences[10, 11]. The results have demonstrated the structural and functional differences of landscape changes along urban-rural gradient[12, 13]. All these studies provide useful information on urban pattern dynamics at both spatial and temporal scales.

However, most studies cited above focus on the urban land use in individual city or region [14]. A systematic and comprehensive comparison between different cities has

* Correspondsing author.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 565–581, 2013.

remained surprisingly scarce. In fact, comparative study on different landscape is very important for researchers and policy makers to examine the effects of local development regulations on landscape structure[15], and to understand the interaction and feedback mechanism among different components[16]. Over the past years there are some publications emphasizing on urban dynamics comparison by using s set of quantitative variables[17, 18]. The results have demonstrated that there are some similarities and differences in urban forms and growths across regions. In this process, great progress has been made in the integration of landscape metrics, remote sensing data, and other kind of social and economic data [19-21].

Urbanization is a vivid phenomenon in China in the past three decades since the government has adopted the "reform and openness" policy in 1978. It is predicted that China's urbanization level would reach 50% with an urban population of 1.5 billion by the end of 2020 [22]. As one of most developed regions in China, the Pearl River Delta (PRD) region has witnessed a rapid urbanization in the past three decades. Some publications have investigated the dynamics of land use and its ecological consequences [4, 23, 24]. However, land use in China is exceptionally dynamic due to the complexities of local, regional, and global forces, which differs from those in market economies [25, 26]. This process reflects the major feature of urban rural interactions in other developing countries.

By combining remote sensing data, landscape metrics, this paper addresses the process of urban rural interactions in some selected cities of the PRD, namely, Guangzhou, Dongguan, Shenzhen, Foshan, Zhongshan, and Zhuhai. The objective of the study is to make a systematic comparison of landscape changes in these cities. The study attempts to answer the following research questions: 1) What are the general temporal and spatial trends of landscape dynamics in the PRD region? 2) Is there any difference or similarity on landscape dynamics among cities in the PRD region? 3) What kinds of indicators can distinguish the dynamics of landscape pattern?

2 Study Area

The PRD is located in south China between 21º 40´ N to 23º N and 112º E to 113º 20´ E, with an area of appropriate 54 744 km^2 (Figure 1). Pearl River is the third largest river in China. The PRD region is formed by the confluence of the Dongjiang, Beijiang and Xijiang rivers flowing down from the Guangdong and Guangxi provinces. Due to the influence of subtropical climate, the region has plenty of rain fall and high temperature, with an average precipitation from 1600 to 2600 mm, and an average annual temperature between 21 and 23℃. Being at the entrance of the South China Sea, the region was historically an intensively agricultural area imbedding with dike-pond system[27].

The PRD region comprises of several cities, inclduing Guangzhou, Shenzhen, Zhuhai, Foshan, Jiangmen, Dongguan, Zhongshan, Huizhou, and Zhaoqing. Since the adoption of "reform and openness" policy in China, the region has experienced a rapidly economic development and social, economic, and spatial transformation over the past three decades. Because of its proximity to Hong Kong and Macao, the region has received much foreign investment and developed large amount of foreign trade.

The value of gross domestic production (GDP) in the region was 2974.56 billion Yuan in 2008, about thirty times that in 1990, which was 100.69 billion Yuan. The PRD accounts for only 30.5 percent of Guangdong's area, but contributes 79.4 percent of its GDP in 2008. The urbanization level in the region had reached 80.5% at the end of 2008, with an average population density of 863/km2[28].

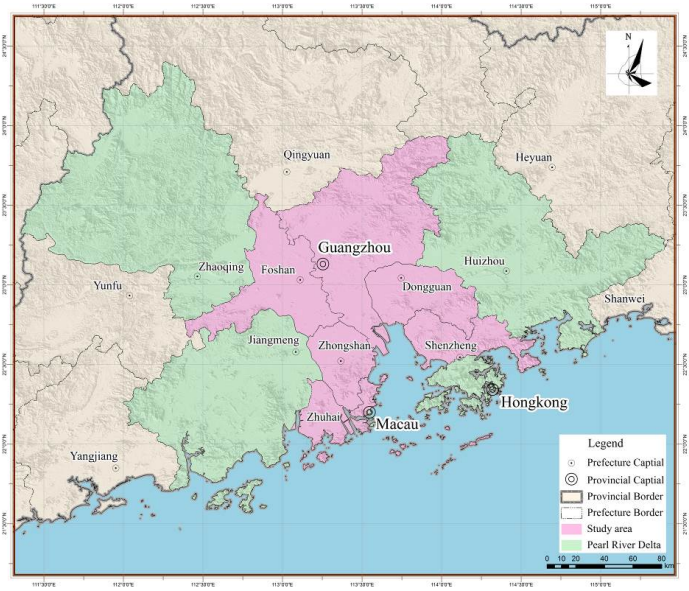


Fig. 1. Map of the study area

3 Data Processing and Methodology

Five sets of Landsat TM satellite images, covering the years 1988, 1993, 1998, 2003, and 2008, were used as basic data source in the study. All data have a ground resolution of 30 m, and were processed using the ERDAS IMAGINE software. The processing includes geometric correction, unsupervised classification and supervised classification, and GIS reclassification. Firstly, the images are rectified to Gauss-Krüger projection based on 1:50000 scale topographic maps. Secondly, unsupervised classification and supervised classification are performed. According to land use classification system in China, six land use types are then categorized, which include cultivated land, orchard, forest, urban built up area, new development area, and water area. According to Yu and Ng [10], new development area refers to an area that is just bulldozed to make way for constructions, which appears high contrast character and suggests the dynamics of urbanization. Finally, a vector road layer is merged with the above classified images. Thus, the final land use classes consists of cultivated land, orchard, forest, urban built up area, new development area, road, and water area (Figure 2).

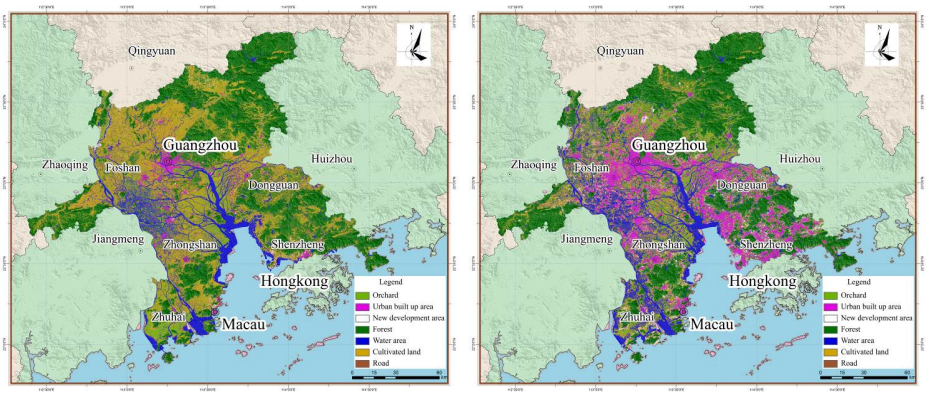


Fig. 2. Land use change in the cities between 1988 and 2008 (Left: 1988, Right:2008)

The classified images are further correct based on a GIS model with the help of a digital elevation model [10]. The accuracy of classification is evaluated by using a confusion matrix. For the 1988, 1993, 1998, and 2003 images, the color aerial photography and topographic maps taken in the corresponding years are used and the overall accuracy are found to be higher than 90%. For the 2008 image, field survey and GPS position are used to establish one-to-one relationship between the sampling sites and visual images. The overall accuracy can reach 95%.

In order to avoid the "salt and pepper" phenomena caused by traditional pixle-based classification [29], patches smaller than nine pixels will be eliminate according to the previous studies [30, 31]. CLUMP and ELIMINATE functions in ERDAS IMAGINE 8.7 are used in this process.

Landscape metrics have been widely used in analyzing and describing landscape characteristics in a region. Ten landscape metrics were selected in the study in order to remove redundant information and reduce correlation, which include NP, MPS, LPI, LSI, AWMPFD, SHDI, SHEI, CONTAG, and COHESION. All these metrics are calculated for the whole study and each city at the class and landscape levels using the FRAGSTATS software [32]. The definition and meaning of the selected metrics can be found in the report produced by McGarigal and Marks[33].

4 Results

4.1 General Trend of Land Use Change in the PRD Area

The negative effects of landscape fragmentation have drawn much attention of planners and policy makers. Landscape fragmentation was mainly caused by the reduction of the total amount of a specific habitat (decrease in their size), and the breaking up of habitats into smaller patches (increase in isolation of the ecosystem patches) [34]. Previous studies have demonstrated that landscape metrics have the potential for describing landscape fragmentation [35, 36]. In the present study, NP, MPS, CONTAG, AI, SHDI, and SHEI are selected to picture the process of landscape fragmentation (Figure 3).

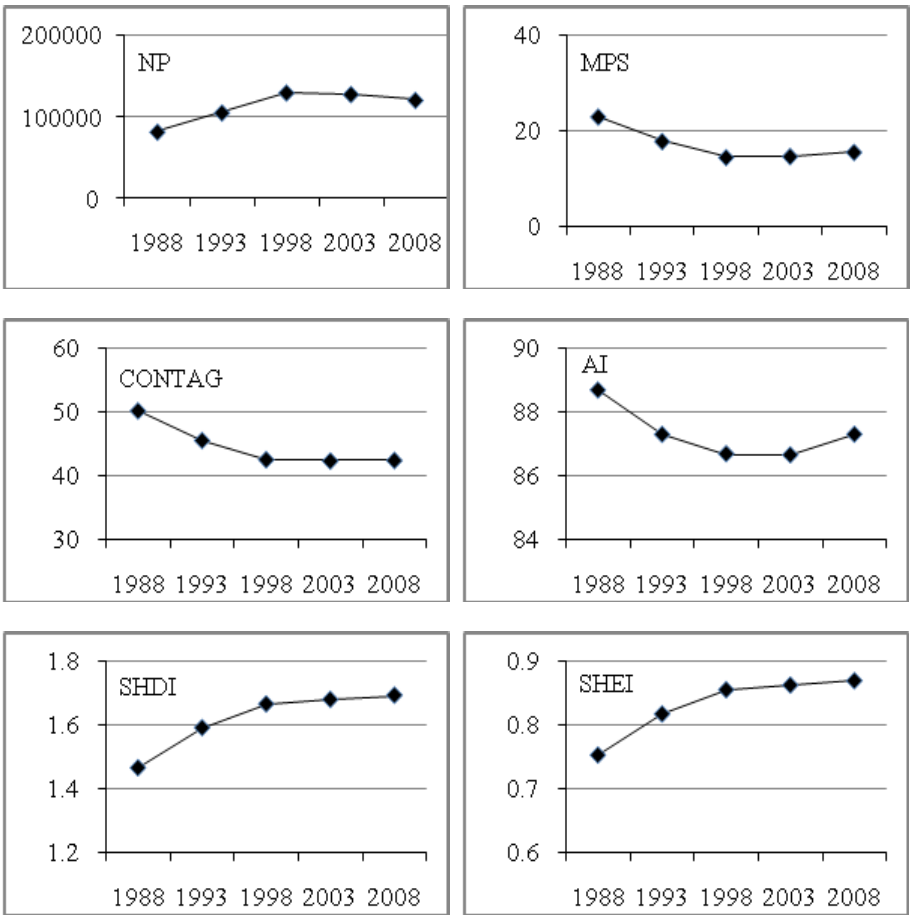


Fig. 3. Variations in landscape metrics at landscape level in the study area

The NP had a steady increase between 1988 and 1998, followed by a slight decrease. In contrast, the MPS decreased first and then increased, with the lowest value in 1998. This process indicated that the highest fragmentation of the landscape appeared in 1998. The variation of the LSI also revealed that the patches in the whole landscape became most disaggregated in 1998. The SHDI had increased from 1988 to 2008, suggesting that the degree of landscape diversity became higher. A similar trend was found in the variation of the SHEI, reflecting the distribution of different land use types in the PRD become increasing even. The decrease of the CONTAG indicated that land use types had been maximally disaggregated over time. The variation of the AI can also reflect this trend.

4.2 General Trend of Land Use Change for the Selected Cities

The PLAND describes the proportional abundance of each patch type in the landscape. Figure 4 shows the temporal changes of cultivated land PLAND in the six

cities. It was highest in 1998 and decreased over the study period, expect that Zhuhai had a slight increase between 1988 and 1993. Subsequently, the MPS has also a general decrease for most cities. The general decrease of the MPS, together with the general increase of the NP, indicated that cultivated land had a declining trend of dominance during the study period. The COHESION had a general decreasing trend for all cities, indicating that the patches of cultivated land became increasingly subdivided on the whole, and hence more physically disconnected.

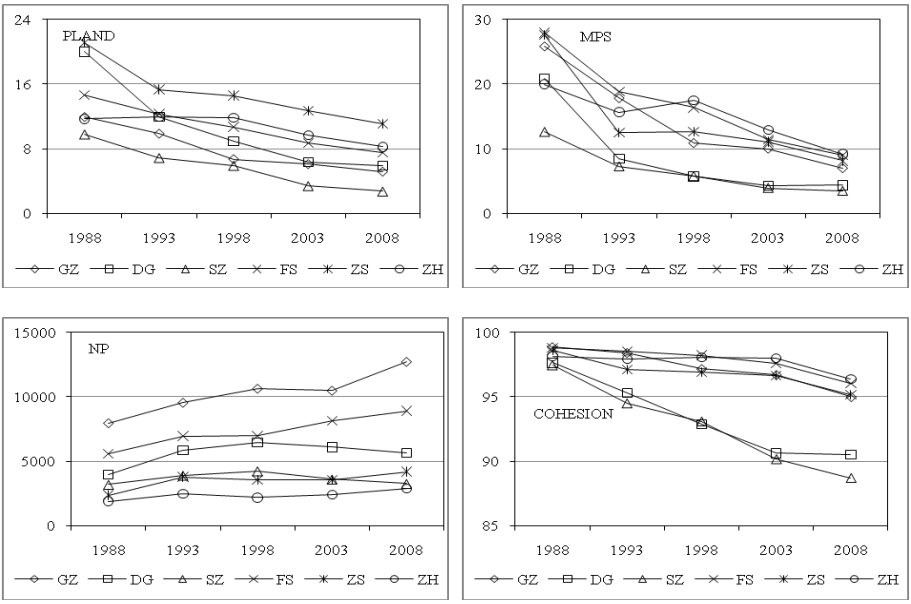


Fig. 4. Variations in landscape metrics for cultivated land in the cities

Compared with the cultivated land, orchard had more complicated variations through time and between cities. For most cities, the NP increased between 1988 and 1998, followed by a declining trend (Figure 5). While the MPS showed the general decrease between 1988 and 1998, and then increase in 2008. Variations in these two metrics suggested that orchard became increasingly fragmented between 1988 and 1998. Between 2003 and 2008, the NP declined with the MPS increased, suggesting that small patches of orchard were replaced by other land use types. Thus, orchard became more clumped and aggregated in its distribution. This can also be reflected by the increases of the COHESION.

The NP of forest had a general increase between 1988 and 1998, and then decreased. While the MPS decreased first and then increased, with the lowest values appearing in 1998 (Figure 6). This process also reflected the trend of fragmentation between 1988 and 1998 and aggregation between 1998 and 2008. The COHESION had a general decrease between 1988 and 1998, followed by an increase. Thus, forest patches became increasingly subdivided and less physically connected first, and then more clumped and aggregated.

The increase of the PLAND over time reflected the progress of urbanization in the PRD region (Figure 7). Urban built up area had a general increase in the variation of the NP. However, the MPS increased over time for most cities, except decline in

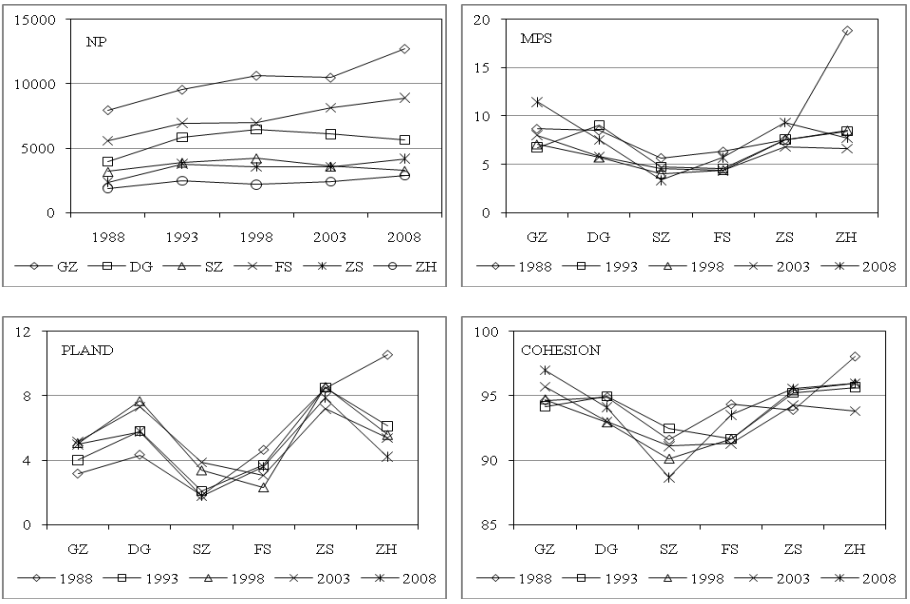


Fig. 5. Variations in landscape metrics for orchard in the cities

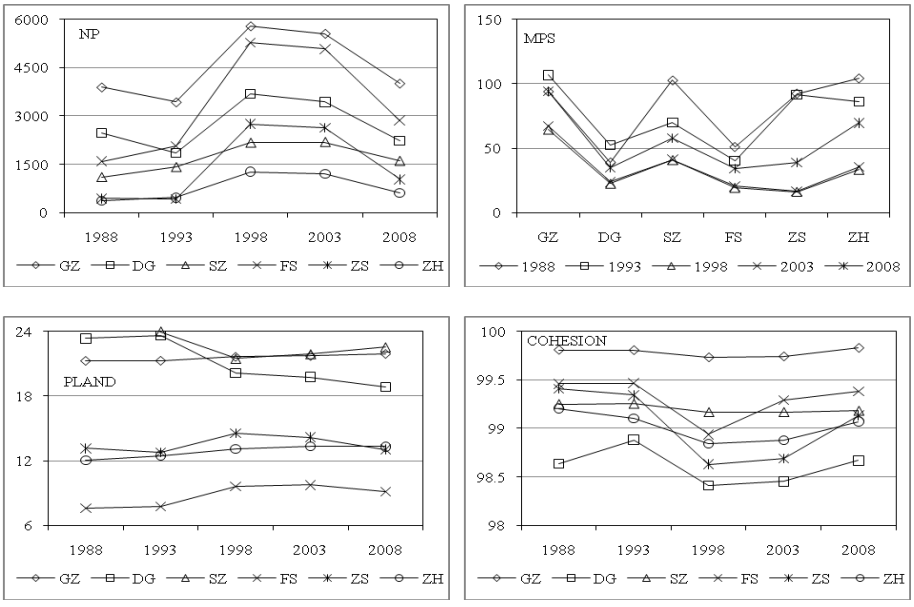


Fig. 6. Variations in landscape metrics for forest in the cities

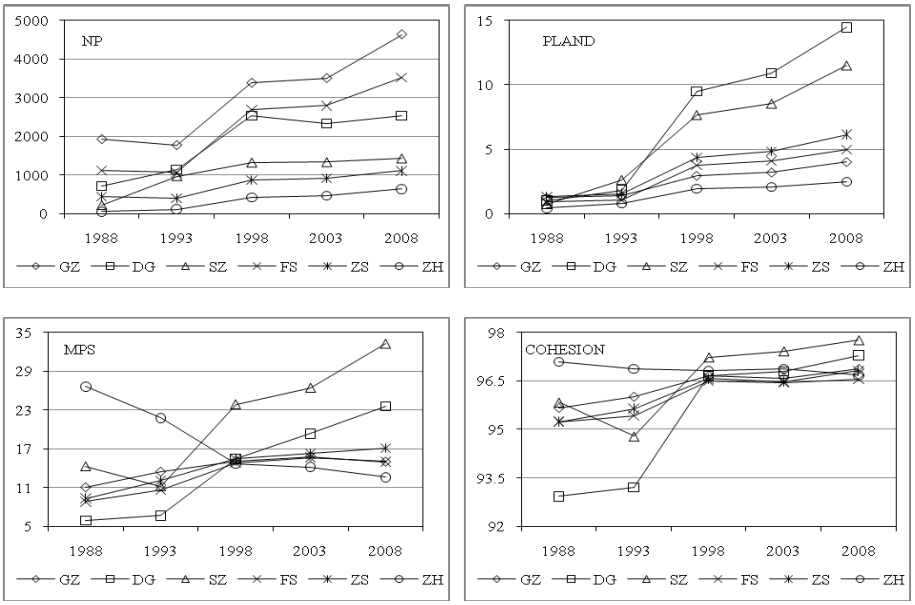


Fig. 7. Variations in landscape metrics for urban built up area in the cities

Zhuhai. The decrease of the MPS indicated that urban patches were formed in a scattered way and new urban areas grew faster that existing urban centres. This progress can also be reflected by the decline of the COHESION in Zhuhai, suggesting that urban patches became increasingly subdivided and less physically connected.

4.3 Comparison of Landscape Metrics Variation Among Six Cities

In order to compare landscape metrics between cities, relative changes of landscape metrics were calculated for each interval period, namely, 1988-1993, 1993-1998, 1998-2003, and 2003-2008. The following formula was used.

$LM_v=[LM^{t+1}-LM^t]/LM^t*100$

where LM_v is the relative change of the corresponding landscape metrics. LM^t and LM^{t+1} are the values of landscape metrics in the before and after year of the interval period, respectively.

Cultivated land, forest, and urban built up area were selected to compare their variability. Across all the six cities cultivated land had a relative decline in the PLAND, MPS, and COHESION while the NP generally increased (Figure 8). This suggested the fragmentation progress as illustrated in section 4.1. However, the degree of their variations was different between cities and over time. Between 1988 and 1993 the variations of the NP and MPS were relatively high in Dongguan and Zhongshan, which suggested that two cities became more fragmented among the six cities. This progress can also be reflected by the decline of their PLAND. Between 1993 and 1998, Guangzhou, Dongguan, and Shenzhen became more fragmented in cultivated land, which can be illustrated by their increase in the NP and decrease in

the MPS. Between 1998 and 2003, Foshan and Zhuhai became more fragmented while Guangzhou and Zhuhai followed this trend between 2008 and 2008.

