

# A Methodology for Estimating Leaf Area Index by Assimilating Remote Sensing Data into Crop Model Based on Temporal and Spatial Knowledge

ZHU Xiaohua<sup>1,2,3</sup>, ZHAO Yingshi<sup>3</sup>, FENG Xiaoming<sup>4</sup>

(1. Key Laboratory of Quantitative Remote Sensing Information Technology, Academy of Opto-Electronics, Chinese Academy of Sciences, Beijing 100094, China; 2. Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China; 3. University of Chinese Academy of Sciences, Beijing 100049, China; 4. Research Center of Eco-environmental Science, Chinese Academy of Sciences, Beijing 100085, China)

**Abstract:** In this paper, a methodology for Leaf Area Index (LAI) estimating was proposed by assimilating remote sensed data into crop model based on temporal and spatial knowledge. Firstly, sensitive parameters of crop model were calibrated by Shuffled Complex Evolution method developed at the University of Arizona (SCE-UA) optimization method based on phenological information, which is called temporal knowledge. The calibrated crop model will be used as the forecast operator. Then, the Taylor's mean value theorem was applied to extracting spatial information from the Moderate Resolution Imaging Spectroradiometer (MODIS) multi-scale data, which was used to calibrate the LAI inversion results by A two-layer Canopy Reflectance Model (ACRM) model. The calibrated LAI result was used as the observation operator. Finally, an Ensemble Kalman Filter (EnKF) was used to assimilate MODIS data into crop model. The results showed that the method could significantly improve the estimation accuracy of LAI and the simulated curves of LAI more conform to the crop growth situation closely comparing with MODIS LAI products. The root mean square error (RMSE) of LAI calculated by assimilation is 0.9185 which is reduced by 58.7% compared with that by simulation (0.3795), and before and after assimilation the mean error is reduced by 92.6% which is from 0.3563 to 0.0265. All these experiments indicated that the methodology proposed in this paper is reasonable and accurate for estimating crop LAI.

**Keywords:** assimilation; temporal and spatial knowledge; Leaf Area Index (LAI); crop model; Ensemble Kalman Filter (EnKF)

**Citation:** Zhu Xiaohua, Zhao Yingshi, Feng Xiaoming, 2013. A methodology for estimating Leaf Area Index by assimilating remote sensing data into crop model based on temporal and spatial knowledge. *Chinese Geographical Science*, 23(5): 550–561. doi: 10.1007/s11769-013-0621-x

## 1 Introduction

Leaf Area Index (LAI), defined as half the total developed area of leaves per unit ground horizontal surface area, is a key parameter for characterizing the structure and the functioning of vegetation. Affected by temporal and spatial discontinuity, the traditional point measurements can just provide LAI distribution in a limited area.

Remote sensing technology with the features of real-time dynamic and wide range monitoring can frequently provide the surface information, which is an important way to estimate regional scale LAI. Due to the restrictions of satellite operating cycle, remote sensing inversion results are temporal discontinuity. Process model can simulate continuous and long time series LAI data. However, influenced by parameterization, atmospheric

Received date: 2012-08-21; accepted date: 2012-12-11

Foundation item: Under the auspices of Major State Basic Research Development Program of China (No. 2007CB714407), National Natural Science Foundation of China (No. 40801070), Action Plan for West Development Program of Chinese Academy of Sciences (No. KZCX2- XB2-09)

Corresponding author: ZHU Xiaohua. E-mail: zhuxh@aoe.ac.cn

© Science Press, Northeast Institute of Geography and Agroecology, CAS and Springer-Verlag Berlin Heidelberg 2013

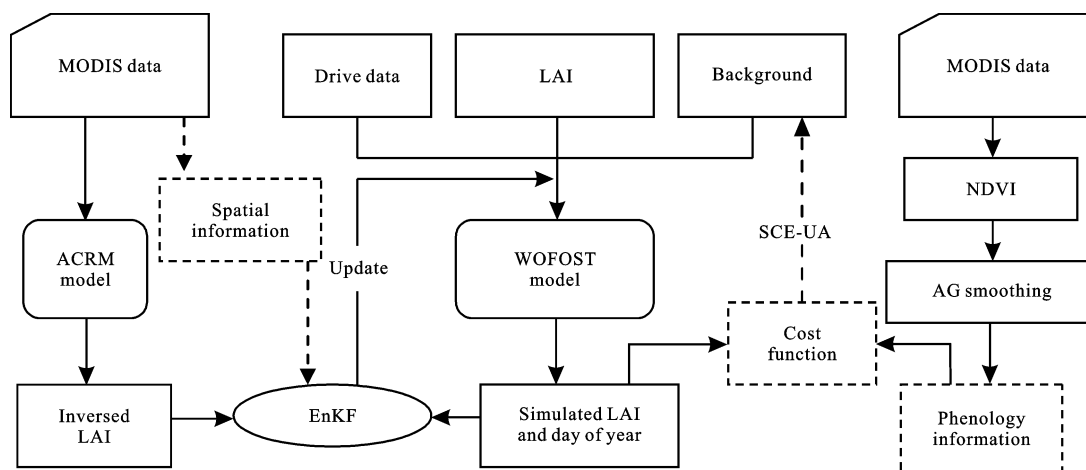
driving conditions and initial model state, there is a certain bias during model simulating process, which will bring much error in the LAI estimation. LAI estimation by remote sensing technology or by process model simulation, each have its own advantages and disadvantages (Dorigo *et al.*, 2007; Wang *et al.*, 2010; Yannick *et al.*, 2012). How to use multi-source data or multi-methods to achieve higher estimation accuracy of LAI is a focus in current research.

Data assimilation is the process by which multi-source information, such as remotely sensed data, field data, are incorporated into a process model to improve the model simulation results. Data assimilation makes full use of the model simulation and remote sensing observation to realize advantageous complementarities. In recent years, data assimilation technology is well applied in the research of atmosphere, ocean, *etc.*, and also attracted the attention of scholars in land surface research. Wang *et al.* (2010) used a crop model to extract the relationship of LAI with the crop growing, which is applied into LAI calculation based on remote sensing inversion model. Hazarika *et al.* (2005) coupled LAI retrieval from remote sensed data into a simulation model of the carbon cycle in land ecosystems (Sim-CYCLE) to accurately monitor the changes of global net primary productivity (NPP). Doraiswamy *et al.* (2003) built a Look-Up-Table (LUT) to calibrate the model parameters by minimizing the LAI value simulated by crop growth model and LAI value inverted by the Moderate Resolution Imaging Spectroradiometer (MODIS) data, and then they could use MODIS data (250 m) to calculate LAI value accurately. In many researches of data assimilation, the application of prior knowledge had attracted the attentions of scholars (Wang *et al.*, 2010; Xiao *et al.*, 2011; Yannick *et al.*, 2012). The prior knowledge here means the field data, spatial information and temporal information. In the study of coupling crop growth model with radiative transfer model, spectral library data (<http://spl.bnu.edu.cn/index.asp>) used as prior information was introduced into the cost function to achieve the optimal extraction of LAI (Wang *et al.*, 2010). In the research of real-time retrieval of LAI from MODIS time series data, Xiao *et al.* (2011) extracted temporal knowledge from MODIS LAI products to construct a dynamic model for providing short-range forecast of LAI, then, they used Ensemble Kalman Filter (EnKF) techniques to update LAI when there is a new observation.

The studies mentioned above have considered the application of prior knowledge, such as field data, temporal information, but still less to consider the influence of surface spatial heterogeneity during data assimilation and the introduction of spatial knowledge into regional assimilation process (Tian *et al.*, 2002; Zhu *et al.*, 2010). For example, crop model is an one-dimensional model and its simulation values are 'point' scale, while remote sensing observation are two-dimensional data and its observed values are 'plane' scale (or 'pixel' scale). So there is no match on spatial scale between process model simulation and remote sensing observations. At the same time, phenology information, taken as a characterization of crop growth process, is an important temporal prior knowledge to be used for estimating crop LAI. If phenology information is introduced into the crop model simulation to optimize the model sensitive parameters, accuracy of real-time simulation can be effectively improved.

In order to make full use of multiple remote sensing information and fully consider the scale effect of assimilating remote sensing data into crop model, an assimilation scheme based on temporal and spatial knowledge is proposed in this paper. The assimilation scheme is implemented in two directions (Fig. 1). One is the temporal direction, in which the parameters of crop model are calibrated, the other one is the spatial direction, in which the spatial information extracted from different resolution data is taken into account during assimilation.