For forest land, the increase of NP and the decrease of the MPS were observed in Shenzhen and Zhuhai between 1988 and 1993 (Figure 9). This process was in conjunction with the fragmentation of forest patches. Subsequently, all cities had the similar tendency in their variations between 1993 and 1998, with the highest variability occurring in Zhongshan and Zhuhai. The increase of the MPS and the decline of the NP across all the cities were observed between 1998 and 2008, with a slight change between 1998 and 2003 and an abrupt variation between 2003 and 2008. The highest variability occurred in Zhongshan and Zhuhai during this period. This process resulted in a more compacted but interspersed forest patches.

Between 1988 and 1993, urban built up area in Shenzhen and Zhuhai witnessed an increase in the NP and a decrease in the MPS. This indicated that two cities became more fragmented compared with the other four cities. Across the whole study period, only Zhuhai followed this tendency. The increase of the PLAND was also high for Shenzhen and Zhuhai, suggesting a rapid urbanization process. However, Foshan and Zhongshan had a decline in the NP and an increase in the MPS, exhibiting a more compacted urban form. Between 1993 and 1998, Dongguan and Foshan exhibited the highest increase in the PLAND. From 1998 to 2008, the variability of landscape metrics was relatively low. But Dongguan exhibited a decrease in the NP and an increase in the MPS between 1998 and 2003, suggesting an increasing compacted urban form. Between 2003 and 2008, except Zhuhai, Guangzhou and Foshan also showed a more fragmented urbanization process (Figure 10).

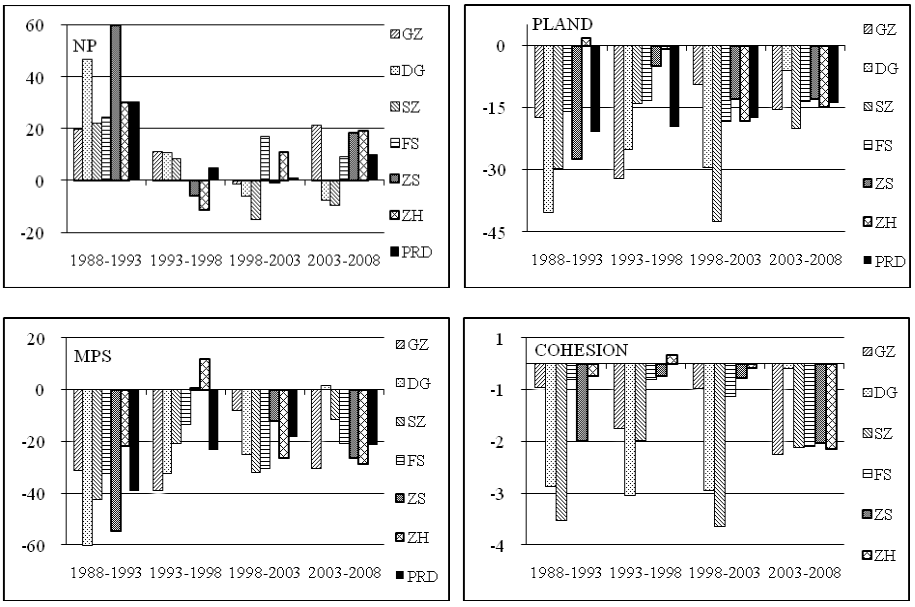


Fig. 8. Relative variability of landscape metrics for cultivated land in the cities

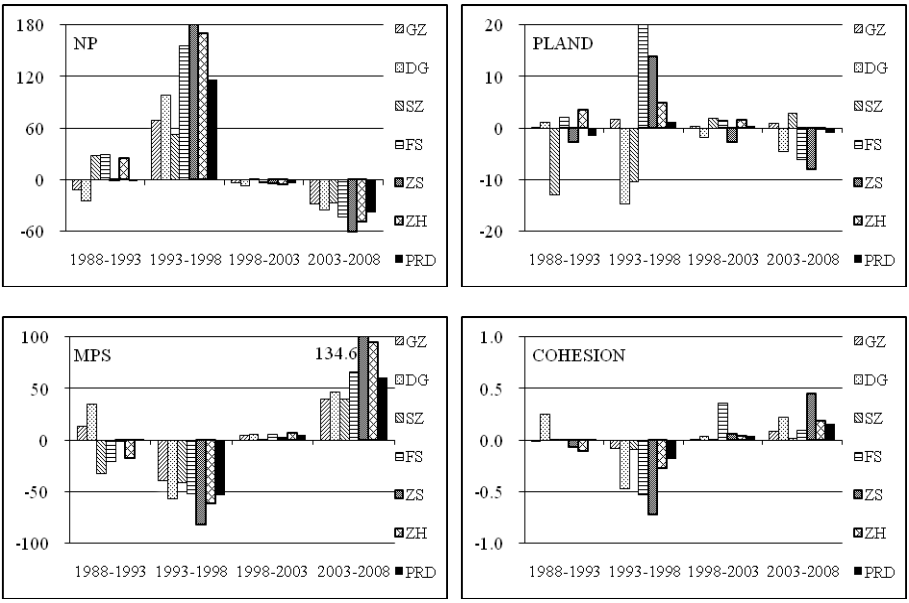


Fig. 9. Relative variability of landscape metrics for forest in the cities

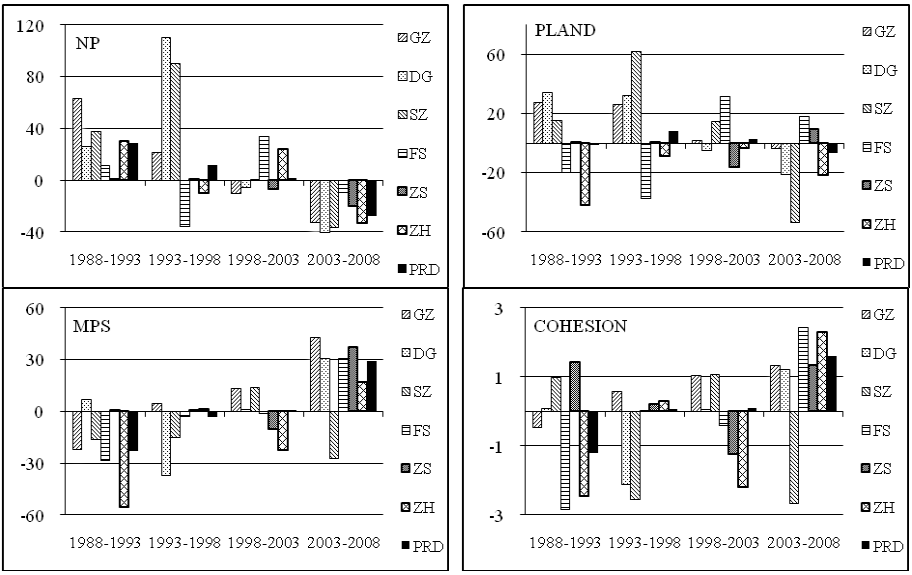


Fig. 10. Relative variability of landscape metrics for urban built up area in the cities

5 Discussions

5.1 Trends in Urbanization in the PRD Region

Urbanization in the PRD was an obvious phenomenon since adopted the market oriented policy in the late 1970s. The period of 1988-1998 witnessed higher urbanization process than that of 1998-2008 (Figure 11). This reflected the increasing urbanization progress since the adoption of "reform and openness" policy in late 1970s. During this period, the speed of urbanization was significant in Dongguan, Shenzhen, Zhongshan, and Zhuhai which is higher than the average of the PRD region. All these cities were formerly small cities or towns in the PRD region. Due to the special status as special economic zones, Shenzhen and Zhuhai had experienced rapid urbanization in the early stage of the reform period. The increasing urbanization in Dongguan and Zhongshan was driven by a new track of spontaneous urbanization[37]. Across the whole study period, the speed of urbanization in Guangzhou was lower than the average speed in the whole PRD region. This also reflected the outcomes of Chinese policy of controlling the growth of large cities[37].

The total area of Guangzhou, Dongguan and Shenzhen, in the eastern part of the PRD, had increased from 385.61 km^2 in 1988 to 2247.32 km^2 in 2008, while Foshan, Zhongshan and Zhuhai, in the western part of the PRD, had increased from 190.88 km^2 in 1998 to 1135.66 km^2 in 2008, respectively. Thus, the increase of urban built up area in the eastern part was significantly higher than that in the western part of the PRD, especially in Dongguan and Shenzhen. Due to the proximity to Hong Kong, the eastern part of the PRD had achieved rapid economic growth and thus brought to spatial disparity as compared with the western part[38]. In order to implement coordinated development of the whole PRD region, Guangdong government has set up the spatial strategy of "promoting the west bank, optimizing the east bank". The connection between Hong Kong and the western part of the PRD will have a great improvement after the construction of Hong Kong-Zhuhai-Macao Bridge. It is believed that this will bring about a new pattern of urbanization in the western part.

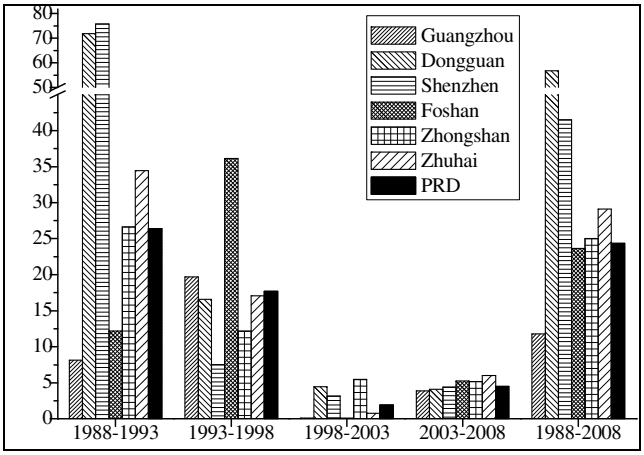


Fig. 11. Rates of urban built up area increase in the cities between 1988 and 2008

5.2 Similarity and Difference Urban Landscape Dynamics between Cities

It was proposed that urban form oscillated between a cycle of diffusion and coalescence[39]. Urban forms in some cities of the PRD region have also been examined in an international context[20, 21]. However, comparison between cities with the similar physical characteristic and national policies in a metropolitan area is still rare. The paper can help better answer this issue.

The MPS decreased in Zhuhai for the whole period and Shenzhen between 1988 and 1993, implying that new urban centres occurred and were growing faster than that of the former urban areas at specific period for two cities[40]. Being a small fishing village and former county, some industrial regions and high-tech zones were established after the "reform and openness". Thus, it is more likely to shape new dispersed and interspersed urban centres in two Special Economic Zones. Increasing NP and decreasing MPS in Guangzhou and Foshan between 2003 and 2008 might be driven by the adjustment of administrative boundary in 2000, which enlarged the extent of urban development areas and resulted in a more diffused progress. In contrast, decreasing NP and increasing MPS between 1988 and 1993 in Guangzhou and Foshan suggested that the former larger cities were firstly expanding along the former urban centres, which can be characterized by a coalesced progress. This tendency was also seen in Dongguan between 1998 and 2003, which was tested in a recent study using a new landscape index [41].

Across all the six cities in the PRD, the ENN_MN had a general decline over time. Thus, all cities can be identified as a diffusion phase at the beginning of a period of twenty years. This can also be reflected by the highest value of the CONTAG at the initial phase. However, the ENN_MN had a minor increase in Guangzhou and Zhongshan between 1988 and 1993, and Dongguan, Shenzhen, Foshan, and Zhongshan between 1998 and 2003, respectively. This suggested that these cities had a slight coalesced progress at the corresponding periods. The CONTAG reached its lowest value in 1998 for Guangzhou, Dongguan, and Shenzhen, suggesting that these cities were most heterogeneous in 1998. Landscape will become more homogenous since then along with coalesced urban pattern. Thus, this time point witnessed a shift from natural or semi natural landscape to urban landscape in these cities.

5.3 Applicability of Landscape Metrics to Capture Urban form Dynamics

The results of the present study can be used to test the hypothesis proposed by Dietzel et al.[39]. The low value in the CONTAG characterized the shift of urban form from diffusion to coalescence. In the present study, variations in the CONTAG can well capture urban diffusion-coalescence process in some cities. However, a discrepancy was also identified for some landscape metrics variation. Dietzel et al.[39] hypothesized that during a cycle of urban diffusion-coalescence process urban patch density (NP in the present study) and ENN_MN will increase first, and then finally decrease. Xu et al.[42] also concluded that the reduction of the area-weighted mean Euclidean nearest neighbour distance (AMENND) is representative of a shift from diffusion to coalescence. In the present study, however, the NP exhibited a wave-like

shape that might imply no one cycle of diffusion and coalescence occurs. The ENN_MN generally decreased through time, suggesting that a general diffusion process is observed. Therefore, variations in landscape metrics values detect a simultaneous diffusion or coalescence process during the same period.

Urban form might not be captured by a single landscape metric while a set of landscape metrics might suggest an alternative result. Such predominance might be due to the limitation of landscape metric. For example, decreasing NP and increasing MPS of urban patches can be representative of a diffusion process. While increasing NP plus an increasing MPS might also be a diffusion process in the case that old urban centre increases obviously faster than new outward urban areas. Thus, a standard diffusion-coalescence-further diffusion story may not be adequate[43]. It is appears more appropriate to be characterized by a partly-stochastic, partly-deterministic, spatially explicit birth and growth process [25]. In fact, the diffusion and coalescence process might simultaneously occur in different locations during the same period. Which process prevails depends on the degree and magnitude of both the old urban centre and new outlying areas. However, the absolute value of landscape metrics on the whole region might not be able to capture this complicated dynamics. Thus, in order to well distinguish two different urban processes, the relative variability of landscape metrics might be more realizable for detecting the degree of urban dynamics at different locations.

5.4 Implications for Land Use Management and Planning

Understanding urban dynamics and urban form is essential for land use management and urban planning. A fundamental dichotomy exists between managing for compact or diffuse cities [44]. Although ecosystem performance declined with increasing urban density [45], recent studies have supported compact city as a sustainable urban form in terms of land use efficiency and energy consumption [46]. Thus, compact city has been regarded as a sustainable form of urban development and widely adopted as a planning approach in developed countries, in particular the EU countries.

Preserving natural habitat and developing green structure are most important in compact city. By clearly distinguishing urban dynamics of individual city at different phase, the results can help policy makers and planning designers making the right decision. For city shifting to coalescence phase, urban form become more compacted as the infill development progresses. Urban fringes or peri-urban area outside the core urban centres will receive much pressure from land transformation. Although this process can help to reduce landscape fragmentation at the final phase of urbanization, it should be borne in mind that landscape background matrix has been transformed [47]. Thus, conserving natural habitat and maintaining greenway at urban fringe can help to reduce the negative impacts of urbanization. For cities at diffusion phase, urbanization generally occurs at the scatter plots far from urban core areas. This will bring about the increase of landscape fragmentation and the loss of biodiversity. With the use of landscape metrics, it can help to investigate the influence of urban diffusion on landscape diversity. Further countermeasure can be taken to minimize the impacts of locally uncontrolled land uses on landscape fragmentation.

5.5 Implications for Regional Ecosystem Conservation Cross Cities

The debate during the early 1990s over urban sustainability had helped to focus attention on the impacts of cities beyond their boundaries [48]. The notion of city-region has been accepted in China and city-region planning has become a new form of Chinese spatial planning [49]. New modern ecological framework with a focus on interactions among multiple sites was adopted in designing New York greenspace system [50]. More recently, great effort has been done to establish ecological network at regional level that is beyond individual city or transcend boundary [51].

Across the whole study area, urbanization is generally regarded as a diffusion process that occurs at remote areas far from urban centres. In peri-urban areas, rapid uncontrolled development might occur in non-urban planned districts and thus result in increased urban sprawl[48], especially along regional transportation networks linking different cities. The increasing land use conversion from non-urban land to urban land will bring about many ecological consequences, such as landscape fragmentation, biodiversity loss. Thus, the designation of long-term natural and agricultural production areas is urgently needed[48]. As one of largest metropolitan region in China, the cities in the PRD are committed to regional cooperation and integrity within the region. It was under this background that "Plan for the coordinated development of the PRD Townships (2004-2020)" was enacted, which is the first regional plan that focuses on city region in the PRD region. In this plan, optimization of the regional eco-security was proposed through conserving natural habitats, establishing eco-corridors, and protecting eco-functional zones[28].

6 Conclusions

The study aims at improving our understanding of landscape dynamics in a city region context. Landscape analysis can fundamentally capture the similarity and difference of urbanization process for the selected cities in the PRD. The temporal landscape changes in the cities was mainly driven by increasing urbanization that was characterized by a partly-stochastic, partly-deterministic process[25]. The spatial difference among cities was mainly related to urban historical trajectories and the influence of both top down polices and bottom up reactions. Urban dynamics analysis can also capture the process of diffusion and coalescence in the cities of the PRD.

A city based landscape analysis can give us an insight on the general process of urbanization and make comparison among cities. For policy makers and planners seeking to manage urban development in a city region context, the comparison among cities can clearly distinguish the similarity and difference of each city within a region. It can help to improve the establishment of ecological network that is beyond the individual city. Thus, there is a need for advancing our knowledge of landscape changes in city region or metropolitan area in a coherent way.

Acknowledgement. This research is financially supported by the General Research Fund, Research Grants Council of Hong Kong (No. HKU 748707H), Science and

Technology Planning Project of Guangdong Province (No.2011B031200011), and Marine Fisheries Science and Technology Promotion Project of Guangdong Province (No. A201001I04, No. A201101I04), China.

References

1. Seto, K.C., Kaufmann, R.K., Woodcock, C.E.: Landsat reveals China's farmland reserves, but they're vanishing fast. Nature 406, 121 (2000)
2. Deng, J.S., Wang, K., Hong, Y., Qi, J.G.: Spatio-temporal dynamics and evolution of land use change and landscape pattern in response to rapid urbanization. Landscape and Urban Planning 92, 187–198 (2009)
3. Du, N., Ottens, H., Sliuzas, R.: Spatial impact of urban expansion on surface water bodies—A case study of Wuhan, China. Landscape and Urban Planning 94(3-4), 175–185 (2001)
4. Weng, Q.: A remote sensing–GIS evaluation of urban expansion and its impact on surface temperature in the Zhujiang Delta, China. International Journal of Remote Sensing 22(10), 1999–2014 (2001)
5. Green, D.M., Baker, M.G.: Urbanization impacts on habitat and bird communities in a Sonoran desert ecosystem. Landscape and Urban Planning 63, 225–239 (2003)
6. McKinney, M.L.: Urbanization, biodiversity, and conservation. Bioscience 52(10), 883–890 (2002)
7. Yeh, C.T., Huang, S.L.: Investigating spatiotemporal patterns of landscape diversity in response to urbanization. Landscape and Urban Planning 93, 151–162 (2009)
8. Feddema, J.J., Oleson, K.W., Bonan, G.B., Mearns, L.O., Buja, L.E., Meehl, G.A., Washington, W.M.: The Importance of Land-Cover Change in Simulating Future Climates Sciences 310, 1674–1678 (2005)
9. Pielke Sr., R.A., Marland, G., Betts, R.A., Chase, T.N., Eastman, J.L., Niles, J.O., Niyogi, D., Running, S.W.: The influence of land-use change and landscape dynamics on the climate system-relevance to climate change policy beyond the radiative effect of greenhouse gases 360, 1705–1719 (2002)
10. Yu, X.J., Ng, C.N.: Spatial and temporal dynamics of urban sprawl along two urban–rural transects: A case study of Guangzhou, China. Landscape and Urban Planning 79(1), 96–109 (2007)
11. Kong, F., Nakagoshi, N.: Spatial-temporal gradient analysis of urban green spaces in Jinan. Landscape and Urban Planning 78, 147–164 (2006)
12. McDonnell, M.J., Pickett, S.T.A.: Ecosystem structure and function along urban-rural gradients: an unexploited opportunity for ecology. Ecology 71(4), 1232–1237 (1990)
13. Luck, M., Wu, J.: A gradient analysis of urban landscape pattern: a case study from the Phoenix metropolitan region of USA. Landscape Ecology 17, 327–339 (2002)
14. Seto, K.C., Shepherd, J.M.: Global urban land-use trends and climate impacts. Current Opinion in Environmental Sustainability 1, 89–95 (2009)
15. Kim, J., Ellis, C.D.: Determining the effects of local development regulations on landscape structure: Comparison of TheWoodlands and North Houston, TX. Landscape and Urban Planning 92, 293–303 (2009)
16. Li, A., Deng, W., Kong, B., Song, M., Feng, W., Lu, X., Lei, G., Bai, J.: A comparative analysis on spatial patterns and processes of three typical wetland ecosystems in 3H area, China. Procedia Environmental Sciences 2, 315–332 (2010)

17. Kasanko, M., Barredo, J., Lavalle, C., McCormick, N., Demicheli, L., Sagris, V., Brezger, A.: Are European cities becoming dispersed? A comparative analysis of 15 European urban areas. Landscape and Urban Planning 77, 111–130 (2006)
18. Tsai, Y.-H.: Quantifying Urban Form: Compactness versus 'Sprawl'. Urban Studies 42(1), 141–161 (2005)
19. Schwarz, N.: Urban form revisited—Selecting indicators for characterising European cities. Landscape and Urban Planning (2010)
20. Huang, J., Lu, X.X., Sellers, J.M.: A global comparative analysis of urban form: Applying spatial metrics and remote sensing. Landscape and Urban Planning 82, 184–197 (2007)
21. Schneider, A., Woodcock, C.E.: Compact, Dispersed, Fragmented, Extensive? A Comparison of Urban Growth in Twenty-five Global Cities using Remotely Sensed Data, Pattern Metrics and Census information. Urban Studies 45(3), 659–692 (2008)
22. Tian, G., Liu, J., Xie, Y., Yang, Z., Zhuang, D., Niu, Z.: Analysis of spatio-temporal dynamic pattern and driving forces of urban land in China in 1990s using TM images and GIS. Cities 22(6), 400–410 (2005)
23. Weng, Q.H.: Land use change analysis in the Zhujiang Delta of China using satellite remote sensing, GIS and stochastic modelling. Journal of Enviromnent Management 64, 273–284 (2002)
24. Seto, K.C., Woodcock, C.E., Song, C., Huang, X., Lu, J., Kaufmann, R.K.: Monitoring land-use change in the Pearl River Delta using Landsat TM. International Journal of Remote Sensing 23(10), 1985–2004 (2002)
25. Fragkias, M., Seto, K.C.: Evolving rank-size distributions of intra-metropolitan urban clusters in South China. Computers, Environment and Urban Systems 33(3), 189–199 (2009)
26. Lin, G.C.S.: The growth and structural change of Chinese cities: A contextual and geographic analysis. Cities 19(5), 299–316 (2002)
27. Ruddle, K., Zhong, G.F.: Integrated Agriculture-aquaculture in South China, the Dike-pond System of the Zhujiang Delta. Cambridge University Press, New York (1988)
28. People's Government of Guangdong Province, Plan for the coordinated development of the Pearl River Delta Townships (2004-2020) (2005), http://zwgk.gd.gov.cn/006939748/200909/t20090915_8884.html
29. Bock, M., Xofis, P., Mitchley, J., Rossner, G., Wissen, M.: Object-oriented methods for habitat mapping at multiple scales – Case studies from Northern Germany and Wye Downs, UK. Journal for Nature Conservation 13(2-3), 75–89 (2005)
30. Asnera, G.P., Keller, M., Pereira, R., Zweeded, J.C.: Remote sensing of selective logging in Amazonia Assessing limitations based on detailed field observations, Landsat ETM+, and textural analysis. Remote Sensing of Environment 80, 483–496 (2002)
31. Knight, J.F., Lunetta, R.S.: An Experimental Assessment of Minimum Mapping Unit Size. IEEE Transactions on Geoscience and Remote Sensing 41(9), 2132–2134 (2003)
32. McCarigal, K., Cushman, S.A., Neel, M.C., Ene, E.: FRAGSTATS: Spatial Pattern Analysis Program for Categorical Maps, Version 3.0. University of Massachusetts, Amherst (2002)
33. McGarigal, K., Marks, B.J.: FRAGSTATS: Spatial Pattern Analysis Program for Quantifying Landscape Structure Gen. Tech. Rep. PNW-GTR-351, 122 p. US Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland (1995)
34. Burel, F., Baudry, J.: Landscape ecology: concepts, methods and applications. Plymouth, Science Publishers, Enfield (2003)

35. Saura, S., Martine-Millan, J.: Sensitivity of landscape pattern metrics to map spatial extent. Photogrammetric Engineering & Remote Sensing 67, 1027–1036 (2002)
36. Tyler, M.W., Peterson, D.L.: Effects of forest policy on landscape pattern of late-seral forest of the Western Olympic Peninsula, Washington. Agriculture, Ecosystems and Environment 101, 289–306 (2004)
37. Shen, J., Feng, Z., Wong, K.: Dual-track urbanization in a transitional economy: The case of Pearl River Delta in South China. Habitat International 30, 690–705 (2006)
38. Li, X., Yeh, A.G.-O.: Analyzing spatial restructuring of land use patterns in a fast growing region using remote sensing and GIS. Landscape and Urban Planning 69, 335–354 (2004)
39. Dietzel, C., Herold, M., Hemphill, J.J., Clarke, K.C.: Spatio-temporal dynamics in California's Central Valley: empirical links to urban theory. International Journal of Geographical Information Science 19(2), 175–195 (2005)
40. Seto, K.C., Fragkias, M.: Quantifying spatiotemporal patterns of urban land-use change in four cities of China with time series landscape metrics. Landscape Ecology 20, 871–888 (2005)
41. Liu, X., Li, X., Chen, Y., Tan, Z., Li, S., Ai, B.: A new landscape index for quantifying urban expansion using multi-temporal remotely sensed data. Landscape Ecology 25, 671–682 (2010)
42. Xu, C., Liu, M., Zhang, C., An, S., Yu, W., Chen, J.M.: The spatiotemporal dynamics of rapid urban growth in the Nanjing metropolitan region of China. Landscape Ecology 22, 925–937 (2007)
43. Seto, K.C., Fragkias, M., Schneider, A.: 20 Years After Reforms: Challenges to Planning and Development in China's City-Regions and Opportunities for Remote Sensing. In: Redman, C., Stefanov, W., Netzband, M. (eds.) Applied Remote Sensing for Urban Planning, Governance and Sustainability, pp. 249–269. Springer (2007)
44. Wu, J., Jenerette, G.D., Buyantuyev, A., Redman, C.L.: Quantifying spatiotemporal patterns of urbanization: The case of the two fastest growing metropolitan regions in the United States Ecological Complexity (2010)
45. Tratalos, J., Fuller, R.A., Warren, P.H., Davies, R.G., Gaston, K.J.: Urban form, biodiversity potential and ecosystem services. Landscape and Urban Planning 83, 308–317 (2007)
46. Holden, E., Norland, I.T.: Three Challenges for the Compact City as a Sustainable Urban Form: Household Consumption of Energy and Transport in Eight Residential Areas in the Greater Oslo Region. Urban Studies 42(12), 2145–2166 (2005)
47. Weng, Y.-C.: Spatiotemporal changes of landscape pattern in response to urbanization. Landscape and Urban Planning 81, 341–353 (2007)
48. Huang, S.-L., Wang, S.-H., Budd, W.W.: Sprawl in Taipei's Peri-urban Zone: Responses to Spatial Planning and Implications for Adapting Global Environmental Change. Landscape and Urban Planning 90, 20–32 (2009)
49. Luo, X., Shen, J.: Why city-region planning does not work well in China: The case of Suzhou–Wuxi–Changzhou. Cities 25, 207–217 (2008)
50. Flores, A., Pickett, S.T.A., Zipperer, W.C., Pouyat, R.V., Pirani, R.: Adopting a modern ecological view of the metropolitan landscape: the case of a greenspace system for the New York City region. Landscape and Urban Planning 39(4), 295–308 (1998)
51. Leienath, M., Blurn, A., Stutzriemer, S.: Transboundary cooperation in establishing ecological networks: The case of Germany's external borders. Landscape and Urban Planning 94, 84–93 (2010)

Analysis of Land Consolidation Potential of Hollowed Villages and Its Eco-Economic Benefits under the Background of Urbanization:——A Case Study on Qin'an County, Gansu Province*

Xiaodong Guo[1], Shisi Yang[1], Libang Ma[2], and Yingfei Li[1]

[1] School of Management, Lanzhou University, 730000 Lanzhou, Gansu Province, China
[2] College of Geography and Environmental Science, Northwest Normal University, Lanzhou, 730000, Lanzhou，Gansu Province, China
{Xiaodong Guo,gxd}@lzu.edu.cn,
{Shisi Yang,449554741}@qq.com

Abstract. Village hollowing is an infaust evolution of village structure caused by the non-agricultural transfer of the rural population under the background of urbanization. Take Qin'an county in loess hilly-gully region as a case, this paper analyzed the potential of hollowed village land consolidation of Qin'an county quantificationally based on the investigation and analysis of six indicators.The authors also made a quantitative analysis of the ecological and economic benefits of the hollow Village' remediation of Qin'an County based on the analysis of the objective reality and the current needs of the rural socio-economic development and eco-environment construction in Qin'an. The results show that Qin'an has huge potential in remedying hollow villages under the urbanization. Actively promote the hollow village land consolidation, reasonable and intensive utilization of land resources is not only the objective demand of the socialist new rural construction, and it also will show great ecological and economic benefits.

Keywords: hollow village, land consolidation, ecological benefit, economic benefit, Qin'an County.

1 Introduction

Urbanization is a process which rural population turns into urban population and rural areas turn into urban areas, and it is also an inevitable phenomenon and process along with the economic and social development [1]. It is substantially a process which the rural labors finish their move from traditional industry to modern industry and their migration from the countryside to the city [2]. Since the reform and openness, the urbanization in China developed rapidly, and the rate of the urbanization increased rapidly by an average annual rate of 0.93 percentage [3]. Especially in the past 10

* The research funded by National Natural Science Foundation of China (NSFC) (40971106).

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 582–592, 2013.

years, as the acceleration of Chinese urbanization process and rural-urban transformation pace, more and more rural population turns into non-agricultural population, which results in an increasingly severe village hollowing phenomenon. People left their hometown and their empty houses; the village continues to expand its living areas but has a "hollow heart"; people in countryside did not remove their old homestead while building a new one. It aroused a high attention from the government and the academic. In this background, there are lots of extensive discussions in the field of geography, economics, sociology and other disciplines about the village hollowing phenomenon in rural places and the resultant waste of land resources, destruction of the living environment and other issues, resulting in a large number of respective research achievements [4-9]. Some scholars believe that the village hollowing phenomenon in rural area is essentially an infaust process of rural areas system evolution during the development of rural-urban transformation, and it is a complex socio-economic process embodied as the physical form of villages(Feng Li, 2008) [9]; Tian Guangjin(2003) pointed out that the widely-distributed and numerous villages resulted in village expansion and astonishing waste of land resources within the special institutional background, which threatened the arable land protection and food security in China and became the primary problem on current rural construction and co-ordinate development of urban and rural places[10]. In recent years, domestic scholars analyzed the remediation potential of the different rural living areas by surveying the current situation of rural residential land, constructing restrictive correction factor and mathematical models [11-14], and there are other scholars who also discussed and analyzed eco-environmental benefits of rural land consolidation [15]. These studies have an extremely important significance on having in-depth understanding of the issues on village hollowing evolution in rural places, probing in and improving the quantitative analysis method of land remediation, and reasonably analyzing the land remediation potential and ecological and economic benefits of hollow villages. However, after analyzing current studies, the existing research mainly focuses on the causes of hollow villages, the mechanism of village hollowing evolution, the characteristics of rural village hollowing evolution, the integrated treatment model and countermeasures of hollow villages, etc. There are just few researches focusing on the remediation potential of the hollow villages, especially on the field of the integrated analysis of combining the remediation potential of the hollow villages and its ecological and economic benefits. In this paper, the authors take Qin'an County---the loess hilly area in the middle of Gansu Province as the empirical research area, and make integrated study and quantitative analysis of the remediation potential and its ecological and economic benefits of the hollow villages in Qin'an County based on the field research and correlational research results.

2 The Profile of the Research Area and the Data Sources

Qin'an County is located in the southeast of Gansu Province, the north of Tianshui, the downstream of Hulu River---the tributary of Wei River, and belongs to the loess hilly-gully area of the middle of Gansu Province. There are many mountains in Qin'an but few flatlands. It is a hilly place, and is one of the eighteen drought areas in

Gansu. It is also one of the national poverty alleviation counties. The total area of Qin'an is 1601 km^2. There are 5 towns, 12 villages, 428 village committees, 6 communities, 1384 groups of villagers, and 159,000 households in Qin'an, including 124,700 rural households. In 2010, the total population in Qin'an is 616,400 including 573,000 rural population. The population density is 384 per square kilometer. The total cultivated area is 70113hm^2， and the per capita cultivated land is 0.115 hm^2. In this paper, the main sources of data are from *Qin'an Statistical Yearbook* (2001-2011, Qin'an Bureau of Statistics) and field survey data. The urban and rural population, rural households and other socio-economic data are from *Qin'an Statistical Yearbook*; the transfer scale of rural surplus labor, the homestead areas and other data are from the government work reports of Qin'an and field research.

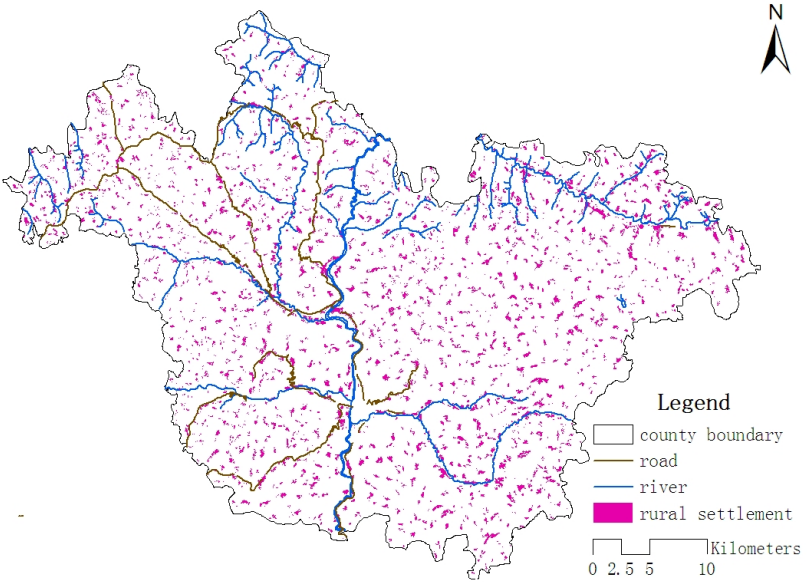


Fig. 1. The spatial distribution of *rural settlements* in Qin'an County

3 The Analysis of the Land Remediation Potential of Hollow Villages

The land remediation potential of hollow villages refers to the amount of the available land areas gained from the remediation of the abandoned homestead in the villages. In recent years, along with the continually intensified land waste phenomenon in the backdrop of rapid urbanization, domestic scholars, focusing on the issues of village hollowing evolution in villages and abandoned settlements, proposed different indicators which can be used to measure the degrees of the village hollowing evolution in villages and vacant and waste settlements such as the abandoned rate of land, population mobility rate, vacancy rate and abandoned rate of present residence, rate of vacancy house and abandoned house, the population density in

villages. By analyzing the current situation, we can tell that the amphibious feature is very common to the surplus rural labor in the loess hilly area in the middle of Gansu, therefore at this stage, the main performances of the village hollowing evolution in villages are the seasonal vacancy of the homestead, the vacant rate of homestead is higher than the rate of abandoned homestead. By analyzing the development trend, we can know that, along with the strategic transformation of China's urbanization, the improvement of the quality of urbanization and the social security system of migrant workers, the rate of stable employment in city and the migration rate of rural migrant workers will gradually improve, therefore the rural homestead of migrant workers will gradually transfer from the vacant to the waste, and the abandoned rate of rural homestead will significantly improve.

3.1 Calculation Method

According to the actual situation of the loess hilly area in the middle of Gansu and its development trend, the paper investigated the abandoned rate of the homestead, analyzed and predicted the abandoned rate of homestead under the background of urbanization, and calculated and analyzed the land remediation potential of the hollow villages in Qin'an in current situation and in future. The formula is:

$$A = A_0 + A_n = R_0 q_0 h + R_n q_n h = h(R_0 q_0 + R_n q_n) \quad (1)$$

$$If\ q_0 = q_n,\ then\ A = hq_0 (R_0 + R_n) \quad (2)$$

Where: A is the land remediation potential of the hollow villages; A_0 and A_n respectively refer to the current and the future land remediation potential of the hollow villages; R_0 and R_n respectively mean the current and the future abandoned rate of the rural homestead; q_0 and q_n respectively refer to the current and the future total rural households in Qin'an; h is the average homestead area.

3.2 Index Analysis

The land remediation potential of the hollow villages involves the scale of the transfer of rural surplus labor and their migration rate, the degree of the urbanization of population, the abandoned rate of the rural homestead and other indicators. According to the regional socio-economic development and the development trend of urbanization, scientifically analyzing and reasonably determining aforesaid indicators are the basic and key to analyze the land remediation potential of hollow villages. ① the current abandoned rate of homestead and homestead area. Qin'an County has 1,627 villages, and 124,700 rural households. According to the field research of Wangfu, Guojia, Liuping , Anfu, Yebao---5 towns, 12 administrative villages, 51 villages in Qin'an, we can see that the current abandoned rate of the homestead in villages is from 1.34% to 19.61% (the average abandoned rate is 6.99%), and the homestead rate is between 200~270m^2; ② rural population and rural households. Affected by the slow growth of rural population, the increase of the proportion of one-child families in rural areas and the process of urbanization, in the past ten years, the growth of the rural households in Qin'an obviously slowed down (Fig.1), therefore

the rural population and rural households in the future will continue to decline with the development of urbanization, while the abandoned rate of homestead will continue to increase; ③ the transfer scale of rural surplus labor and migration rate. Based on the field research, in recent years, the transfer scale of rural surplus labor in Qin'an remained stable. The annual transfer number is about 100,000 (the number is 106,800 in 2010), which accounted for 18.6% of the rural population. By analyzing the variation trend of the number of the rural labor in Qin'an (Fig.2), it showed that the growth of the rural surplus labor would be very limited, therefore the transfer scale of the rural labor will also be stabilized. In particular, the relatively stable transfer scale of the rural surplus labor in Qin'an is based on the high amphibious rate and low migration rate. The migratory move of the surplus labor is the main reason for the relatively stable transfer scale of the rural labor; ④ county urbanization. Urbanization is divided into local and remote urbanization. Qin'an is an undeveloped

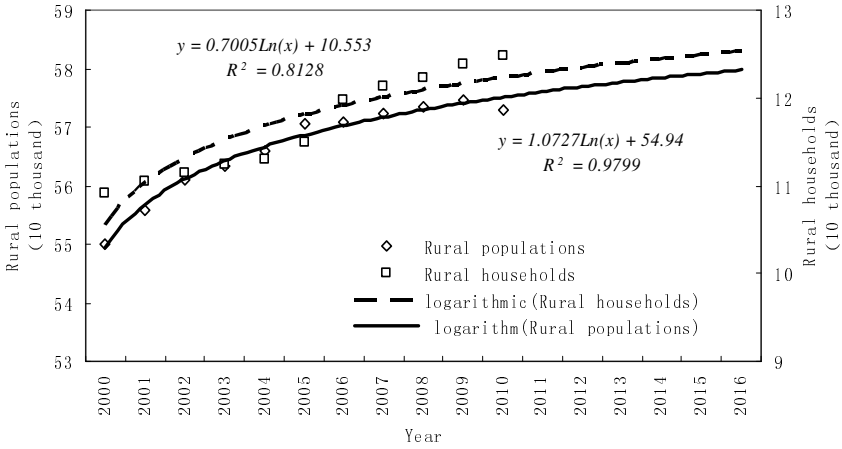


Fig. 2. The Analysis of the Variation Trend of the *Rural Population* and *Rural Households* in Qin'an

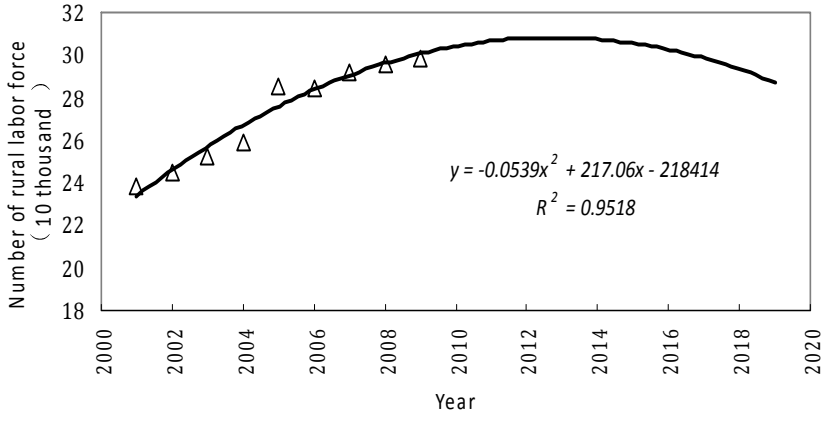


Fig. 3. Analysis of the growth trend of the *number of rural labor* in Qin'an

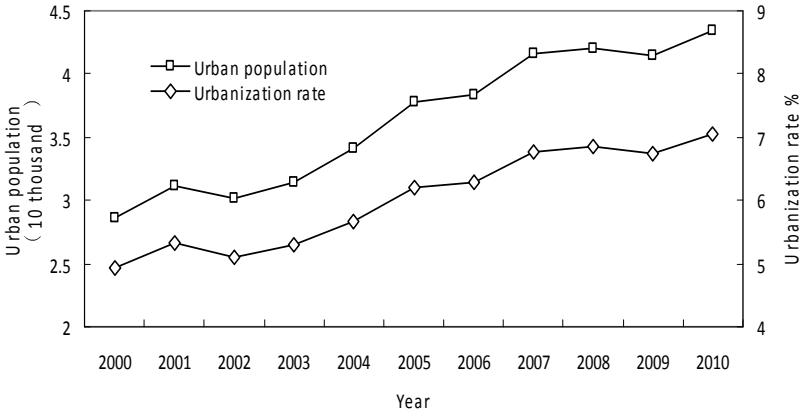


Fig. 4. The growth trend of the *urban population* and the *urbanization rate* in Qin'an

county, and the county urbanization (the local urbanization) is in a low level. In 2010, the county urbanization rate was only 7.04%, which is far lower than 36.1%, the average level in Gansu. With the analysis of the development trend, the county urbanization will greatly increase with the strategic transformation of China's urbanization and the promotion of the process of new urbanization.

3.3 The Calculation of the Remediation Potential

According to the aforesaid index analysis, we can gain from the calculation of formula (1) and formula (2) that, the current land remediation potential in Qin'an (A_0) is 204.8 hm^2. Calculated with the relevant indicators which maybe achieve in the future, we know that the remediation potential of the newly increased land in 2010 and 2030 will respectively reach 540.5 hm^2 and 1402.95 hm^2 (Tab.1).In table 1，the homestead area of the research villages in Qin'an is between 200~270m^2, and in the analysis of land remediation potential, we calculated with the median number---235m^2，the local urbanization rate from 2010 to 2020 is estimated with the data which increase by an average annual growth of 0.8%, and from 2020 to 2030 by 1%，the average household size is calculated for 4 people.

Table 1. The analysis of the urbanization of population and the remediation potential in Qin'an

	Year	The rate of urbanization (%)	Ratio of labor transfer	Total decrease number of the rural households(10 thousand)	Total number of the abandonned homestead(10 thousand)	The added value of the homestead remediation potential(hm^2)
Current number	2010	7.04	-	-	0.87	204.8
Prediction number	2020	15	0.1	2.30	2.30	540.50
	2030	25	0.3	5.97	5.97	1402.95

4 The Analysis of the Ecological and Economic Benefits of the Land Remediation of the Hollow Villages

Qin'an is located in the loess plateau area, was undeveloped in economy and had serious soil erosion. The long-standing PPE circle (Poverty- Population growth-Environmental degradation vicious cycle) seriously threatened the regional social and economic sustainable development. In past ten years, with the implementation of the national policy of reversion of arable land to forest, Qin'an has had great development in economy, and now has become one of the major production bases of fruits in north China, and has been awarded as *The Top Ten County Which Made Great Development of Economic Forest, The Famous Peach Town in China* and *The National Demonstration Counties*. The construction of the economic forest is an effective way to combine the rural economic development with the ecological environment construction. By the combination of the rearrangement of villages, constructing the center village and land remediation of hollow villages, through the land replacement[1] and the adjustment of the land layout, Qin'an can take use of the land gained from the remediation of hollow villages to plant economic forest which will produce huge ecological and economic benefits.