Based on the parameter 'LAI', the crop growth model World Food Studies (WOFOST) (Boogaard *et al.*, 1998) and the reflectance model A two-layer Canopy Reflectance Model (ACRM) (Kuusk, 2001) are combined in this assimilation scheme by using the EnKF (Evensen, 1994) and the Shuffled Complex Evolution method developed at the University of Arizona (SCE-UA) method (Duan *et al.*, 1993). The MOD09A1 is used as observation data, then the spatial knowledge is extracted on the basis of its own multi-scale feature, and the spatial knowledge will be introduced into the assimilation process to adjust the mismatch between point data from WOFOST model simulation and pixel data from ACRM model inversion. The MOD09Q1 is used for calculating the Normalized Difference Vegetation Index (NDVI) data, which will be applied to extracting the phenology information, and then the phenology information is used as prior temporal knowledge to establish a cost function



**Fig. 1** Flowchart of assimilating MODIS data into WOFOST model based on prior knowledge. MODIS is Moderate Resolution Imaging Spectroradiometer; ACRM is a two-layer Canopy Reflectance Model; LAI is Leaf Area Index; EnKF is Ensemble Kalman Filter; WOFOST is crop growth model, World Food Studies; SCE-UA is Shuffled Complex Evolution method developed at University of Arizona; AG is a symmetric Gaussian function; NDVI is Normalized Difference Vegetation Index

with the fertility parameters of WOFOST model. Based on the cost function, the sensitive parameters will be calibrated by using SCE-UA optimization algorithm. At last, calibrated WOFOST model is used as forecast operator and ACRM inversion results is used as observation operator. Meanwhile, the EnKF algorithm embedded with spatial knowledge is used to update the crop LAI values.

## 2 Data and Methods

### 2.1 Study area

The Heihe River Basin is the second largest inland river basin in Northwest China, with a drainage area about 128 700 km<sup>2</sup>. Affected by the high-latitude westerlies circulation and the polar cold-air mass, the average annual precipitation is scarce, about 108 mm, the annual average temperature is of 6°C and the elevation of the Heihe River Basin varies between 1000–2000 m. The study area is mainly concentrated in the Yingke Oasis in the middle reaches of the middle Heihe River, with an area about 200 km<sup>2</sup>. The area, located at 8 km south of Zhangye City in Gansu Province, is a typical farmland ecosystem region, and the main crops are maize and spring wheat. There are two automatic meteorological stations established for long-period observation, recording air temperature, wind speed, relative humidity, precipitation and other atmospheric forcing data. From April to July in 2008, the Heihe remote sensing experiment measured part vegetation leaves, canopy and soil

parameters and so on. The locations of field data are shown in Fig. 2. The assimilation approach was tested at point 'p' and the crop type is maize.

### 2.2 Data and processing

MODIS products of MOD09A1 and MOD09Q1 were used as the remote sensing data in this paper (Table 1). MOD09A1 contains the first seven bands which are more sensitive to the land vegetation and its spatial resolution is 500 m. NDVI time series data is calculated by MOD09Q1. MOD09Q1 has a characteristic of relatively high time resolution and lower impact of cloud and its spatial resolution is 250 m. The ground observation data were obtained from the 'Watershed Airborne Telemetry Experiment Research of Chinese Academic of Science Action Plan for West Development Program'. Among them, The LAI collection method is as follows. Lengths and widths of sample leaves were measured manually and by LAI-2000 and LAI-3000 instruments. Correction factors for each crop were calculated by comparing the manual measured value and that from the instrument (after conversion). Crop LAIs were obtained through multiplication of crop leaf length and width by the correction factors. The sample size in Yingke Oasis is either 180 m × 180 m or 240 m × 240 m and the resolution of MODIS is low, so during the results analysis, the validation data are set by the average of field point data within a MODIS pixel (250 m). Two automatic meteorological stations (Yingke and Huazhaizi) are established for measuring and recording temperature, precipitation and