4.1 The Analysis of Ecological Benefits

Economic forest is a specific type of woodland, just like other woodland, it also has multiple ecological benefits like landscaping, air purification, climate regulation, soil and water conservation, etc. In this paper, the authors will make a quantitative analysis of its ecological benefits from four aspects.

The Value of Air Purification.
Woodland can absorb CO^2, release O^2, absorb and prevent dust and harmful gases, so as to purify the air. It can use replacement cost approach to calculate the benefits of the air purification from woodland [15]. The formula is:

$$PVB = \sum_{t=0}^{t} \frac{B_t}{(1+r)^t} \tag{3}$$

Where: PVB means the present value of the total utility; B_t refers to the benefit of the year t; r is the discount rate (the social discount rate is 10%); t is time. Assuming that every year has same benefit and the county can benefit from it in an infinite period after finishing the land remediation, namely the value of t is infinity, and then the formula (1) can be simplified as:

[1] Namely, to achieve the centralization and scale of the scattered abandoned land (homestead) by closing and merging villages and central village construction; remote area and the waste land in villages, through the replacement of the farmland in the farmland in remote area can be used for develop economic forest, the waste land in villages can be used for arable land, and the rational allocation and the effective use of land resources can be achieved.

$$PVB = \lim \frac{(1+r)^{t+1}-1}{r(1+r)^{t}} B_t = \frac{1+r}{r} B_t \quad (4)$$

Related data show that one hectare woodland can absorb 1005kg CO^2 per day and release 735kg O^2; The costs of maintaining CO^2 and releasing O^2 are respectively 273.3 RMB/T and 369.7 RMB/T. According to the above index and the different homestead remediation potential in different periods in Qin'an, we can respectively calculate the value of the air purification of the economic forest transferred from remediation land, and calculate the current value of its purified air by using formula (4), the results are shown in Table 2.

Table 2. The Measurement of the Value of Puried Air of Hollow Villages' Remediation Potential in Qin'an

year	Land remediation potential (hm^2)		Maintain CO_2		Release O_2		B_t (10 thousand)	PVB (10 thousand)
	Added value	Totals	Weight (T)	Value (10 thousand)	Weight (T)	Value (10 thousand)		
2010		204.8	75125.8	2053.2	54942.7	2031.2	4084.4	44928.6
2020	540.5	745.3	273394.7	7471.9	199945.4	7392.0	14863.9	163502.5
2030	1402.9	1607.8	589762.9	16118.2	431319.1	15945.9	32064.1	352705.0

4.1.1 The Value of Water Conservation

Forest has significant rainfall and water conservation function. Domestic and foreign researches show that the added value of annual water conservation is $3000m^3/hm^2$, and its benefit can be calculated by replacement cost approach. In Gansu province, the current agricultural irrigation water price is 0.06-0.15 RMB/m^3. With the increasingly prominent contradiction between water supply and demand, we use the high value (0.15 RMB/m^3) to calculate the replacement value of water conservation of forest. Calculating with the above indicators, it showed that after Qin'an finished transferring its abandoned homestead to forest, its replacement value of water conservation is 92,200 RMB, and the present value of its benefits of water conservation function calculated by using formula (2) is 1,013,800 RMB (Table 3).

4.1.2 The Value of Water and Soil Conservation

Forest has great function of soil and water conservation. On one hand, the crown can directly reduce the impact of rain on the ground, so as to effectively reduce soil and water loss; on the other hand, through the function of dispersing, blocking and interception of slope runoff, forest can reduce the surface runoff of 60%, well-structured forest can reduce soil erosion more than 90%. According to the measurement of research, every hectare forest can annually conserve soil 22.5-30.0t, so calculated with the lower limit value 22.5t, after Qin'an finishes its land

remediation of abandoned homestead and transfers into forest, it can reduce soil erosion of 4608t. Calculated with 10 RMB/t[15], we can use formula (2) to calculate that the present value of the benefits of its soil and water conservation function is 506,900 RMB (Table 3).

4.1.3 The Value of Biodiversity

Forest creates the condition of life and reproduction for biological population, so that biodiversity can be maintained and show great biodiversity value. The value of biodiversity includes direct value, indirect value, option value, genetic value and existence value, so it is very complicated to assess. Research shows that the annual biodiversity value in intact forest is 15 $/hm2 [15], and the annual direct value in use of biodiversity in China is 1.8 trillion, and the indirect value in use is 37.31 trillion. Calculated with these indicators, after Qianan finishes its transfer, its biodiversity value will be up to 3,072 $, according to the current exchange rate it will be 19,300 RMB, so the present value of its benefits of biodiversity protection by using formula (2) is 211,900 RMB (Table 3).

4.2 Economic Benefits Analysis

Qin'an is one of the national forest demonstration counties. By the end of 2008, the fruit planting area in Qin'an had reached 47,900 hm^2, and the annual output value was 441 million RMB. The per-capita income of fruit farmers was 771 RMB, which accounted for 35.6% of rural per-capita income. Economic forest has shown enormous economic benefits. According to the economic forest area in Qin'an and its annual output value, the average annual production value on unit area of forest is 9206.68/hm^2. Based on these indicators, we can infer that after finishing the land transfer, Qin'an can increase its annul economic benefits by 1,885,500 RMB; the economic benefits in 2020 and 2030 respectively will be 6,861,700 RMB and 1,480,200 RMB. From the formula (2), we can calculate that the present value of economic benefits in three periods respectively is 20,740,500 RMB, 75,478,700 RMB and 162,822,000 RMB (Table 3).

Table 3. The Estimation and Measurement of the Ecological and Economic Benefits of the Land Remediation of hollowing Villages in Qin'an

year	Economic benefits		Ecological benefits(10 thousand)							
			Air purification		Soil and water conservation		Maintaining of soil and water		biodiversity	
	Bt	PVB	Bt	PVB	Bt	PVB	Bt	PVB	Bt	PVB
2010	188.55	2074.05	4084.42	44928.62	9.22	101.38	4.61	50.69	1.93	21.19
2020	686.17	7547.87	14863.86	163502.46	33.54	368.92	16.77	184.46	7.04	77.47
2030	1480.2	16282.20	32064.09	352704.99	72.35	795.84	36.17	397.92	15.12	166.33

5 Conclusion and Discussion

(1) The factors affecting the village hollowing evolution in rural places are complicated. At different stages of development, the dominant factors, driving forces and forms are all different. Since China's reform and opening-up, the rural economy in China developed rapidly. The increase of farmers' income and their desire to improve their living conditions promote rural housing boom. Meanwhile, a lot of original homesteads are abandoned. It is very common for the people in rural place to build a new homestead but do not demolish the old one. Outside expansion and inside decaying are the basic features of the village hollowing evolution in villages. In the new era, with the accelerating development of urbanization, strategic transformation of urbanization and the continuous improvement of the quality of urbanization in China, the urbanization of population will be the main driving factor of the village hollowing evolution in the future, and the village hollowing evolution will exacerbate with the high level of urbanization, therefore the land remediation potential is huge. Calculated with the county urbanization level and migration rate of rural surplus labor which may reach in the future, it showed that the remediation potential of rural homestead in 2020 and 2030 would respectively increase by 540.5 hm^2 and 1402.95 hm^2.

(2) The construction of economic forest is an effective way to achieve the combination of the development of rural economy and ecological environment construction. By the combination of the rearrangement of villages, constructing the central village and land remediation of hollow villages, through the land replacement and the adjustment of the land layout, Qin'an can take use of the land gained from the remediation of hollow villages to plant economic forest which will have huge ecological and economic benefits. The results show that the economic benefits of transferring the abandoned homestead in Qin'an into economic forest will be 20,740,500 RMB, and the comprehensive ecological benefits will be 41,001,800 RMB, and the present value of ecological benefits will reach 451,018,800 RMB.

(3) Urbanization is one of the basic socio-economic development strategies in China, and its improvement mainly depends on the expansion of the non-agricultural transfer scale of rural surplus labor and the increase of the stable urban employment rate and migration rate. However, due to the incomplete social security system in China, especially the incomplete system focusing on migrant workers, there are low stable urban employment rate and migration rate of rural surplus labor, which results in an amphibious phenomenon. Therefore, at this stage, the village hollowing evolution mainly performs as the seasonal vacancy. The vacant rate of rural homestead is higher than the abandoned rate. By analyzing the development trend, we can tell that, with the strategic transformation of urbanization in China, the implementation of new urbanization strategy and the improvement of the social security system for rural workers, the urban migration rate of rural surplus labor will gradually improve, and the rural homestead will also converted from the vacant to the abandoned.

The scarcity of land sources in China is severe, with the advance of the process of urbanization and the increase of the construction land in urban area, the contradiction

of people and land is increasingly prominent. Therefore, we should aim for protecting economic and social sustainable development, make a good use of land resources, strengthen the rural housing construction planning and management, strengthen land remediation of rural homestead, improve the land remediation in villages and the construction of a new countryside, combine the economic development and ecological construction, and promote the sustainable and healthy economic development of rural society.

References

1. Chu, Y.: The Relationship of Urbanization Process and Regional Economic Development. Nankai Economic Studies 42, 23–28 (1994)
2. Meng, X.: On the Mechanism of Urbanization. ACTA Geographica Sinica, 441–450 (1992)
3. Lu, X., Zhu, S., Zhang, Z.: The Empirical Analysis of the Transfer Scale of Rural Labor. Chinese Rural Economy, 35–41 (2006)
4. Ma, X., Li, Q., Shen, Y.: Morphological Difference and Regional Types of Rural Settlements in Jiangsu Province. ACTA Geographica Sinica, 516–525 (2012)
5. Liu, Y., Liu, Y., Zhai, R.: Geographical Research and Optimizing Practice of Rural Hollowing in China. ACTA Geographica Sinica, 1193–1202 (2009)
6. Zhang, Z.: On the Theory of Vacant Village in Hebei Province. Journal of Hebei Normal University (Natural Science), 573–576 (1998)
7. Xue, L.: Study on the Inner-decaying Village and the Countermeasures with Jiangsu Province as the Case. City Palnning Review, 8–13 (2001)
8. Cheng, L., Fen, W., Jiang, L.: The Analysis of Rural Settlement Hollowing System of the Southeast of Taiyuan Basin. ACTA Geographica Sinica, 437–446 (2001)
9. Feng, L.: Hollow Village Phenomenon and Regulatory Mechanism Under the Background of Urbanization. Theory Horizon, 174–175 (2008)
10. Tian, G.: The Spatial Distribution of Rural Settlements and Their Dynamic Change in China Using GIS. Remote Sensing Information, 32–35 (2003)
11. Ji, G., Wang, A.L., Zhu, Z., Li, Y.: Study On Land Reclamation Potential of Rural Residential Area——Taking Muping County as an Example. Journal of Shandong Agricultural University (Natural Science), 597–602 (2011)
12. Li, X., Gan, Z., Ou, M., et al.: Potential Estimation and Layout Optimization of Rural Residential Land Consolidation: A Case Study on Jiangdu City in Jiangsu Province. Scientia Geographica Sinica, 150–156 (2013)
13. Kong, X., Liu, Y., Zou, Y.: Calculation of land consolidation potential and optimization off rural residential areas based on households' willingness. Transactions of the Chinese Society of Agricultural Engineering, 296–300 (2010)
14. Song, W., Zhang, F., Chen, X.: The Calculation Method of Land Arrangement Potential in Rural Residential Areas in China. Guangdong Land Science, 43–46 (2006)
15. Wang, Y., Li, X., et al.: The Evaluation of the Eco- environment Benefits of the Land Consolidation in Village. Journal of Shanxi Educational College (Natural Science Edition), 79–81 (2005)

The Behaviors of Landscape Metrics to Changing MMU and Suggestions for Landscape Analysis

ChoNam Ng[1], Xijun Yu[2,*], Yujing Xie[1], and Jian Yang[2]

[1] Department of Geography, the University of Hong Kong, Hong Kong, China
[2] South China Institute of Environmental Sciences, Ministry of Environmental Protection, Guangzhou, P.R. China
yuxijun@scies.org

Abstract. Scale issue is one of the key issues in landscape ecology and related environmental science. Previous studies have addressed the effects of landscape metrics to changing spatial scale, grain size, and class resolution. However, Minimum mapping unit (MMU) as one of scale issues has not been well investigated. Using a real landscape, this paper examines the effects of changing MMU on landscape metrics. The sensitivity of landscape metrics to changing MMU is also calculated. The results show that changing MMU is different from changing spatial scale and grain size. The paper further provides some criteria or suggestions for selection of landscape metrics in landscape analysis. These findings are crucial for the understanding of the behaviours of landscape metrics across different scales and for selecting the appropriate metrics in landscape analysis or related environmental studies.

Keywords: Landscape metrics, Minimum mapping unit, Sensitivity, Criteria for selection.

1 Introduction

Scale refers to the spatial or temporal dimension of an object or a process[1]. In landscape ecology, scale is often characterized by grain and extent. Grain refers to the finest spatial resolution within a given data set. Extent refers to the size of the overall study area. Scale is one of the key issues in landscape ecology and has attracted much attention both in theory and in practice. When aggregating landscape units to relatively larger homogeneous unit, it is unavoidable that information will be lost. This process may further result in different influences on landscape pattern. To understand how landscape pattern affect landscape function, scale dependence should be firstly defined and well examined.

The importance of scale issue has been reflected in the actual discussion of several case studies in landscape ecology and related environmental science[2-5]. Much more research has been done into the effects of changing scale, which includes changing grain size (resolution)[6], changing extent[7, 8], and changing them both[9, 10]. Most landscape metrics can be categorized into three types: (1) metrics showed predictable

* Correspondsing author.

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 593–610, 2013.

responses with changing scale; (2) metrics exhibited staircase-like responses that were less predictable; and (3) metrics behaved erratically in response to changing scale[3, 9]. Wu[10] further distinguished these effects at class level, which showed simple scaling functions and unpredictable behavior, and at landscape level, with the third being staircase pattern.

However, Minimum Mapping Unit (MMU) as one aspect of scale issues has been neglected when comparing landscape dynamics from different data sources. A MMU is defined as "the smallest size areal entity to be mapped as a discrete entity" [11]. For remotely sensed data, it refers to the size of single grain or pixel. In vector data it is similar to the smallest "land unit" that is defined as ecologically homogeneous at the scale level concerned[12]. Changing MMU refers to the definition of areal units by progressively aggregating the original data into fewer, larger patch units for analysis. It is different from changing grain size because their effects on the spatial characteristics of landscape are different from each other[13, 14]. Changing pixel-based MMU will not modify the resolution of the original data and thus shape and edge information can be best maintained. Changing grain size will exponentially reduce the interior edge length of pixels within a patch. This difference results in the different effects on landscape metrics calculated on pixel attributes.

In a pioneer study based on simulated landscape by Saura[15], MMU size has been shown to be one of the factors influencing the values of landscape metrics. Results have illustrated that some landscape metrics cannot capture the characters of landscape changes if land cover data with different MMU or the distribution of patch size frequencies are to be compared. But the author selected only six landscape metrics based on simulated landscape, which cannot fully reflect the responses of commonly used metrics to changing MMU. Moreover, the differences between changing grain size and changing MMU have not been compared in the previous study. Considering a variety of landscape metrics and their wide usefulness in practice, more studies are needed to examine the behaviors of landscape metrics based on both simulated image and real data.

The paper provides an empirical study on the effects of changing MMU on landscape metrics both by increasing MMU size within a landscape. A set of fixed MMUs is defined, which will be used to systematically investigate the effects of changing MMU on landscape metrics. The behavior and sensitivity of landscape metrics will then be systematically examined and compared with changing grain size.

2 Data Processing and Methodology

One Landsat TM satellite image taken in 2002 was used as basic data source in the study. The data has a ground resolution of 30m, and were processed using the ERDAS IMAGINE software. It is proceeded through geometric correction, unsupervised classification and supervised classification, and GIS reclassification. In order to reduce the error of classification, the classified image is further correct based on a GIS model with the help of a digital elevation model[16]. The accuracy of classification is evaluated by using a confusion matrix and the overall accuracy is found to be higher than 90%. In order to avoid the "salt and pepper" phenomena caused by traditional pixle-based classification[17], patches smaller than nine pixels

will be eliminate according to the previous studies [18, 19]. CLUMP and ELIMINATE functions in ERDAS IMAGINE 8.7 are used in this process.

Six land use types are then categorized, which include cultivated land, orchard, forest, urban built up area, new development area, and water area. New development area refers to an area that is just bulldozed to make way for constructions[16], which appears high contrast character and suggests the dynamics of urbanization. Finally, a vector road layer is merged with the above classified images. Thus, the final land use classes consists of cultivated land, orchard, forest, urban built up area, new development area, road, and water area (Figure 1).

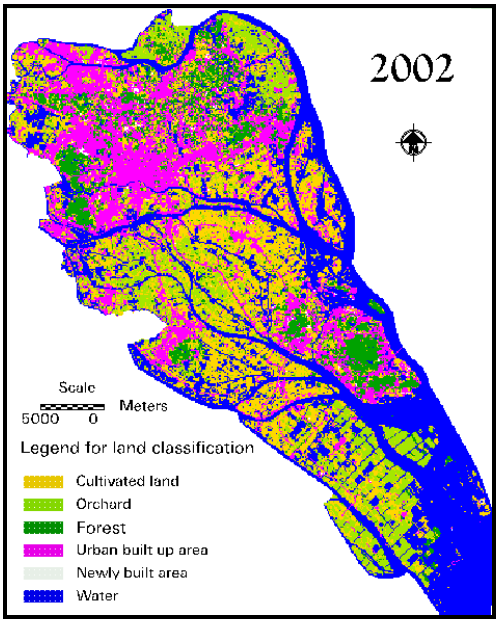


Fig. 1. Map of the study area

To investigate the effects of changing MMU on landscape metrics behavior, we systematically changed MMU size from 3 by 3 to 9 by 9 pixels with the spatial extent and classification kept constant. We select 9 pixels as the starting point because it can avoid "salt and pepper" phenomena without losing much information sensitive to highly localized surface features[20]. Newly built up area is approaching thorough elimination with increasing MMU size. Thus, we select 81 pixels as the end point in our study so as to keep land use classification consistent. This process can be achieved by using ELIMINATE command with majority algorithm in ERDAS IMAGINE software.

Landscape metrics were selected in the study in order to examine their behaviors to changing MMU. The definition and meaning of the selected metrics can be found in the report produced by McGarigal and Marks[21]. All these metrics are calculated for the whole study and each city at the class and landscape levels using the FRAGSTATS software[22].

3 Results

3.1 Effects on Area/Density/Edge Metrics

Area/density/edge metrics represent a loose collection of metrics that deal with the number and size of patches and the amount of edge created by these patches[23]. They are the most important and useful composition metrics in describing the basic aspect of landscape pattern. The commonly used metrics of this group at class and landscape levels include the CA, PLAND, NP, MPS, PD, TE, ED, LPI, PSSD, and PSCV.

When increasing MMU sizes at different levels, a direct effect is the removal and merging of small patches smaller than the MMU. This will have a significant influence on those patch-based metrics such as patch density, patch size and variability metrics. Compared with the original data, the absolute vale of the NP has a significant decrease with increasing MMU size (Table 1). It decreases from 7815 at original data to 4723 at MMUs1 at landscape level, a loss of about 40%. The rapid loss is observed when increasing MMUs and only 10% of the original NP can be identified when MMU size increases to MMUs5. On the contrary, the MPS has an obvious increase with increasing MMUs. This kind of patch-based metrics is either related to absolute value of patch types or standardized by total area, which includes PD, PSSD, PSCV, TE, ED, and LPI.

Table 1. Variation of patch-based metrics at landscape level with changing MMUs

	Orig. Data	MMUs1	MMUs2	MMUs3	MMUs4	MMUs5
NP	7815	4723 (-39.6)	2394 (-69.4)	1455 (-81.4)	963 (-87.7)	676 (-91.4)
PD	2.68	1.62	0.82	0.50	0.33	0.23
MPS	15.75	26.06	51.39	84.54	127.65	181.84
PSSD	347.88	475.27	733.12	1052.18	1340.59	1642.90
PSCV	2208.19	1823.75	1426.46	1244.56	1050.18	903.51

Overall, these metrics mainly represent the basic composition aspects of landscape pattern and are easily to be influenced by changing MMU size. The results obtained in the present study are similar to the findings of Saura[15], who used the same MMU size for all land cover classes based on simulated landscape. The higher sensitivity of these metrics is also reported when changing grain (or pixel) size or changing spatial context [10, 24].

3.2 Effects on Shape Metrics

Shape metrics are important for understanding the configuration character of a landscape at the patch, class, and landscape levels. In the present study, patch level metrics is excluded from analysis. Shape metrics compute the relativity of shape complexity, which is very important for characterizing the configurational aspects of the landscape. Most frequently used metrics of this group include the PAFRAC, LSI, MSI, AWMSI, MPFD, and AWMPFD.

When increasing MMUs size, landscape patches will become more regular and the internal length of edge will decrease. Thus, increasing MMUs size will decrease the LSI values at both class and landscape levels (figure not shown here). This tendency is consistent with changing grain size reported by Wu[10]. However, the MSI will increase with increasing MMUs size at both class and landscape levels (Figure not shown here), which is different from the observations of Wu[10] when changing grain size. This might be due to the fact that the MSI is calculated by equally weighting all the patches, independently of their size[15]. The MPFD also shows different responses between changing MMUs size and grain size (Figure 2). It increases with the increasing MMUs size, except that forest decreases at MMUs5. However, it shows generally decreasing tendency or insensitive to changing grain size[3]. Thus, these metrics should be further tested by more data.

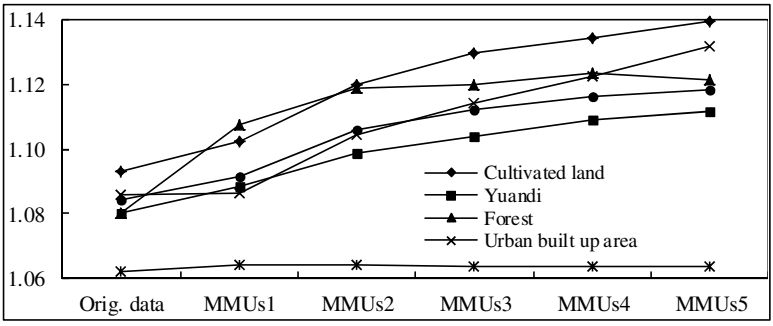


Fig. 2. Variation in MPFD with changing MMUs

The AWMSI is very robust to changes in MMU because it overcomes the limitations of the MSI by giving more weight to big patches[15]. It will increase first with increasing MMU and then decline. In the present study, it doesn't show the same tendency with changing MMU for different types (Figure 3). However, its variations to the original data are only minor. The AWMPFD shows similar trend as that of the AWMSI with a minor fluctuation because the former is able to suppress somewhat the abrupt large fluctuation that occurs with the AWMSI[10]. Thus, the AWMPFD is more insensitive to changing MMUs size than the AWMSI, which would be more appropriate for landscape comparison among different information sources. They have been used to measure the fragmentation of each land use class [25].