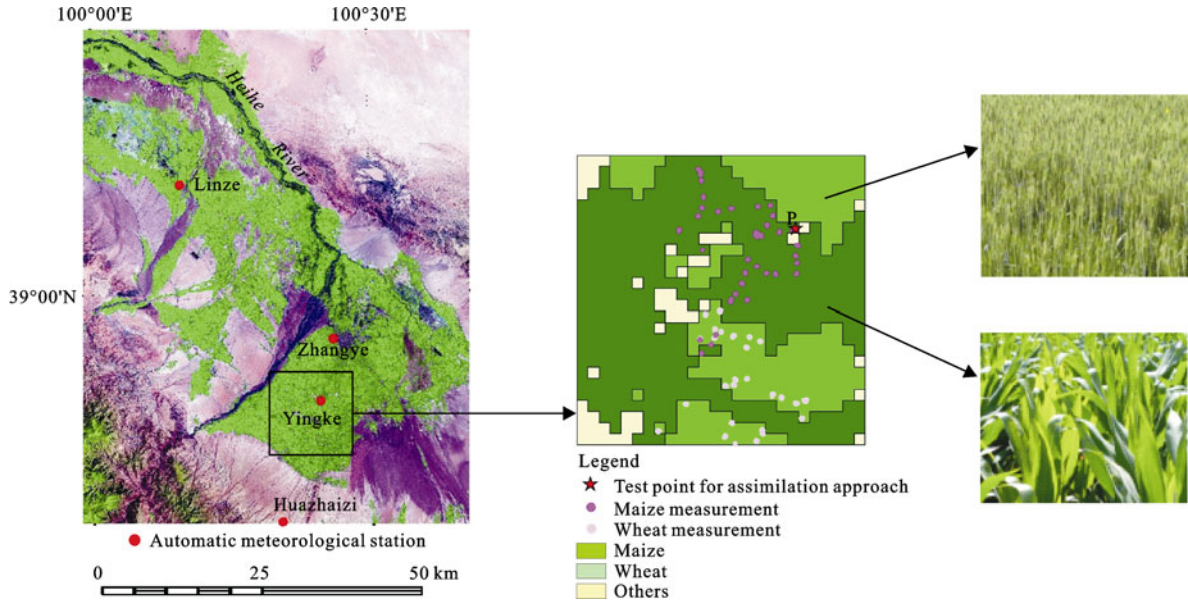


Fig. 2 Sketch map of Heihe River Basin, classification map of measurement location, and crop types

Table 1 Overview of remote sensed data

Product	DOY	Spatial resolution (m)
MOD09A1	129, 137, 153, 161, 185, 201, 225, 241	500
MOD09Q1	1–366	250

Note: DOY is day of year

other meteorological data. All the data are resampled into 250 m, corresponding to WOFOST model simulation scale.

## 2.3 Methods

### 2.3.1 Crop growth model

WOFOST model is a tool for quantitative analysis of the crop growth and its time step is 'day'. In this model, the radiation, temperature and crop characteristic parameters are used for simulating crop dry matter accumulation status from emergence to maturity (Xie et al., 2006). The model uses crop organs growth rate such as roots, stems, leaves and ears and their appearance order to identify the crop growth period, the whole growth period can be divided into three developmental stages, planting-emergence, emergence-flowering and flowering-mature. In WOFOST model, the dry matter is accumulated based on canopy CO<sub>2</sub> assimilation rate, which is calculated by the incident radiation and crop leaf area. One part of carbohydrates is used to maintain respiration of crop living tissue, the other part is used

for forming roots, stems, leaves and other organs of crop. The part which is assigned to leaves will determine the growth of leaf area and light interception (Ba, 2005).

$$D_{rdj} = \frac{T_{edi}}{TSUM_{dj}} \quad (1)$$

$$\begin{aligned} \Delta W &= C_e(kA_d - R_{m,T}) \\ &= kC_e \left( \frac{D(A_{h,-1} - 1.6A_{h,0} + A_{h,1})}{3.6} \right) - C_e R_{m,T} \end{aligned} \quad (2)$$

$$A_h = \frac{LAI(A_{-1} - 1.6A_0 + A_1)}{3.6} \quad A_L = A_m(I - e^{-\epsilon L/A_m}) \quad (3)$$

where  $D_{rdj}$  is the development rate at  $dj$ th development stage;  $T_{edi}$  is the effective temperature at  $d$ th day of year;  $TSUM_{dj}$  is the effective temperature needed at  $dj$ th development stage;  $\Delta W$  is the dry matter growth rate;  $C_e$  is assimilate conversion factor;  $k$  is the conversion coefficient between CO<sub>2</sub> and CH<sub>2</sub>O;  $A_d$  is the daily gross assimilation rate;  $R_{m,T}$  is the maintenance respiration rate at actual temperature ( $T$ );  $D$  is the length of day;  $A_{h,-1}$ ,  $A_{h,0}$ , and  $A_{h,1}$  are instantaneous assimilation rate in three Gaussian integral points;  $A_{-1}$ ,  $A_0$ , and  $A_1$  are gross assimilation rate in three Gaussian integral point;  $A_h$  is instantaneous assimilation rate at time  $h$ ;  $A_L$  is the gross assimilation rate;  $A_m$  is the maximum gross assimilation rate;  $I$  is the adsorbed radiation by leaf layer  $L$ ;  $\epsilon$  is the initial light use efficiency of individual leaves;  $LAI$  is Leaf Area Index.