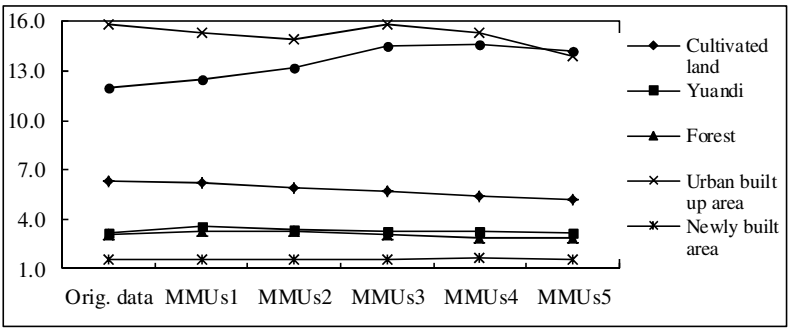
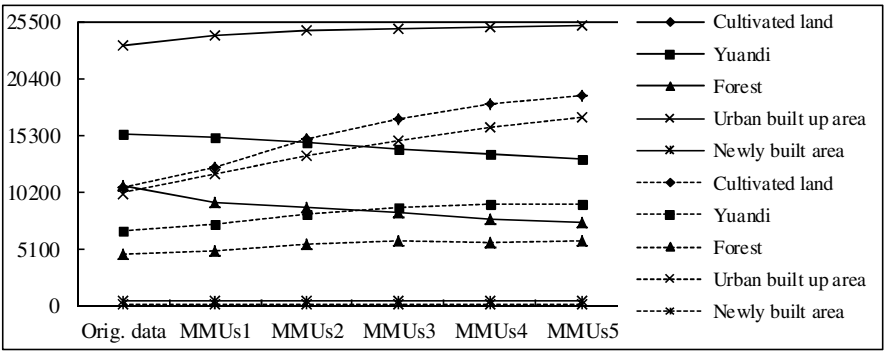


Fig. 3. Variation in AWMSI with changing MMUs

3.3 Effects on Core Area Metrics

Core area metrics can be considered as the counterparts of area, density and edge metrics because they are computed in the same manner as those area, density and edge metrics after eliminating the specified edge from all patches. Thus, a specified edge depth should be defined before calculation. 60 m is selected in this study to represent edge width for comparative study. Core area metrics are also important for the study of edge effects. Main core area metrics include the TCA, CPLAND, CASD1, CACV1, CASD2, CACV2, TCAI, and, MCAI.

According to its definition, the TCA should be directly confined by the CA, together with the edge depth assigned. However, comparison between two metrics doesn't show the same tendency among different land use types (Figure 4). When increasing MMU size, the CA increases for urban built up area but decreases for cultivated land, yuandi, and forest whereas the TCA increases for all four land use types. The differences between them are influenced by their shape regularity. Generally, patches with high interior/edge ratio (for example square or almost square) might have higher TCA than those with low interior/edge ratio (for example, narrow



Note: —— (CA) - - - - (TCA)

Fig. 4. Comparison of CA and TCA with increasing MMUs

elongated patches). The increase of the TCA for all four land use types with increasing MMU size indicates that the landscape patches become more compacted and regular after eliminating small patches. This change subsequently leads to the similar direction of increasing CPLAND value, which is also different from the variation of PLAND as shown above. The TCAI and MCAI will have a consistent increase with increasing MMUs.

3.4 Effects on Contrast Metrics

The contrast between the focal patch and its neighborhood patches can influence a number of important ecological processes[26]. Contrast metrics is a set of indices based on edge contrast at the patch, class, and landscape levels, which are measured by assigning different weights to each land use type within the landscape under consideration[23]. Thus, edge contrast weighting matrix need to be defined before calculating this kind of metrics. Frequently used contrast metrics include CWED, TECI, MECI, and AWMECI.

The TECI increases with increasing MMUs size at both class and landscape levels (Figure 5). Its approaching to 100 means all edge becomes more contrasted. On the contrary, the CWED exhibits an opposite trend with increasing MMU size, indicating that the amounts of edge in the landscape decreases. The variations of these two metrics show that landscape will become more contrasted together with the decline of the amount of all edges. Thus, elimination of small patches will result in high edge contrast and low edge length of the corresponding land use types. The AWMECI also increases with increasing MMUs size, which accords with the variation of the TECI.

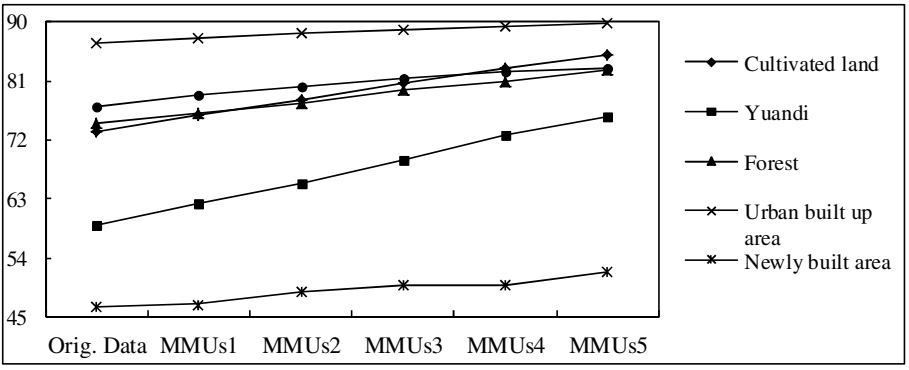


Fig. 5. Variations of TECI (up) and CWED (down) with changing MMUs

3.5 Effects on Contagion and Interspersion Metrics

Contagion and interspersion metrics are used to reflect the overall texture of the entire landscape mosaic[23]. Interspersion is based on edge amounts associated with patch type adjacencies. Contagion describes the tendency of patch types to be spatially

aggregated, which subsumes interspersion in that it reflects both the dispersion and intermixing of patch types. Major metrics of this group include AI, IJI, DIVISION, and CONTAG.

The AI is calculated from an adjacency matrix, which shows the frequency with which different pairs of patch types (including like adjacencies between the same patch types) appear side-by-side on the map[23]. The IJI metric defines how class patches are located in relation to other patches of the same or other classes within the landscape. The AI increases with increasing MMUs size at both class and landscape level (Figure 6). However, the IJI doesn't show the similar trend to the AI and its behaviors are erratic for different land use types (Figure 7). The CONTAG is only calculated at landscape level and the increase tendency is found with increasing MMUs, which indicates that landscape will become more aggregated due to the elimination of small patches (figure is not shown here).

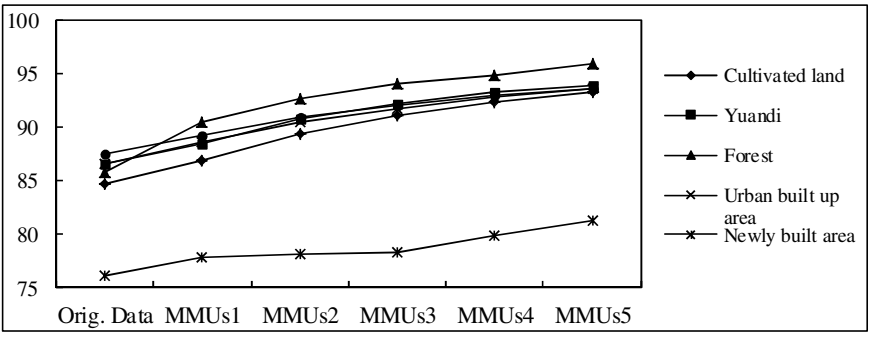


Fig. 6. Variations of AI with increasing MMUs

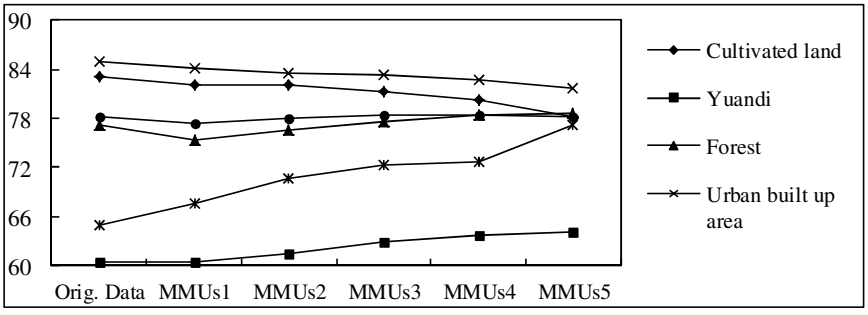


Fig. 7. Variations of IJI with increasing MMUs

3.6 Effects on Connectivity Metrics

Connectivity refers to the degree to which a landscape facilitates or impedes ecological flows[23]. FRAGSTATS provides three metrics at class and landscape

level. As the threshold distance and resist file are difficult to define in this study, only CONHESION is calculated in the study. It shows consistently increasing trend with changing MMUs size at both class and landscape levels (Figure 8). Thus, the landscape will become more physically connected with the elimination of small patches.

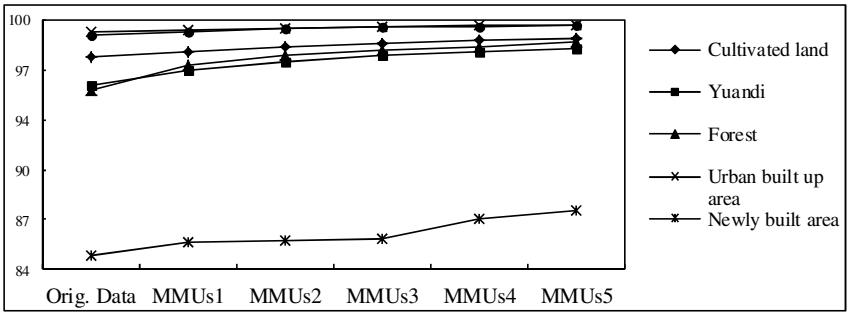


Fig. 8. Variation in CONHESION with changing MMUs

3.7 Effects on Diversity Metrics

Diversity metrics are composition metrics that quantify landscape pattern based on richness and evenness of patch types[23]. Richness refers to the number of patch types present whereas evenness refers to the distribution of area among different types. These metrics are also used to evaluate the heterogeneity of a landscape. Diversity metrics calculated in FRAGSTATS include PRD, SHDI, SIDI, MSIDI, SHEI, SIEI, and MSIEI.

All diversity metrics decrease with the increasing MMU size, but the variations among different MMU sizes are minor (Table 2). As the number of patch types is constant in the study, the decline of diversity metrics means the proportional distribution of area among patch types becomes more inequitable with increasing MMU. The SIDI shows the smallest variation among all these metrics while the MSIDI varies more than others, but it is still not very high in contrast with other

Table 2. Percentage variation of diversity metrics to the original data

	MMUs1	MMUs2	MMUs3	MMUs4	MMUs5
SHDI	-1.03	-1.44	-1.85	-2.15	-2.30
SIDI	-0.70	-1.08	-1.51	-1.94	-2.31
MSIDI	-1.59	-2.40	-3.34	-4.25	-5.04
SHEI	-1.04	-1.44	-1.85	-2.15	-2.30
SIEI	-0.71	-1.08	-1.52	-1.94	-2.31
MSIEI	-1.59	-2.40	-3.34	-4.25	-5.04

categorized metrics. Thus, although these kinds of metrics are criticized for not providing information on the actual composition of a landscape and its elements[23], the small variations with changing MMU size make them more appropriate for conducting comparison analysis from different data sources.

3.8 Sensitivity of Landscape Metrics to Changing MMU

To compare the sensitivity of different spatial metrics to MMU, Saura[27] compared different methods to evaluate their validities. Both absolute variations and the range of variation of a particular metrics cannot clearly describe the sensitivity because each of the metrics has different ranges of variation, and not all spatial metrics have a defined finite range of variation[27]. This makes the results difficult to compare among different landscape metrics. Studies have shown that landscape metrics generated from landscape TM image with MMU at 9 pixels can basically reflect the characters of landscape pattern without losing much information[20]. Thus, landscape metrics values calculated at this MMU size can be treated as the reference value for each land use class or the whole landscape. This study quantified the mean sensitivity of landscape metrics across different MMU sizes by comparing percentage change at each MMU size with those reference data[24]. The equation is as follow.

$$S_C = \frac{\sum_{i=1}^{5}\sum_{j=1}^{5}\frac{1}{M_i^R}\left|M_j - M_i^R\right|}{5 \times 5} \times 100\%$$

Where S_C is the sensitivity of each metrics to changing MMU

M_i^R is the reference value for land use class i at original data

M_j is the value of each metrics obtained at MMUs = j

i= number of land use classes

j = number of set of MMUs size

When calculating sensitivity at landscape level, the above formula will be changed as the following:

$$S_L = \frac{\sum_{i=1}^{5}\left|M_i - M^R\right|}{5}$$

Where S_L is the sensitivity of landscape metrics at landscape level

M^R is the reference value at landscape level at original data

High percentage change indicates greater sensitivity to changing MMU and vice versa. The results are shown in Table 3.

Table 3. Sensitivity of landscape metrics to changing MMU size

Category	Acronym	Percentage change (%)	
		Class level	Land. level
Area/density/edge metrics	CA	9.40	
	PLAND	9.40	
	NP	74.52	73.87
	PD	74.52	73.87
	MPS	673.82	498.56
	PSSD	229.54	201.49
	PSCV	42.85	41.60
	LPI	25.37	24.89
	TE	39.36	34.15
	ED	39.36	34.15
Shape metrics	LSI	38.39	33.44
	MSI	33.52	31.62
	AWMSI	10.02	14.71
	MPFD	2.62	2.24
	AWMPFD	0.75	0.78
	PAFRAC	5.85	3.44
Core area metrics	CPLAND	32.04	
	TCA	32.04	28.53
	CASD1	295.93	206.32
	CACV1	49.86	52.64
	CASD2	144.68	56.24
	CACV2	22.02	35.74
	TCAI	39.92	45.91
	MCAI	358.28	1.72
Contrast metrics	CWED	35.91	29.26
	TECI	7.37	5.19
	MECI	6.63	5.03
	AWMECI	7.42	5.40
Contagion and interspersion metrics	AI	5.85	4.83
	IJI	3.03	0.40
	DIVISION	0.14	0.67
	CONTAG		9.77
Connectivity metrics	COHESION	1.15	0.44
Diversity metrics	SHDI		1.75
	SIDI		1.51
	MSIDI		3.32
	SHEI		1.76
	SIEI		1.51
	MSIEI		3.32

There are considerable differences in the sensitivity of each of landscape metrics to the adjustment of MMU size. This is consistent at both class and landscape level. Among the seven categories, area, density, and edge metrics are the most sensitive metrics to changing MMU size. For example, the NP, PD, MPS, PSSD, and PSCV have the higher percentage change in this group, among which the MPS and NP have been suggested as very poor indicators based on simulated landscape analysis[15]. The result is also consistent with the previous studies on the behaviors of landscape metrics to changing grain size [28]. In fact, a common feature of these metrics is that they are patch based metrics, which will be evidently affected by the elimination of small patches caused by increasing MMU.

However, the TE and ED, which are both sensitive to grain size of the landscape, are relatively insensitive to changing MMU in this group because changing MMU just affects the edge length corresponding to small patches[15]. Therefore, the TE and ED are not as sensitive as the NP and MPS. At class level, the CA and CPLAND are of low sensitivity to changing MMU size and thus can be good composition metrics. However, this might be more related the statistical variation of each land use type during the processing of MMU adjustment since they are determined by the sum of the areas of the corresponding patch type. Similarly, core area metrics also have higher sensitivity as opposed to other categorized metrics.

Shape metrics are of relatively low sensitivity to changing MMU size, especially for those area weighted metrics such as the AWMSI and AWMPFD. This might be the consequence of the little weight giving to the smallest patch in comparison to the bigger ones[15]. The AWMSI is also recommended as a robust indicator for comparing landscape pattern with different MMU sizes[15]. In this study, however, the AWMPFD appears to be more insensitive to changing MMU than the AWMSI and is more appropriate for comparing landscape pattern. Compared with the AWMSI, the AWMPFD is able to suppress somewhat the abrupt large fluctuation[10]. The PAFRAC provides information on the irregularity of patch types[23]. It is also not sensitive to changing MMU, thus it has an advantage over edge density and other edge measures[29]. The behaviors of some metrics shown here are also different from those obtained from changing spatial extent. For example, the AWMSI is very sensitive to map spatial extent while presents a very robust behavior with changing MMU[8, 15].

Most of contagion and interspersion metrics are insensitive to changing MMUs. Study has demonstrated that contagion has a strong, negative correlation with the ED[29]. However, they have shown different sensitivity in the present study. Although contagion is affected by the number and evenness of patch types, spatial arrangement of patches, and grain size[3], its sensitivity to changing MMU seems to be relatively low in the present study. In contrast, the ED has higher sensitivity. At landscape level, the AI is computed simply as an area-weighted mean class aggregation index by weighting each class as its proportional area in the landscape[30]. The IJI and DIVISION are also insensitive to changing MMU, which can be used in landscape analysis when MMU issue is of concern.

Edge contrast metrics are mainly based on the percentage of the total length between patches. As mentioned above, changing MMU just affects the edge length

corresponding to small patches[29]. Because the total length of the corresponding small patches is very small as compared with the large patches, the effect on the percentage variation is not so high. Thus, edge contrast metrics are relatively insensitive to changing MMU size. Similarly, diversity metrics are also based on the percentage variation of each patch. Their sensitivity to changing MMU is also very low in general. This is consistent with the report by Wickham and Riitters[6]. However, some studies suggested that the use of the SHDI and SIDI might not be appropriate for comparing landscape change [31, 32]. Thus, it is highly recommended to carefully select landscape metrics even if MMU issue has been concerned.

4 Discussions

4.1 Comparison of Responses of Landscape Metrics to Different Parameters

The differences between changing grain size and changing spatial extent have been compared in some studies. Considering a variety of landscape metrics and their wide usefulness in practice, changing MMU should be considered based on both simulated image and real data. Based on literature review, comparison between changing MMU and changing grain size and spatial extent is well documented. The results are shown in Table 4.

The above results show that there are some differences among changing grain size, changing spatial extent, and changing MMU. Patch based metrics are consistent with those from changing grain size as both processes result in the loss of small patches. When increasing both grain size and MMU size, decreasing trends are found for some of area/density/edge metrics, such as the NP, PD, PSCV, TE, and ED but increase for the MPS, PSSD, and LPI. However there are some differences for shape metrics between changing these two parameters. When changing grain size, decreasing trend is found for the MSI, AWMSI, and AWMPFD. But they show increasing or inconsistent responses with increasing MMUs. The different response can be also found for contagion and interspersion metrics. The IJI will increase with increasing grain size but decrease with increasing MMUs. The CONTAG declines with increasing MMUs in this study, but decline or no consistent behavior is found when increasing grain size[3, 33]. A great difference is found in diversity metrics. Most of them exhibit inconsistent behavior with changing grain size, but decrease with changing MMU.

The above results further confirm that landscape metrics are very sensitive to scale issue and no one solution can solve all problems. Any landscape analysis should select the preferential metrics based on the specific requirement. It is strongly agreed with Baldwin et al.[24] that metrics responses should not be evaluated with the same threshold for all landscape metrics and for all locations. Similarly, metrics selection should also involve considering relevance for use in specific environmental monitoring application[34]. Thus, it is highly recommended that MMU issue should be firstly considered when landscape comparison between different data sources is of concern.

Table 4. Comparison of some reported landscape metrics responses from the literature

Metrics	Parameters		
	Grain or pixel	Spatial extent	MMU
NP	Dec.[3]	Inc. [3]	Dec.
PD	Dec.[3]	Dec. or NC [3]	Dec.
TE	Dec. [3]	Inc. [3]	Dec.
ED	Dec. [3]	Inc. then Dec.[29] or NC[3])	Dec.
MPS	Inc. [3]	Dec. or NC[3]	Inc.
PSSD	Inc. [3]	Inc.[3]	Inc.
PSCV	Dec.[3]		Dec.
LPI	Inc.[3]	NC [3]	Inc.
PLAND			NC
LSI	Dec.[3]	Inc.[3]	Dec.
MSI	Dec. or NC [3]	NC [3]	Inc.
AWMSI	Dec. [3]		NC
PAFRAC	NC [3]	NC[3]	Inc.
MPFD	Dec.[3]	NC[3]	NC.
AWMPFD	Dec. [3]	Dec. [9]	NC
TCA	Inc. then Dec.	Inc.	Inc.
IJI	Inc.	Inc.	Dec.
CONTAG	Dec.[3, 33] or NC [10]	Dec. or NC [3]; or Inc. [33]	Inc.
PR*	Dec.	Inc.	Dec.
SHDI	NC or Dec. [3]	NC or Inc.[3]	Dec.
SIDI	NC	Inc.	Dec.
SHEI	NC	Inc.	Dec.
SIEI	NC	Inc.	Dec.

Notes: "Inc."—Increasing, "Dec."—Decreasing, "NC"— Inconsistent or erratic.
PR is not calculated in the present study

4.2 Selection of Metrics Insensitive to Changing MMU

To be useful for landscape comparison, landscape metrics must be both reliable and valid for the questions under investigation. As landscape pattern is sensitive to increasing MMU as demonstrated above, using landscape metrics that have low sensitivity to MMU effects and most sensitivity to the actual landscape pattern changes through time would be more appropriate in practice. From this point, patch based metrics might not be appropriate for landscape comparison when considering this criterion, although they are widely used to analyze landscape dynamics that are mainly derived from the same data set [35, 36]. Saura[33] selected the LD (DIVISION in the present study) and LPI among the six indices to directly compare landscape fragmentation using data with different spatial resolutions. These two metrics are also robust to variations in MMU size and thus can be used for comparison between different data sources. In addition, landscape metrics should

capture landscape variations through time and thus those metrics shown little temporal variability will be removed from further analysis. In this study, it is recommended to select shape metrics, contagion and interspersion metrics, connectivity metrics and diversity metrics.