The model contains a lot of parameters about crop

genetic characteristics and crop varieties, such as development parameters (temperature needed for different developmental stages, photoperiod influence factor, *etc.*), growth parameters (maximum photosynthetic rate, specific leaf area, dry matter partition coefficient, *etc.*), soil parameters. Parts of those parameters can be set by some literature such as Boogaard *et al.* (1998), Wu *et al.* (2003), Wang Tao *et al.* (2010), and experimental data observed from the Heihe River Basin. However, due to too many parameters and limited field data, some parameters are difficult to be set. And different parameters have different effects on the simulation of LAI, for some very sensitive parameters, unreasonable values will cause the simulation result deviate, inaccurate, or even wrong. Therefore, the model parameters must be adjusted to respond to the combined effect of simulation environment and crop varieties. Uncertainty and Sensitivity Matrix (USM) (Li *et al.*, 1997) is introduced to evaluate the sensitivity of model parameters by using the value of the parameter increased or decreased by 15% while the other parameters values are fixed on their expectations to calculate difference of simulated maximum and minimum LAI values.

$$USM_{DOY, para} = \frac{|LAI_{DOY, para \times (1 + 15\%)} - LAI_{DOY, para \times (1 - 15\%)}|}{LAI_{DOY, para}} \quad (4)$$

where *DOY* is the day of year; *para* is WOFOST model parameters;  $LAI_{DOY, para \times (1 + 15\%)}$  means *LAI* value that calculated when the value of parameter *para* increased by 15% and other parameters are fixed on their expectations;  $LAI_{DOY, para \times (1 - 15\%)}$  means *LAI* value that calculated when the value of parameter *para* decreased by 15% and other parameters are fixed on their expectations;  $LAI_{DOY, para}$  means *LAI* value that calculated when all the parameters are fixed on their expectations.

The experiment result from test site shows that LAI is sensitive to the effective temperature  $TSUM_1$ ,  $TSUM_2$ , the specific leaf area SLATB, the initial dry weight TDWI, the maximum photosynthetic rate AMAXTB and the assimilates conversion efficiency CVL, all of those parameters will be optimized during the assimilation process.

### 2.3.2 Reflectance model

The turbid medium, Markov chain canopy reflectance

model (a two-layer canopy reflectance model) (Kuusk, 2001) incorporates Markov properties of stand geometry, making it applicable to plant canopies largely composed of vertical elements such as maize. The ACRM model contains the PROSPECT leaf optical model (Jacquemoud and Baret, 1990), simulating the reflectance and transmittance from 400 nm to 2500 nm of a leaf, through the relationship function of leaf structure parameters and biochemical parameters. The model has a high computational efficiency and can explain the specular reflection of a leaf surface. The calculation of multiple scattering in the model is the same as in the Scattering by Arbitrarily Inclined Leaves (SAIL) model (Verhoef, 1984). The ACRM model is a one-dimensional, bidirectional turbid medium radiative transfer model that has been modified to take into account the 'hot spot' and 'leaf mirror reflection' effects of plant canopy reflectance (Kuusk, 1985). Use of a turbid medium defines the canopy as a horizontally homogenous and semi-infinite layer, consisting of small vegetation elements that act as absorbing and scattering particles of given geometry and density. Giving canopy structure and environmental parameters, the model can calculate canopy reflectance for any sun altitude and observation direction. The Price model (Price, 1990) is incorporated in the ACRM to calculate soil reflectance. The soil reflectance spectra in the model are approximated as a function of four basis vectors, where the first two vectors explain 94.2% of spectral variability in soil reflectance.

In order to avoid the situation that remote sensing observation bands is too few to cause inversion error, the MOD09A1 product is chosen as the remote sensing data. The product contains the first seven bands of MODIS, which are more sensitive to the land vegetation, and its spatial resolution is 500 m.