4.3 Integration of Empirical and Expert Knowledge

The above criterion provides a basic for selecting landscape metrics used in this study. However, an object based method is still needed to provide more accurate information. Landscape pattern metrics are commonly compared between remote sensing data and simulation generated landscape[37]. The combination of landscape metrics, remote sensing, and modeling provides a new direction for understand of the spatial and temporal forms of land use changes. However, it also results in much uncertainty because policy dimension is not often linked well with the modeling process. It should be noted that remote sensing is based on raster data with texture different from that of human generated planning maps. Herzog and Lausch [32] identified the issues of "hard borders" and "soft borders" characterized by vector based map and digital satellite image, respectively.

Expert knowledge should also be synthesized in selecting landscape metrics and conducting landscape comparison[32]. As landscape metrics have not been explicitly linked with ecological processes, experts could help to select the reliable and valid landscape metrics that are linked to ecological consequences under consideration. Study has also demonstrated that class level metrics are more appropriate for most ecological or planning consideration than landscape-level metrics [10]. In the present study, the effects of changing MMU to landscape metrics are well examined at class level, which can provide theoretical basis for landscape comparison. However, little consensus has been obtained at landscape level. This is partly because landscape metrics at landscape level will vary due to the differences of numbers of patch classes. The difference of MMU size for each patch class might also influence the behaviors of landscape metrics, which has not been systematically investigated. Thus, class level metrics might be more appropriate in the present study.

4.4 To Maintain an Acceptable Level of Accuracy when Changing MMU

In this study, both unsupervised classification and supervised classification algorithms are used to obtain land use categories. However, this traditional pixel-based classification process will result in a pixelised (salt and pepper) representation of land cover[38]. Many class patches with only one or two pixels will sharply increase the number of total patches and decrease the mean area of each land use class. This will further result in the fragmentation of land use[26, 39], and decrease the overall percent accuracy[20]. Thus, patches smaller than a certain number of pixels should be eliminated to avoid the "salt and pepper" phenomena without losing much specificity. Small patch types will be dissolved and merged into neighboring patch types. However, dissolving those small patch types will result in the statistical changes in each land use class, and thus might influence landscape pattern. In this study, the NP

will be greatly reduced if MMUs increase to a high level, for example at MMUs4 and MMUs5. This will also result in the great variation of land use statistics as compared with the original data, which cannot reflect the pattern of the original data. Thus, low MMUs might well reflect the actual landscape change in terms of both accuracy consideration and representation of landscape pattern.

5 Conclusions

Based on a real landscape, this paper investigates the response of commonly used landscape metrics to changing MMU size. The results suggest that changing MMU size will first result in the variations of land use statistics of each land use class. It has significant effects on landscape metrics, especially for those area/density/edge metrics. Patch based metrics appear the most sensitive to changing MMU and hence not recommended as indicators for landscape comparison, especially for different data source. In contrast, area weighted metrics and those that reflect relative values seem appropriate for measuring landscape pattern obtained from different data sources. Some new findings that are different from changing grain size and changing spatial extent are also obtained in this paper. The tendencies of landscape metrics variation are different between changing MMU size and changing grain/pixel size for most of shape metrics, contagion and interspersion metrics, and diversity metrics. Accordingly, the sensitivity of landscape metrics to different changing parameters is also different, which is then compared with some of the reported landscape metrics.

The sensitivity of landscape metrics to changing MMU is also calculated in the paper. It is recommended that those metrics with low sensitivity can be used in landscape comparison study. Some principles for selecting the suitable metrics for the specific study objectives are suggested. However, it must be noted that the response of landscape metrics to changing MMU size is a very complex and difficult issue, which cannot fully be resolved in this paper. More systematic simulation and empirical studies are needed to further evaluate and verify the generality and robustness of landscape metrics responses to changing MMU.

Acknowledgement. This research is financially supported by the General Research Fund, Research Grants Council of Hong Kong (No. HKU 748707H), Science and Technology Planning Project of Guangdong Province (No.2011B031200011), and Marine Fisheries Science and Technology Promotion Project of Guangdong Province (No. A201101I04), China.

References

1. Turner, M.G., Gardner, R.H., O'Neill, R.V.: Landscape Ecology in Theory and Practice: Pattern and Process, vol. 291. Springer, New York (2001)
2. Wiens, J.: Spatial scaling in ecology. Functional Ecology 3, 385–397 (1989)
3. Wu, J., Shen, W., Sun, W., Tueller, P.T.: Empirical patterns of the effects of changing scale on landscape metrics. Landscape Ecology 17, 761–782 (2002)

4. Leitao, A.B., Ahern, J.: Applying landscape ecological concepts and metrics in sustainable landscape planning. Landscape and Urban Planning 59, 65–93 (2002)
5. Lausch, A., Herzog, F.: Applicability of landscape metrics for the monitoring of landscape change: issues of scale, resolution and interpretability. Ecological Indicators 2, 3–15 (2002)
6. Wickham, J.D., Riitters, K.H.: Sensitivity of landscape metrics to pixel size. International Journal of Remote Sensing 16, 3585–3594 (1995)
7. Turner, M.G.: Landscape ecology: the effect of patterns on process. Annual Review of Ecology and Systematics 20, 171–191 (1989)
8. Saura, S., Martinez-Millan, J.: Sensitivity of landscape pattern metrics to map spatial extent. Photogrammetric Engineering and Remote Sensing 67(9), 1027–1036 (2001)
9. Shen, W., Jenerette, G.D., Wu, J., Gardner, R.H.: Evaluating empirical scaling relations of pattern metrics with simulated landscapes. Ecography 27, 459–469 (2004)
10. Wu, J.: Effects of changing scale on landscape pattern analysis: Scaling relations. Landscape Ecology 19, 125–138 (2004)
11. Lillesand, T., Kiefer, R.: Remote Sensing and Image Interpretation, vol. 157. John Wiley & Sons, Inc., New York (1994)
12. Zonneveld, I.S.: The land unit - a fundamental concept in landscape ecology, and its application. Landscape Ecology 3(2), 67–86 (1989)
13. Davis, W.A., Peet, F.G.: A method of smoothing digital thematic maps. Remote Sensing of Environment 6, 45–49 (1977)
14. Goodchild, M.F., Quattrochi, D.A.: Scale, multiscaling, remote sensing, and GIS. In: Quattrochi, D.A., Goodchild, M.F. (eds.) Scale in Remote Sensing and GIS, pp. 1–11. CRC Lewis, Boca Raton (1997)
15. Saura, S.: Effects of minimum mapping unit on land cover data spatial configuration and composition. International Journal of Remote Sensing 23(22), 4853–4880 (2002)
16. Yu, X.J., Ng, C.N.: Spatial and temporal dynamics of urban sprawl along two urban–rural transects: A case study of Guangzhou, China. Landscape and Urban Planning 79(1), 96–109 (2007)
17. Bock, M., Xofis, P., Mitchley, J., Rossner, G., Wissen, M.: Object-oriented methods for habitat mapping at multiple scales – Case studies from Northern Germany and Wye Downs, UK. Journal for Nature Conservation 13(2-3), 75–89 (2005)
18. Asnera, G.P., Keller, M., Pereira, R., Zweeded, J.C.: Remote sensing of selective logging in Amazonia Assessing limitations based on detailed field observations, Landsat ETM+, and textural analysis. Remote Sensing of Environment 80, 483–496 (2002)
19. Knight, J.F., Lunetta, R.S.: An Experimental Assessment of Minimum Mapping Unit Size. IEEE Transactions on Geoscience and Remote Sensing 41(9), 2132–2134 (2003)
20. Knight, J.F., Lunetta, R.S.: An experimental assessment of minimum mapping unit size. IEEE Transactions on Geoscience and Remote Sensing 41(9), 2132–2134 (2003)
21. McGarigal, K., Marks, B.J.: FRAGSTATS: Spatial Pattern Analysis Program for Quantifying Landscape Structure Gen. Tech. Rep. PNW-GTR-351, 122 p. US Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland (1995)
22. McCarigal, K., Cushman, S.A., Neel, M.C., Ene, E.: FRAGSTATS: Spatial Pattern Analysis Program for Categorical Maps, Version 3.0. University of Massachusetts, Amherst, MA (2002)
23. McGarigal, K., Cushman, S.A., Neel, M.C., Ene, E.: FRAGSTATS: Spatial Pattern Analysis Program for Categorical Maps. Computer software program produced by the authors at the University of Massachusetts, Amherst (2002), http://www.umass.edu/landeco/research/fragstats/fragstats.html

24. Baldwin, D.J.B., Weaver, K., Schnekenburger, F., Perera, A.H.: Sensitivity of landscape pattern indices to input data characteristics on real landscapes: implications for their use in natural disturbance emulation. Landscape Ecology 19(3), 255–271 (2004)
25. Herold, M., Scepan, J., Clarke, K.C.: The use of remote sensing and landscape metrics to describe structures and changes in urban land uses. Environment and Planning A 34, 1443–1458 (2002)
26. Forman, R.T.T., Godron, M.: Landscape Ecology. Wiley and Sons, New York (1986)
27. Sadler, B.: From environmental assessment to sustainability appraisal. In: Environmental Assessment Yearbook 2002 (2002)
28. Saura, S.: Effects of remote sensor spatial resolution and data aggregation on selected fragmentation indices. Landscape Ecology 19(2), 197–209 (2004)
29. Hargis, C.D., Bissonette, J., David, J.L.: The behavior of landscape metrics commonly used in the study of habitat fragmentation. Landscape Ecology 13, 167–186 (1998)
30. O'Neill, R.V., Hunsaker, C.T., Jones, K.B., Riitters, K.H., Wickham, J.D., Schwartz, P.M., Goodman, I.A., Jackson, B.L., Baillargeon, W.S.: Monitoring environmental quality at the landscape scale. BioScience 47(8), 513–520 (1997)
31. Yue, T.X., Haber, W., Herzog, F., Cheng, T., Zhang, H.Q., Wu, Q.H.: Models for DLU strategy and their applications. Ekologia 17(suppl. 1), 118–128 (1998)
32. Herzog, F., Lausch, A.: Supplementing land-use statistics with landscape metrics: Some methodological considerations. Environmental Monitoring and Assessment 72, 37–50 (2001)
33. Turner, M.G., O'Neill, R.V., Gardner, R.H., Milne, B.T.: Effects of changing spatial scale on the analysis of landscape pattern. Landscape Ecology 3, 153–162 (1989b)
34. Trani, M.K., Giles, J.R.H.: An analysis of deforestation: metrics used to describe pattern change. Forest Ecology and Management 114, 459–470 (1999)
35. Griffith, J.A., Martinko, E.A., Price, K.P.: Landscape structure analysis of Kansas at three scales. Landscape and Urban Planning 52, 45–61 (2000)
36. Liu, Y.B., Nishiyama, S., Kusaka, T.: Examining landscape dynamics at a watershed scale using Landsat TM imagery for detection of wintering hooded crane decline in Yashiro, Japan. Environmental Management 31(3), 365–376 (2003)
37. Herold, M., Goldstein, N.C., Clarke, K.C.: The spatiotemporal form of urban growth: measurement, analysis and modeling. Remote Sensing of Environment 86, 286–302 (2003)
38. Bock, M., Xofis, P., Mitchley, J., Rossner, G., Wissen, M.: Object-oriented methods for habitat mapping at multiple scale - case studies from Northern Germany and Wye Downs, UK. Journal for Nature Conservation 13, 75–89 (2005)
39. Burel, F., Baudry, J.: Landscape Ecology: Concepts, Methods and Applications. Plymouth, Science Publishers, Enfield (2003)

Inter-annual Above-Ground Biomass Dynamics by Differential Association Model with Remotely Sensed Data in Xilingol River Basin, Inner Mongolia, China

Zongyao Sha[1] and Yongfei Bai[2]

[1] Department of Spatial Information & Digital Engineering, Wuhan University, Wuhan, China
zongyaosha@wuhan.edu.cn
[2] Laboratory of Quantitative Vegetation Ecology, Institute of Botany,
Chinese Academy of Sciences, Beijing 100093, P.R. China
yfbai@ibcas.ac.cn

Abstract. Grasslands occupy a great part of area in the global terrestrial landscapes and are important resources from both agronomic and ecological perspectives. Understanding vegetation dynamics is critical to make informed decisions for grassland management. Remote sensing technique has been widely applied in long-term grassland vegetation mapping. In this paper, we introduced the combination of remotely sensed data (including Landsat ETM+/China's small environment satellite HJ-1A/B and MODIS NDVI) with the aid of ground data in the study of vegetation dynamics in a typical grassland area (Xilingol River Basin, Inner Mongolia, China) over multiple years (from 2000 to 2012). The main work included two aspects. First, a differential association model (DAM) was proposed, which was able to reflect inter-annual above-ground biomass (AG-Biomass) fluctuations. Second, the remotely sensed imagery and field data were used to derive AG-Biomass fluctuation through the DAM with respect to time and space. Research findings from AG-Biomass fluctuation can provide key insight for administrators and local farmers to better understand vegetation growth and to take vegetation protection measures, which may have both economical and ecological benefit for the grassland.

Keywords: remote sensing, grassland, biomass, NDVI.

1 Introduction

Grasslands are important resources from both agronomic and ecological perspectives, as more than 40% of the global terrestrial landscape consists of grasslands and savannas [1]. Because of climate variability and human disturbances, vegetation degradation on arid and semi-arid grasslands can lead to serious ecological problems [2, 3]. Above-ground biomass (AG-Biomass) is one of the indicators that reflect vegetation health. Understanding spatio-temporal AG-Biomass patterns can help make informed decisions for sustainable development of grassland ecology. Moreover, AG-Biomass provides food for grassland pasture animals and animal feeding schema is primarily dependent on vegetation productivity. However, obtaining reliable estimates

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 611–622, 2013.

of spatio-temporal dynamics of vegetation properties especially over a large area still remains a challenge [4]. The most direct method to get AG-Biomass data is to conduct field surveys, which is usually point-based and requires that field researchers have specialized training and rich field work experience. Thus the field direct survey is time-consuming and expensive; therefore, biomass data acquired in such a way are scarce and often restricted to a very narrow scope [5]. Recent development of geo-spatial techniques raises the possibility for acquiring vegetation cover from remotely sensed data. A good case in geo-spatial techniques is remote sensing which offers an effective means to study grassland vegetation over large areas in a systematic way [5, 6].

Remotely sensed imagery varies from spatial and temporal resolutions [7]. To map vegetation cover, various remotely sensed products can be selected for better performance based on their characteristics. For example, the Landsat series are in medium spatial resolution and is frequently used for mapping land cover at local or regional scale [8, 9]. However, due to the low temporal resolution of the product (16 days), finding qualified Landsat scenes especially in vegetation growing season is always difficult due to heavy cloud cover, which makes such imagery unsuitable for continuous vegetation monitoring on grassland. Therefore, other studies are now resorting to remotely sensed imagery from platforms which may have higher temporal resolution while, at the same time, possesses similar or even higher spatial resolution. A good case in this point is China's small environment satellite HJ-1A/B program which was launched on September 6th, 2008 [10]. HJ-1A/B provides the same spatial resolution of imagery as Landsat (30m) but has much higher temporal resolution than Landsat (re-visit cycle in 48 h) [11]. This high temporal resolution imagery can possibly avoid the limitation of cloud contamination in vegetation monitoring. Coarse spatial resolution satellite images like AVHRR or MODIS have been acquired routinely for years, and their acquisition will continue in the future. Multi-period analysis of surface reflectance based on those images allows the examination of changes in AG-Biomass on grassland because of the high temporal resolution.

Grassland degradation has been observed in arid or semi-arid areas around the world [12]. In the semi-arid area in northern China, for example, grassland was experiencing degradation due to human activity and climate changes [3, 13]. To make informed decisions for grassland management (e.g., feeding schema for pasture animals), the grassland administrators and practitioners are interested to know spatial and temporal variability of AG-Biomass fluctuation. The aim of this study was to find a cost-effective and easy-to-apply solution for acquiring long-term AG-Biomass fluctuation in the typical grassland in Xilingol River Basin, Inner Mongolia, China.

2 Methodology

2.1 Study Area

Xilingol River Basin, situated 43°26′-44°29′N and 115°32′-117°12′E, is one of the most representative steppe zones in northern China. More than 90% of the land in the region is covered by steppe grassland [3]. From southeast to northeast, the altitude

gradually decreases from the highest value of 1608m to the lowest 902m above sea level (Figure 1). The basin has a total area of about 18,000 square kilometers (km^2) with an annual average temperature 1.71℃. In the past decade, the basin has an annual mean precipitation of 231 mm (from 288 mm in the south-east to 226 mm in the north), 60-80% of which occurs in May, June, July and August, almost coinciding with growing season. Over the last several decades, the human population keeps growing and urbanization becomes a critical concern for local government and grassland researchers due to its impact on vegetation degradation. The Xilingol river, which starts from three branches in the south-east, passes through Xilinhot city and ends at Chagan lake in the north. Herd husbandry is still the main income for the local farmers and important for the local economy. Over-grazing has been noticed and proper number of pasture animals based on vegetation capacity or an appropriate animal feeding schema is desired for sustainable development of the grassland. It is prerequisite for the local policy makers and practitioners (farmers) to have AG-Biomass fluctuation as well as information about how the related factors influence the spatial pattern and temporal dynamics of the fluctuation.

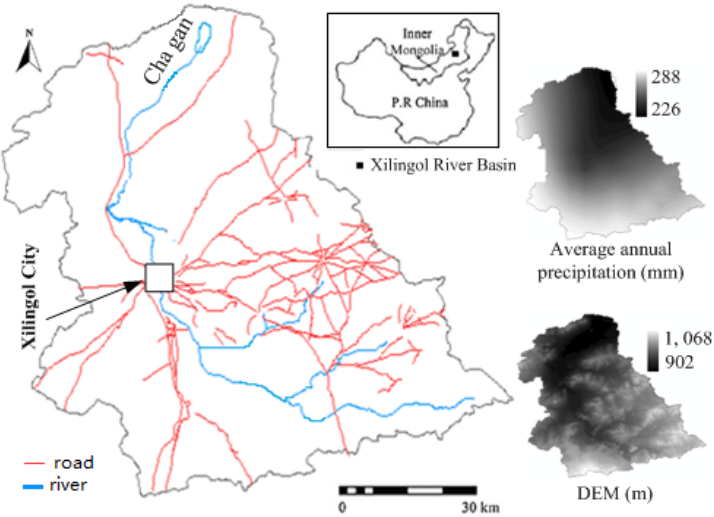


Fig. 1. The research region

2.2 Data Source and Methods

AG-Biomass data at field sites were collected in two separate years. The first field trip was done in the late August in 2004 (referred to as base period) and the second was done in the late July in 2012 (referred to as target period) which also used similar data collection approach as that of the first one to make sure that the collected samples were representative and comparable [14]. It took a few days to complete the field work for both periods. The dates were decided by local biologists on the evaluation that vegetation had reached peak greenness by data collection time. To inverse

regional AG-Biomass distribution, remotely sensed imagery from Landsat Enhanced Thematic Mapper (ETM+) and China's small environment satellite HJ-1A/B were used with the field data for the year 2004 and 2012, respectively. For the year of 2012, this study took use of HJ-1A/B. HJ-1A/B has comparable spectral bands to Landsat TM/ETM+. The HJ-1B image (Row: 05, Path: 60) on July 28th, 2012 was selected when there was little cloud; radiation correction was followed using Atmospheric Correction Of Hyperspectral/Multispectral Imagery (FLAASH) model. The HJ-1B image was geometrically corrected to the Landsat ETM+. MODIS/Terra MOD13A3 products (Vegetation Indices monthly L3 Global 1km, Collection 5), which contain, among other bands, monthly NDVI based on MVC and a Pixel Reliability criterion for pixels according to their quality, were downloaded from USGS (the United States Geological Survey) website in the study of mapping AG-Biomass fluctuation from the beginning of year 2000 to end of 2012 on a monthly basis. Other data includes the digital elevation model (DEM) with spatial resolution of 30m. Meteorological data (in monthly interval) were from the China's Meteorological Bureau (CMB, http: //cdc.cma.gov.cn). Preliminary analysis on the climate data found that the average temperature in the growing season varied within a very limited scope (within 2.5℃) during the study period and thus only monthly precipitation was included. Because CMB collects monthly data from more than 700 meteorological stations, spatial interpolation was performed to produce areal precipitation maps from year 2000 to 2012 following the method proposed by [15] and the maps were then clipped using the study area boundary.

The steps involved in the study were outlined in Figure 2. In the first step, the preprocessed Landsat ETM+ and HJ-1B images were used to derive NDVI maps ($NDVI_{ETM+}$ and $NDVI_{HJ}$) following the function,

$$NDVI = \frac{NIR - RED}{NIR + RED} \tag{1}$$

where NIR and RED are reflectance from the near Infrared band and Red band of the images, respectively. In the second step, the derived NDVI maps were combined with field AG-Biomass data, along with DEM, DEM derived slope and aspect, and reflectance bands of the images (Landsat ETM+ and HJ-1A/B), to produce regional AG-Biomass map AGBETM+ (biomass map from ETM+ for August 2004) and AGBHJ (biomass map from HJ for July 2012) using artificial neural network (ANN) model, respectively. The performance of each ANN model was evaluated using RMSE (total root–mean–square errors) and RMSEr (relative RMSE) as the indicators, which were defined as follows.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2}$$

$$\mathrm{RMSE}_r = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{\bar{y}} \times 100 \tag{3}$$

where $\hat{y}_i$ is the modeled AG-Biomass value, y_i is the observed AG-Biomass value, $\bar{y}$ is the mean of the observed values, and n is the number of samples.

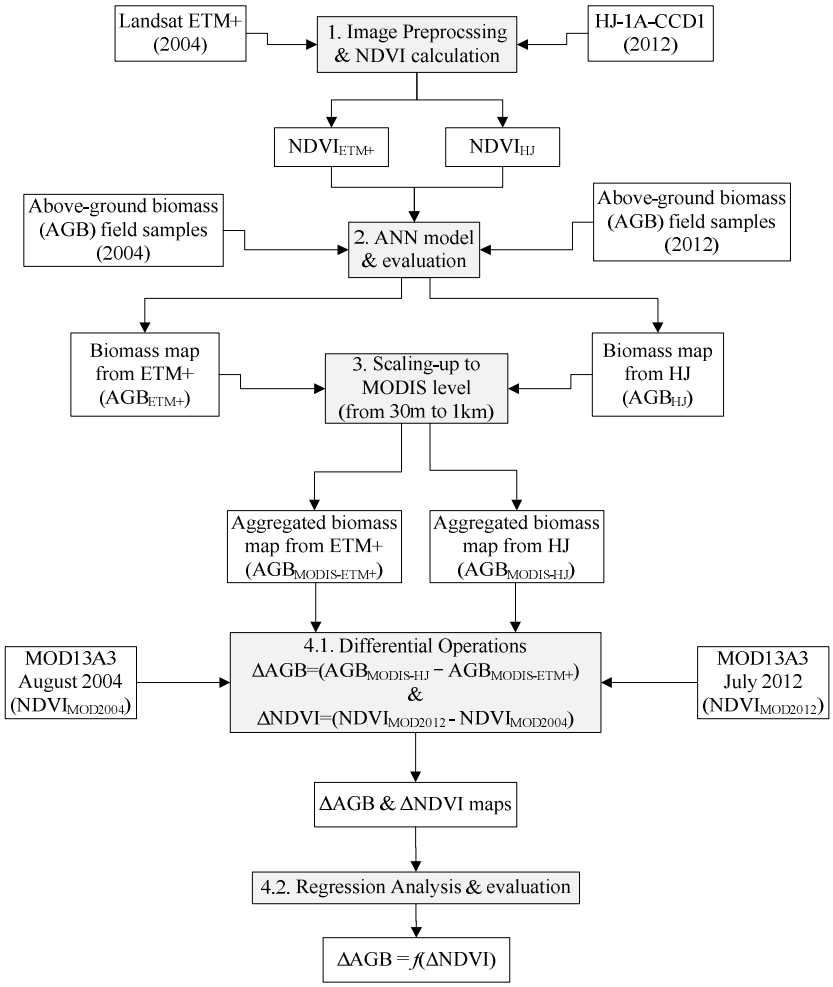


Fig. 2. Working flow used in the study

To take use of MODIS high temporal feature and to study spatio-temporal vegetation dynamics of the region, the AG-Biomass maps (AGB_{ETM+} and AGB_{HJ}) were resampled to a coarser spatial resolution at MOD13A3 level (1km) based on pixel average strategy in the third step. Two 1 km × 1 km biomass maps, $AGB_{MODIS-ETM+}$ and $AGB_{MODIS-HJ}$, were therefore generated in this step. In the fourth step, the $AGB_{MODIS-ETM+}$ and $AGB_{MODIS-HJ}$, along with MODIS/Terra MOD13A3 data for the base period (August 2004) and target period (July 2012), were used to obtain a linear

regression relationship between the differential AG-Biomass (ΔAGB) as dependent variable and differential NDVI (ΔNDVI) as independent. The process of preparing variables of the differential AG-Biomass (ΔAGB) and differential NDVI (ΔNDVI) as well as fitting the linear regression function for ΔAGB and ΔNDVI is termed as Differential Association Model (DAM) in this study. Both ΔAGB and ΔNDVI were processed through pixel differential with AG-Biomass and MOD13A3 NDVI, respectively, between the base period and target period. In this step, only pixels with Pixel Reliability marked as 0 (QA=0, indicating good data) in MOD13A3 for both periods were used. When the regression function was determined, the evaluation of the regression was followed using function (2) and (3) where the observed value was the original and the modeled value was the computed ΔAGB from the model.

Available literature indicates that AG-Biomass could be directly estimated by NDVI from remotely sensed imagery through empirical mathematical functions [5, 16]. However, our preliminary study showed that to best estimate AG-Biomass for the years having field data support, various different functions could be fit, which makes it difficult to estimate AG-Biomass for other years in a long-term AG-Biomass study when no field data is available. This difficulty inspired us to find other ways and therefore the above mentioned differential association model was proposed.

To map AG-Biomass dynamics for all periods in a long-term AG-Biomass study, we applied NDVI differential association from MOD13A3 between any two periods (in a monthly step) (Figure 3). Let $t_{y0\text{-}mi}$ be the base period, representing month i in year y_0. For any other target period $t_{y\text{-}mj}$ (j^{th} month in year y), we have AG-Biomass fluctuation (ΔAGB) from the base period $t_{y0\text{-}mi}$ to target $t_{y\text{-}mj}$

$$\Delta AGB \text{ (i.e., } AGB\text{-}t_{y\text{-}mj} - AGB\text{-}t_{y0\text{-}mi}) = f(NDVI\text{-}t_{y\text{-}mj} - NDVI\text{-}t_{y0\text{-}mi}) \quad (4)$$

where $AGB\text{-}t_{y\text{-}mj}$ and $NDVI\text{-}t_{y\text{-}mj}$ mean AG-Biomass and NDVI at period $t_{y\text{-}mj}$, $AGB\text{-}t_{y0\text{-}mi}$ and $NDVI\text{-}t_{y0\text{-}mi}$ mean AG-Biomass and NDVI at $t_{y0\text{-}mi}$ (the base period), and the function *f* is determined in Step 4. In this work, instead of using the absolute AG-Biomass value ($AGB\text{-}t_{y\text{-}mj}$ or $AGB\text{-}t_{y0\text{-}mi}$), relative AG-Biomass fluctuation was used to represent AG-Biomass increase (ΔAGB>0) or decrease (ΔAGB<0) for the target period relative to the base period and therefore can provide comparative AG-Biomass dynamics for different periods. Interannual AG-Biomass fluctuation was computed further by averaging the monthly AG-Biomass fluctuation in the growing season based on the result of monthly fluctuation of each year. After one year is set as base period, an average line relative to the base period (or the averaged ΔAGBs for all the years except the base period) can be calculated.

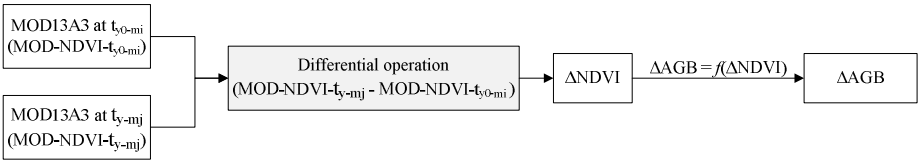


Fig. 3. AG-Biomass fluctuation based on differential association model

3 Results

3.1 Above-Ground Biomass Mapping from Landsat ETM+/HJ

The remotely sensed imagery from Landsat ETM+/HJ was used to derive the regional maps of AG-Biomass distribution for each period (August 2004 and July 2012), respectively. The process and ANN model were explained in detail in the previous study [14]. Specifically, AG-Biomass map for both periods was modeled by ANN in a similar process, including training, testing, and producing ultimate AG-Biomass distribution maps until reasonable accuracy for both periods was obtained (Table 1). The ANN models selected for both periods had similar architecture which was decided by its performance balancing against its complexity and used to create final AG-Biomass maps. It can be seen from the coefficient of determination that the modeled biomasses showed much close to the observed biomasses (R^2 = 0. 817, n=461 in 2004; R^2 = 0.702, n=289 in 2012).

Table 1. Result evaluation of AG-Biomass modeling from Landsat ETM+/HJ

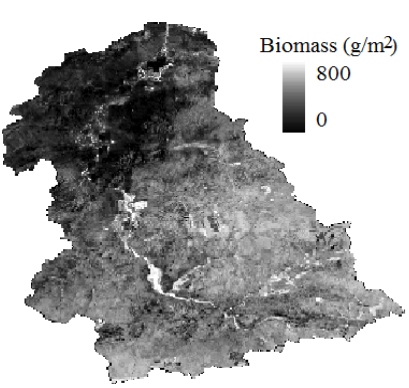
Year	number of samples	RMSE	$RMSE_r$(%)	R^{2*}	Note
2004	461	60.01	40.61	0.817	ANN model in [14]
2012	289	69.12	47.63	0.702	Current study

* Coefficient of determination of linear regression between the observed biomasses and the modeled biomasses for entire dataset (n=461 in 2004 and n=289 in 2012)

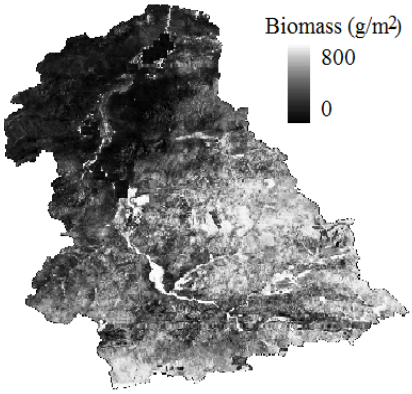
Both Landsat ETM+ and HJ-1A/B derived AG-Biomass maps have medium spatial resolution (30m). To facilitate the use of high temporal MODIS products, they were then resampled to a coarser level at spatial resolution as that of MOD13A3 (1km). To better understand the spatial distribution of vegetation productivity, the averaged NDVI for the study periods (from year 2000 to 2012) in the growing season is displayed together with two upscaled AG-Biomass maps (Figure 4).

Generally vegetation shows better condition in the south-east area than that in the north and north-west area for both periods. When comparing the maps to the precipitation and DEM maps (Figure 1), those maps present similar spatial pattern, i.e., high vegetation density (indicated by high NDVI or high AG-Biomass) is mainly noticed in area with high DEM or high precipitation in the south-east. Another important phenomenon observed is the impact of Xilingol river on vegetation growth. This dominant river originates from three branches located in the south-east and extends to the north until *Chagan* lake (Figure 1). Along the two sides of the river beds (especially the branches in the upper stream in the south and south-east), the vegetation has obvious better cover than the area away from the river. Due to the dry climate in the study region, upper stream of the river usually is under alternate dry and wet conditions even in high precipitation season, which provides favorable environment for vegetation growth along the wide sides of river channel. When arriving at lower part after passing the urban area (Xilinhot city), the river becomes deeper and narrower; its influence to vegetation seems to become less significant. In addition to topography (DEM) and river, human activities to vegetation growth

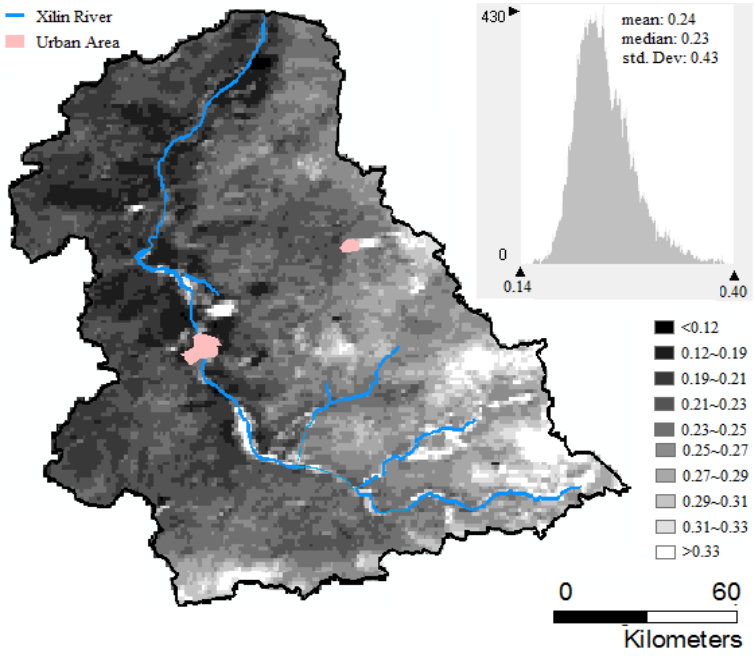
deserve attention. It was reported that human activities had a complicated impact on vegetation dynamics, depending on spatial and temporal scales at which assessments were conducted [17]. In the current study, we only included road network in our analysis.



(a) Upscaled AG-Biomass map estimated by ANN in 2004



(b) Upscaled AG-Biomass map estimated by ANN in 2012



(c) Average NDVI in growing seasons

Fig. 4. Upscaled AG-Biomass distribution maps and average NDVI in growing seasons (from 2000 to 2012)

Because road network can lead to easier accessibility to remote regions which were not accessible in the past, it is usually recognized as an important factor to reflect the intensity of human activities [18, 19]. Existing literature has proved that the road distribution demonstrated side effect on vegetation status in this region [13]. From the regional scale, however, the high density of road network in the central and south-east did not lead to poor vegetation growth if compared to other sparsely road-covered area. From micro-scale point of view, the influence of road network on vegetation still deserves further study. For example, the vegetation along the sides of road seemed to show lower vegetation productivity (low NDVI or AG-Biomass value). In summary, the vegetation status is controlled by many factors such as the climate (precipitation), river, topography (DEM) and road network, and those factors seem to have interacted impact on vegetation productivity.

3.2 Above-Ground Biomass Fluctuation by the Differential Association Model

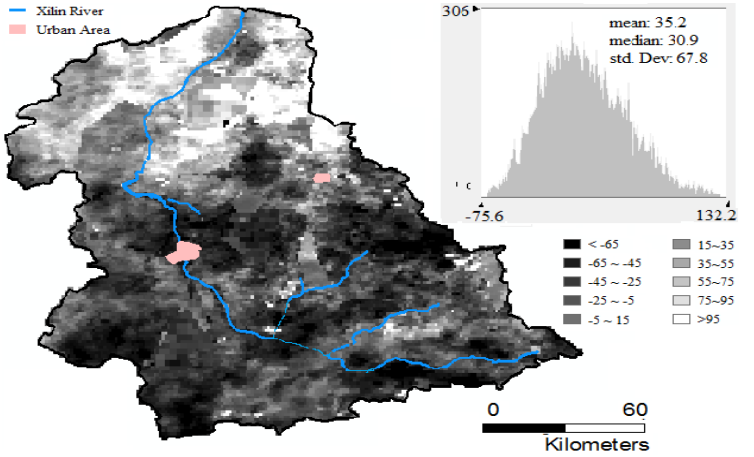
The DAM conducted on NDVIs and the upscaled AG-Biomass maps between the base period (late August, 2004) and target period (late July, 2012) produced a nearly normal distribution of a differential NDVI (ΔNDVI) map which has mean value close to zero (0.022) and a standard deviation 0.91 (Figure 5). More than 99% pixels of the map have values between the range of -0.27 and 0.37. This map was further used to explain AG-Biomass increase/decrease (ΔAGB) from the base period to the target. Linear regression model was built for the study region using ΔNDVI as the independent variable and AG-Biomass fluctuation (ΔAGB) as the dependent. Since our objective was to evaluate relative AG-Biomass fluctuation (not the absolute AG-Biomass) over the years, this regression model can serve the purpose well since ΔNDVI between any interested periods is readily available from MODIS product (MOD13A3) by DAM.

Linear regression functions were established taking the formula as

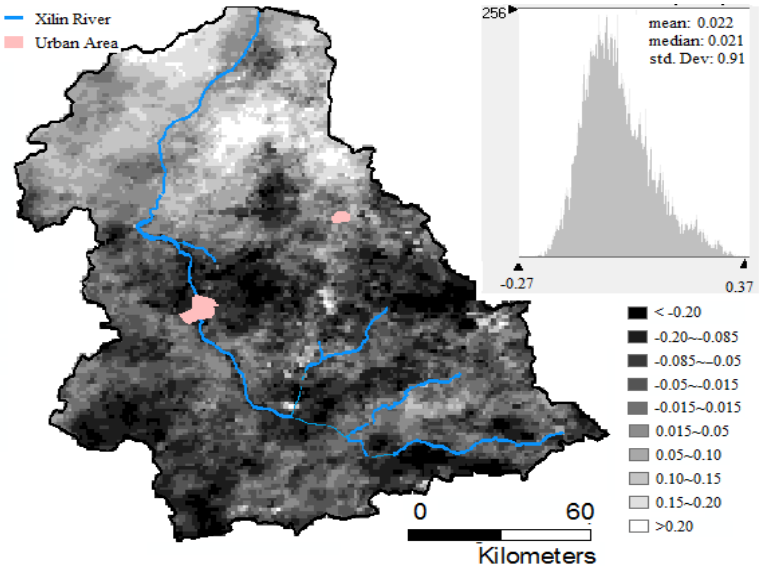
$$\Delta\text{AGB} = a \cdot \mathbf{\Delta}\text{NDVI} + b \tag{5}$$

where a is the coefficient for the differential NDVI (**Δ**NDVI) between any period interval (the base period and target period) and b is the intercept. Two regression models were built and their performance was compared. One is referred to as optimized model with the intercept b was set to 0, which is intended to facilitate consistent computation of AG-Biomass fluctuation for all pairs among multiple periods (regardless of which period is set as base period), and the other referred to as best fitting model in which b is set based on the best fitting function (by least square method). Both RMSE (total root–mean–square errors) and $RMSE_r$ (relative RMSE) were computed for performance comparison. In addition, statistical t-test was used to compare the difference between the estimated ΔAGB results from the two models (Table 2). Both of them show that ΔAGB is positively related to **Δ**NDVI (a>0). When the intercept was intentionally set to 0 (for the optimized model), the performance showed little decrease compared to the best fitting one (in terms of RMSE, $RMSE_r$ as well as R^2). However, the t-test verified that there was no significant difference between ΔAGB either estimated by the best fitting model or by the optimized.

The fit model (regression equation with b=0) was selected to derive the AG-Biomass fluctuations using the year 2000 as the base line year for the period from 2000 to 2012. Because the differential NDVIs (ΔNDVI) between any year can be easily obtained from the MODIS/Terra MOD13A3 product, the inter-annual AG-Biomass fluctuations over the years were acquired, which provides key insight to better understand vegetation growth.



(a) Differential AG-Biomass (ΔAGB)



(b)Differential NDVI (ΔNDVI)

Fig. 5. Differential results (ΔNDVI for NDVI and ΔAGB for AG-Biomass) between the base period (August 2004) and target period (July 2012)

Table 2. Performance comparison of AG-Biomass estimation based on two linear regressions between the relative AG-Biomass fluctuation (ΔAGB) and differential NDVIs (ΔNDVI)

Regression type	Regression function		RMSE	$RMSE_r$(%)	R^2	t-test*
	a	b				
Best fitting	189.34	8.82	17.45	27.51	0.852	0.07
Optimized	232.73	0	20.87	31.32	0.821	

*t-test is used for evaluating the null hypothesis of "no difference" between two ΔAGBs estimated by each regression function (best fitting and optimized) at the significance level of p = 0.05. The t-test value (0.07) means acceptance of the hypothesis because it is much less than the critical value of t distribution (1.96, two-tailed and n=18, 269) and thus the performance of the two regressions have no obvious performance in ΔAGB estimation.

4 Conclusion

Remote sensing technique has been widely applied in long-term grassland vegetation mapping. In this paper, we introduced the combination of remotely sensed data (including Landsat ETM+/China's small environment satellite HJ-1A/B and MODIS NDVI) with the aid of ground data in the study of vegetation dynamics in a typical grassland area (Xilingol River Basin, Inner Mongolia, China) over multiple years (from 2000 to 2012). We first proposed a differential associaton model (DAM) which was able to reflect inter-annual above-ground biomass (AG-Biomass) fluctuations. Remotely sensed imagery and field data were then used to derive AG-Biomass fluctuations through the DAM over the years from 2000 to 2012 in the study area. Research findings from AG-Biomass fluctuation can provide key insight for administrators and local farmers to better understand vegetation growth and to take vegetation protection measures, which may have both economical and ecological benefit for the grassland.

Acknowledgments. This work was Supported by the "Strategic Priority Research Program - Climate Change: Carbon Budget and Relevant Issues" of the Chinese Academy of Sciences (No. XDA05050402) and the Natural Science Foundation of China (Grant Nos. 41071249 and 41371371).

References

1. Blanco, L., Aguilera, M., Paruelo, J., Biurrun, F.: Grazing effect on NDVI across an aridity gradient in Argentina. J. Arid Environ. 72, 764–776 (2008)
2. Beerling, D.: Long-term responses of boreal vegetation to global change: an experimental and modelling investigation. Global Change Biol. 5, 55–74 (2008)
3. He, C., Zhang, Q., Li, Y., Li, X., Shi, P.: Zoning grassland protection area using remote sensing and cellular automata modelling: A case study in Xilingol steppe grassland in northern China. J. Arid Environ. 63, 814–826 (2005)

4. Cutler, N., Belyea, L., Dugmore, A.: The spatiotemporal dynamics of a primary degradation. J. Ecol. 96, 231–246 (2007)
5. Moreau, S., Bosseno, R., Gu, X., Baret, F.: Assessing the biomass dynamics of Andean bofedal and totora high-protein wetland grasses from NOAA/AVHRR. Rem. Sens. Environ. 85, 516–529 (2003)
6. Nordberg, M.L., Evertson, J.: Monitoring Change in Mountainous Dry-heath Vegetation at a Regional Scale Using Multitemporal Landsat TM Data. Ambio 32, 502–509 (2003)
7. Xie, Y., Sha, Z., Yu, M.: Remote sensing imagery in vegetation mapping: a review. J. Plant Ecol. 1, 9–23 (2008)
8. Mäkeä, H., Pekkarinen, A.: Estimation of forest stand volumes by Landsat TM imagery and stand-level field-inventory data. Forest Ecol. and Management 196, 245–255 (2004)
9. Röder, A., Udelhoven, T., Hill, J., Barrio, G., Tsiourlis, G.: Trend analysis of Landsat-TM and -ETM+ imagery to monitor grazing impact in a rangeland ecosystem in Northern Greece. Rem. Sens. Environ. 112, 2863–2875 (2008)
10. Li, Y., Xue, Y., He, X., Guang, J.: High-resolution aerosol remote sensing retrieval over urban areas by synergetic use of HJ-1 CCD and MODIS data. Atmospheric Environ. 46, 173–180 (2012)
11. Lu, S., Wu, B., Yan, N., Wang, H.: Water body mapping method with HJ-1A/B satellite imagery. Int. J. Appl. Earth Observation and Geoinformation 13, 428–434 (2011)
12. Nakayama, T.: Factors controlling vegetation degradation in Kushiro Mire. Ecol. Modelling 215, 225–236 (2008)
13. Xie, Y., Sha, Z.: Quantitative Analysis of Driving Factors of Grassland Degradation: A Case Study in Xilin River Basin, Inner Mongolia. The Scientific World J., Article ID 169724, 14 pages (2012), doi:10.1100/2012/169724
14. Xie, Y., Sha, Z., Yu, M., Bai, Y., Zhang, L.: A comparison of two models with Landsat data for estimating above ground grassland biomass in Inner Mongolia, China. Ecol. Modelling 220, 1810–1818 (2009)
15. Deng, X., Dong, X., Bo, H.: Research on Rainfall Spatial Interpolation Based on SRTM DEM. Disaster and Control Eng. 67, 55–60 (2009) (in Chinese with English abstract)
16. Yang, Y.H., Fang, J.Y., Pan, Y.D., Ji, C.J.: Aboveground biomass in Tibetan grasslands. J. Arid Environ. 73, 91–95 (2009)
17. Xu, D., Kang, X., Zhuang, D., Pan, J.: Multi-scale quantitative assessment of the relative roles of climate change and human activities in desertification—a case study of the Ordos Plateau, China. J. Arid Environ. 74, 498–507 (2010)
18. Peeters, D., Thisse, J., Thomas, I.: Transportation Networks and the Location of Human Activities. Geographical Analysis 30, 355–371 (1998)
19. Fore, L.S., Karr, J.R.: Assessing invertebrate responses to human activities: evaluating alternative approaches. J. N. Am. Benthol. Soc. 15, 212–231 (1996)

A New Spatial Interpolation Approach Based on Inverse Distance Weighting: Case Study from Interpolating Soil Properties

Jiaogen Zhou[1] and Zongyao Sha[2]

[1] Changsha Research Station for Agricultural & Environmental Monitoring, Institute of Subtropical Agriculture, The Chinese Academy of Science, Changsha 410125, China
zhoujg@isa.ac.cn
[2] Wuhan University, Wuhan, 430079, China
zongyaosha@wuhan.edu.cn

Abstract. This paper analyzed and discussed the effect of three similarity measure modes, i.e., spatial distance measure (SDM), non-spatial attribute similarity measure (NSASM) and their hybrid (SDM+NSASM), on interpolation accuracy by inverse distance weighting algorithm (IDW). The newly proposed approach differed from the traditional IDW one in that it also took use of non-spatial information (or environmental co-variables, either NSASM or SDM+NSASM mode) in the weighting vector computation for known spatial samples which are used to make interpolation for unknown locations. The proposed approach was tested on a case study in soil property (soil pH) interpolation. The result from cross-validation demonstrated that the NSASM and SDM+NSASM helped to decrease the prediction error compared to traditional SDM method, with good correlation between its environmental variables and the interpolation variable. Meanwhile, the IDW modeling result by NSASM and SDM+NSASM revealed closer local spatial relationships between soil pH and its environmental factors in comparison with the result from the traditional SDM.

Keywords: Attribute similarity measure, inverse distance weighting, spatial interpolation.

1 Introduction

Prediction of soil properties for interesting locations through a limited known soil samples at different locations has been interesting and attractive topic in the research field of spatial analysis and prediction. One typical approach for spatial prediction is through spatial interpolation. Most of spatial interpolation techniques involving with the hot issue are mainly based on spatial autocorrelation theory or soil-landscape model, as having been reviewed in literature [1-3]. The inverse distance weighting (IDW) method is considered as one of popular methods and has been implemented in many GIS software packages [4]. IDW is simple and easy to use, but has the apparent limitation of a constant distance-decay relationship [5-7]. Many methods were

F. Bian et al. (Eds.): GRMSE 2013, Part II, CCIS 399, pp. 623–631, 2013.

proposed to solve the problems related to this limitation [7-8]. In the paper, we try to address the issue which leads to possible high predictor error in interpolation process through IDW.

IDW method can be simplified to such a process that the estimation of an interesting spatial variable (e.g., soil property value) at an unmeasured location is calculated by the observations of a number of the measured samples spatially close to the objective location. Generally, the more similarity of soil properties between the close neighbors to the unknown location, the better prediction result will be obtained. Given an unobserved soil location, IDW method searches its k-nearest neighbors in the spatial space based on spatial distance measure (or SDM). An underlying assumption for SDM is that predicted variable of samples close to each other are more similar than the ones with larger distance. However, there have been reports that for spatial objects with spatial and non-spatial attributes, their non-spatial attributes may not be similar even close to each other in geographic space [9-10].

Considering spatial (given by geographical coordinates) and environmental variables both influencing spatial variables (e.g., soil properties) at every measured or unmeasured soil point, we extended the IDW approach which considers spatial distance only and presented two new versions' spatial similarity modes, i.e., AIDW and SAIDW for calculating IDW similarity metric. Different from IDW, AIDW and SAIDW search k-nearest neighbors in non-spatial attributes space (non-spatial attribute similarity metric, NSASM) and spatial and non-spatial attributes space (spatial and non-spatial attribute similarity metric, SDM+NSASM), respectively, and thus is believed to have more power in spatial interpolation. Our contributions are: (1) it first proves that NSASM and NSASM+SDM can present better the similarity of some environmental variables (e.g., soil properties) than SDM; (2) it proposes two new methods to build similarity measure metric in IDW modeling, i.e., AIDW and SAIDW. The following of the paper is organized as follows. Section 2 introduces three IDW similarity models based on different similarity measure modes. Section 3 presents a case study on real soil pH data. Finally, section 4 makes a conclusion.

2 IDW Models with Different Similarity Measure Modes

The key idea of traditional IDW approach is to search k-nearest neighboring known samples in the spatial space and to build a weighting vector for those neighboring samples based on their spatial distance to the unknown locations. The close distance indicates higher weighting for that known sample in the spatial interpolation process for the unknown location. Clearly, this process for calculation the weighting vector only takes spatial distance (usually in the form of Euclidean distance). In this paper, we extended IDW method by proposing two new methods, AIDW and SAIDW, for calculating the weighting vector for known samples. The main differences between AIDW, SAIDW and the traditional IDW way lies in the different similarity measure modes during the estimation of the weighting vector for the found k-nearest neighbors. The AIDW searches k-nearest neighbors for the un-sampled soil points

based on non-spatial attributes space, and calculate weigh coefficients through the difference between the un-sampled points and their neighbors in non-spatial attributes space. For SAIDW, given an unknown location, it tries to find its k-nearest neighbors and to build weight vector based on both spatial and non-spatial attribute space.

The above-mentioned three methods estimate the unknown value $\hat{z}(x_0)$ at the location of x_0 can all be written by the following equation,

$$\hat{z}(x_0) = \sum_{i=1}^{k} \lambda_i z(x_i) \tag{1}$$

where the weight coefficient λ_i is assigned to each of the k sampled locations nearest to un-sampled location x_0, and is limited to sum to 1, and $z(x_i)$ is the value at the x_i location. The weight coefficient λ_i is determined by:

$$\lambda_i = \frac{1/d_{i0}^{p}}{\sum_{i=1}^{k} 1/d_{i0}^{p}} \tag{2}$$

where p is a power parameter generally assumed to be 2, and d_{i0} is the Euclidean distance of the location x_0 to the measured location as Eq.(3).

$$d_{i0} = \sqrt{(Px_0 - Px_i)(Px_0 - Px_i)^T} \tag{3}$$

Here Px_0 and Px_i stand for the row vector of attributes at the location x_0 and one at x_i, respectively. Noting that Px_0 and Px_i differ in three interpolation methods, that is that Px_0 and Px_i are row vectors of geographical coordinates for IDW, ones of non-spatial attributes for AIDW, and ones of geographical coordinates and non-spatial attributes for SAIDW, respectively.

Additionally, R*-tree index technology [11] is used to reduce the complexity of k-nearest neighbors query for the three methods.

3 Case Study

3.1 Data Preparation

Two data sets were collected at landscape and watershed level, respectively, i.e., Pantang area (PTA) of 4.5km^2 and Jinjing watershed (JJW) of 133km^2 in Changsha County, both located in Hunan province, China (Fig.1). They were used as case study in this work.

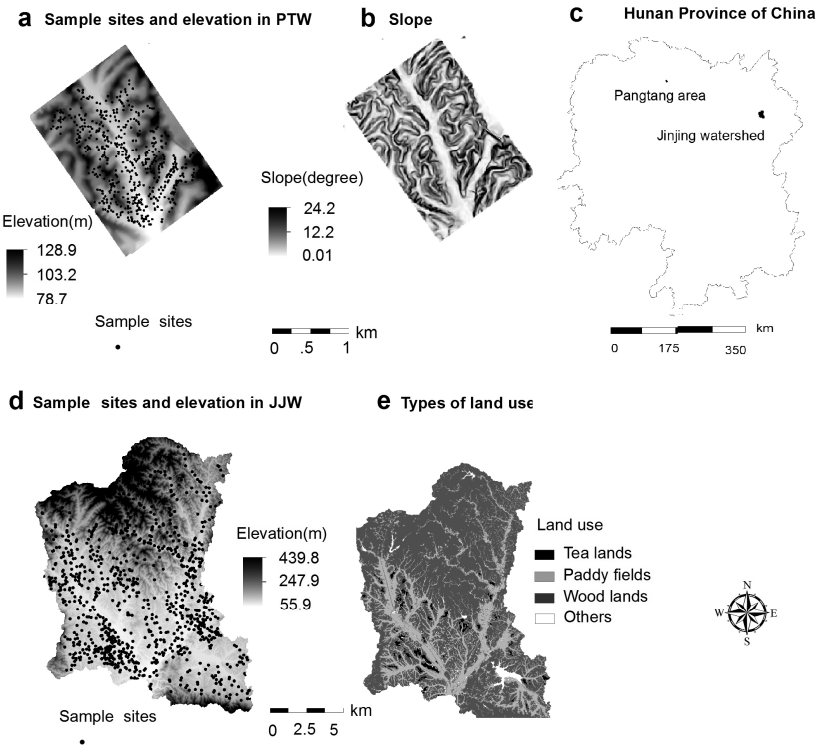


Fig. 1. Sampling sites and the environmental factors selected for soil pH interpolation in PTA and JJW

The most recent digital elevation model (DEM) with 5m spatial resolution for the two study areas and the land-use map in JJW were acquired from the land surveying department in Hunan province, China. Both the 523 soil points in PTA and the 955 soil points in JJW have the attributes of geographical coordinates, environmental factors (elevation and slope) and interpolation variable (e.g., soil pH).

Considering the field of land use is categorical not numeric, the averaged pH value of soil samples contained in each land-use patches was calculated, and then assigned to the corresponding land use class. To generate pH perdition map for JJW, a grid map with 20m×20m spatial resolution were generated. The environmental factors at the unmeasured points were achieved by overlaying the grid layer on corresponding raster maps, respectively. All operations on spatial data were implemented with Arcgis9.0 software. The pH value of the soil samples in PTA and JJW were first log-transformed before entering prediction models, and then the predicted values were back-transformed.

3.2 Model Calibration and Validation

To assess the performances of SDM, NSASM and SDM+NASDM on representing the similarity between the soil property (soil pH), the two real datasets were randomly split into two parts: 75% as training data and 25% as validation data, respectively. The mean similarity index (MSI) was calculated to assess how close the pH value of every soil point in validation dataset was to that of its neighbors in training data in Eq. (4):

$$MSI = \frac{1}{p \times k} \sum_{j=1}^{p} \sum_{i=1}^{k} | Z_{ij} - Z_j | \tag{4}$$

where k is the number of the nearest neighbors close to the jth sample, p for that of validation samples, Z_j for the pH value of the jth validation sample, and Z_{ij} is the pH value of the ith neighbor close to the jth sample. For a couple of soil samples, the smaller the value of *MSI*, the more similarity of their soil pH is.

The ten-fold cross-validation was used to evaluate the prediction performance of three IDW models with different similarity measure modes. For this, the two real data sets were randomly split into ten equally sized sub-datasets, respectively. Of the10 sub-datasets, a sub-dataset was retained as the validation data for testing the models, and the remaining 9 sub-datasets were used as training data. The cross-validation process was then repeated ten times, with each of the 10 subsamples used exactly once as the validation data. The prediction performance of the three models was evaluated by determining how close the predicted value $\hat{z}(x_0)$ was to the measured value at every point x_i. The two indices of mean absolute error (MAE), and root mean square error (RMSE) were calculated as:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| \hat{z}(x_i) - z(x_i) \right| \tag{5}$$

$$RSME = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left[\hat{z}(x_i) - z(x_i) \right]^2} \tag{6}$$

where $n, \hat{z}(x_i)$ and $z(x_i)$ are the number of validated samplers, the values of the predicted and measured, respectively.

3.3 The Effect of Different Measure Modes on Similarity between Soil Locations

The smaller values of MSI indicate more similarity of the objective attribute (e.g., soil pH) between soil points. In PTA, the values of MSI with NSASM higher than those with SDM and SDM+NSASM at every KNN size, while the MSI values of SDM+NSASM are close to that of SDM (Fig.2-a). In JJW, SDM made higher MSI than SDM and SDM+NSASM, and NSASM and SDM+NSASM had similar MSI (Fig.2-b). Additionally, Due to the more heterogeneous soil points as KNN size

increases, the similarity of soil pH between soil points decreased. As a result, the values of MSI for SDM, NSASM and SDM+NSASM had upward trends in PTA and JJW with increasing of KNN size.

Generally, SDM and SDM+NSASM helped to search more similar k neighbors for a given soil point than NSASM for soil pH in PTA, while NSASM and SDM+NSASM performed over SDM for soil pH in JJW. The comparative results indicate that the superiority of NSASM and SDM+NSASM over SDM may be limited by the correlation between its environmental co-variables and the soil pH variable. For example, the environmental factors of elevation and land use significantly correlated with soil pH in JJW, while a weak relationship among elevation, slop and soil pH was observed (Table.1). This implies that NSASM and SDM+NSASM can represent similarity between the pH variable of soil points with significantly good correlation.

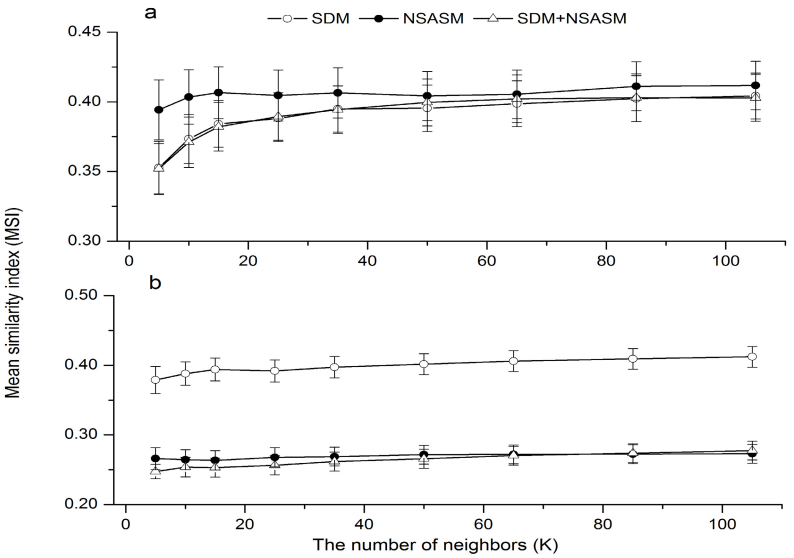


Fig. 2. Mean similarity index with different measure modes on two real datasets

Table 1. Correlation coefficient between soil pH and its co-variables

Soil pH	Co-variables		
	Elevation(m)	Slope(degree)	Land use
PTA_pH	0.08	–0.03	
JJW_pH	-0.20*		0.73*

PTA_pH, JJW_pH are for soil pH in Pantang area and Jinjing watershed, respectively. * Significant at 0.05 probability level.

3.4 Comparison of Prediction Accuracy on Validation Sets

The two indexes of MAE and RMSE were used to evaluate the prediction uncertainty of the three IDW models. Good prediction should have the lower values for MAE and RMSE. In the case of the dataset in PTA, IDW provided better predictions as lower MAE and RMSE than AIDW and SAIDW was observed and for the dataset of JJW, SAIDW gave a better prediction, followed by AIDW and IDW (Table.2). So, it is reasonably to believe that NSASM and SDM+NSASM helps to improve the prediction accuracy with good relationship between soil pH and its environmental co-variables based on our analysis.

Table 2. Root mean squared error (RMSE) and mean absolute error (MAE) on SOC and TSN of validation sets in Pantang area and Jinjing watershed

Soil pH	Items	Prediction methods		
		IDW	AIDW	SAIDW
PTA_pH	MAE	0.26	0.29	0.30
	RMSE	0.33	0.38	0.38
JJW_pH	MAE	0.29	0.21	0.19
	RMSE	0.37	0.30	0.27

PTA_pH, JJW_pH are for soil pH in Pantang area and Jinjing watershed, respectively.

3.5 Comparison of Spatial Prediction Maps

Good prediction map by a given spatial interpolation method should first give a range of the predicted inside the range of the certain soil property to be predicted, and it is more important that it should reflect the spatial patterns of its environmental factors influencing the soil property. For prediction of pH contents at un-sampled locations in JJW, three models gave various ranges, 3.3~5.6 for IDW, 3.3~4.6 for AIDW, and 3.3~4.6 for SAIDW, respectively. In the comparison with the range of pH observation values (3.1~5.8) of the 955 soil points, the estimations of the three IDW models are inside the range of observations. There is a trend that high pH with paddy field (averaged value of 4.2), middle pH (averaged value of 3.8) with forest land and relatively low pH with tea field (averaged value of 3.5) across the JJW.

The predicted map on soil pH by IDW model did not follow the trend (Fig.3-a), and it showed a high smoothing effect in the result of spatial interpolation. Different from IDW, AIDW and SAIDW reduced the smoothing effect because they took environmental variables into account in the prediction process (Fig.3-b and c). Meanwhile, AIDW and SAIDW produced pH maps with more reasonable and characteristic spatial patterns as influenced by the elevation and land use. First, the predicted pH values in the maps by AIDW and SAIDW generally followed the trend that soil pH is a significantly negative correlation ($p<0.05$) with elevation across the

JJW, except some regions in the northern parts, that is, the low pH prediction values with high elevation and relatively high ones with low elevation. Second, the predicted maps by AIDW and SAIDW both reflected that the spatial patterns of land use affected soil pH across the JJW as well. It is very clear that high pH values appear mainly in paddy fields, middle ones in forest fields, and low ones in tea fields. From this perspective, it is reasonably believe that NSASD and SDM+ NSASM helped to improve prediction precision than SDM.

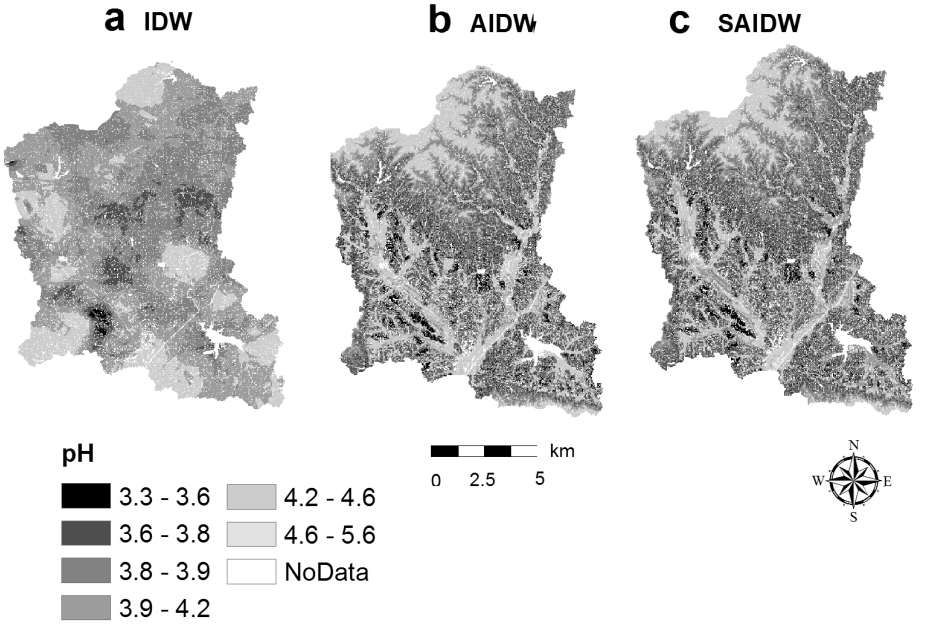


Fig. 3. The prediction maps by IDW, AIDW and SAIDW

4 Conclusions

In the paper, we discussed and analyzed the effect of the three attribute-similarity measure modes of SDM, NSASM and their hybrid on the prediction performance of inverse distance weighting approach. The experimental results on real datasets proved that NSASM and SDM+NSASM not only presented a better similarity of soil properties between soil points when compared to traditional SDM, but also helped to produce interpolation maps which present spatial patterns of other environmental variables influencing on soil properties. Considering that above-mentioned results obtained at landscape and watershed scales, the larger spatial-scale data sets are required to test our results in the future. It is also worth further discussion that how to adaptively choose optimal model parameters.

Acknowledgments. The research is supported by National Natural Science Foundation of China (41201299), Shanghai Natural Science Foundation of China (11ZR1432700), and the "Strategic Priority Research Program - Climate Change: Carbon Budget and Relevant Issues" of the Chinese Academy of Sciences (No. XDA05050402).

References

1. Goovaerts, P.: Geostatistics in soil science: state of the art and perspectives. Geoderma 89, 1–45 (1999)
2. McBratney, A.B., Odeh, I.O.A., Bishop, T.F.A., Dunbara, M.S., Shatar, T.M.: An overview of pedometric techniques for use in soil survey. Geoderma 97(3-4), 293–327 (2000)
3. McBratney, A.B., Santos, M.L.M., Minas, B.: On digital soil mapping. Geoderma 117, 3–52 (2003)
4. Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W.: Geographic Information Systems and Science. Wiley, Chichester (2001)
5. Lloyd, C.D.: Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain. Journal of Hydrology 308, 128–150 (2005)
6. Goovaerts, P.: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. Journal of Hydrology 228, 113–129 (2000)
7. Lu, G.Y., Wong, D.W.: An adaptive inverse distance weighting spatial interpolation technique. Computers & Geosciences 34(9), 1044–1055 (2008)
8. Tomczak, M.: Spatial interpolation and its uncertainty using automated anisotropic inverse distance weighting (IDW)-cross-validation/jackknife approach. Journal of Geographic information and Decision Analysis 2(2), 18–30 (1998)
9. Lin, C.R., Liu, K.H., Chen, M.S.: Dual clustering: integrating data clustering over optimization and constraint domains. IEEE Transactions on Knowledge and Data Engineering 17(5), 628–637 (2005)
10. Jiao, L.M., Liu, Y.L., Zou, B.: Self-organizing dual clustering considering spatial analysis and hybrid distance measures. Science China Earth Sciences 54(8), 1268–1278 (2011)
11. Beckmann, N., Kriegel, H.P., Schneider, R., Seeger, B.: The R*-Tree: An efficient and robust access method for points and rectangles. In: 1990 ACM SIGMOD International Conference on Management of Data, pp. 322–331. IEEE Press (1990)

Author Index