### 2.3.3 Phenology information

The crop model is a one-dimensional simulation model by which the different crops growth situation is simulated. The crop model simulation result is 'point' scale. Yingke Oasis has a large area of planting area, however, for pixel of MOD09A1 data with a resolution of 500 m, there may exist maize and wheat simultaneously. Therefore, considering that the growth change situation of maize and wheat is different, in the framework of assimilating MODIS data into crop model based on tem-

poral-spatial knowledge, the MODIS product MOD09Q1 is used for optimizing crop model sensitive parameters.

In order to reduce the impact of noise on phenological information extraction, a Asymmetric Gaussian function (AG filtering) of Timesat 3.0 is used to smooth NDVI time series data. The AG fitting is a process from local fitting to overall fitting, which uses segmented Gaussian function to simulate the process of vegetation growth, and finally connect each Gaussian fitting curve to achieve the time series reconstruction (Per and Eklundhc, 2004). First, local fitting functions are taken to describe NDVI data between maxima and minima, then, the local functions are merged to a global function to follow the change of NDVI time series during crop growth.

The mutation of NDVI value is a significant change signal of vegetation photosynthesis, indicating the great change of surface vegetation growth status, which is the key point for extracting phenological parameters. In this study, an improved two-side Logistic model (Zhang et al., 2003) is chosen to simulate the seasonal variation curves ( $y(t)$ ) of crop NDVI, and the key phenological phases are confirmed by calculating the curvature change of the simulation curve.

$$y(t) = \frac{a_3}{1 + e^{a_1 + a_2 t}} + a_4 \quad (5)$$

$$dc = \frac{ds}{dt} = \frac{\sqrt{(1 + e^{a_1 + a_2 t})^4 + (a_3 a_4 e^{a_1 + a_2 t})^2}}{(1 + e^{a_1 + a_2 t})^2} \quad (6)$$

where  $y(t)$  is the NDVI value at time  $t$ ;  $a_1$ ,  $a_2$ ,  $a_3$  are the fitting coefficients;  $a_4$  is the initial NDVI background value;  $dc$  is the arc differential, namely curvature change situation;  $ds$  is the unit arc length of logistic curve. The phenological parameters extracted in this study include those at the emergence stage (during which the crop is beginning to grow), the anthesis stage (during which NDVI achieves maximum and sustains) and mature stage (during which NDVI beginning to attenuate).

### 2.3.4 Parameters optimization based on SCE-UA

Shuffled Complex Evolution method developed at the University of Arizona (SCE-UA) (Duan et al., 1993) is a global optimization algorithm, proposed by Dr. Duan in 1992, and can solve the parameters optimization problem of nonlinear, multi-extreme, no specific function expression, interval constraint, and so on. SCE-UA has

the advantages of deterministic search, random search and biological competition evolution and introduces the population concepts. Compared with genetic algorithm and other optimization algorithms, SCE-UA method shows better on convergence speed and computational efficiency, and has better stability (Song et al., 2009). The specific process and application of SCE-UA algorithm please refer to the findings of Duan et al. (1993; 1994).

SCE-UA in this paper is used for optimizing the sensitive parameters of WOFOST model (TSUM<sub>1</sub>, TSUM<sub>2</sub>, SLATB, TDWI, AMAXTB and CVL). Based on prior knowledge (phenological information), the cost function ( $F$ ) is established as follows:

$$F = \sum_{DVS=0}^n [(DAY_{DVS}^{cal} - DAY_{DVS}^{pri})^2 + (LAI_{DVS}^{cal} - LAI_{DVS}^{pri})^2] \quad (7)$$

where  $DVS$  is model growth stage code;  $DAY_{DVS}^{cal}$  is the  $DOY$  calculated by WOFOST model;  $LAI_{DVS}^{cal}$  is the corresponding  $LAI$  at growth stage, which are calculated by crop model;  $DAY_{DVS}^{pri}$  and  $LAI_{DVS}^{pri}$  are the prior  $DOY$  and the corresponding prior  $LAI$ . Limited by the field data, the prior  $LAI$  data are taken from the spectral library (<http://spl.bnu.edu.cn/index.asp>) statistical values during crop emergence, anthesis and mature stages.

### 2.3.5 Spatial knowledge

In this paper, spatial knowledge is defined as surface spatial heterogeneity information, describing the scale bias of different-resolution remote sensing data, expressed by the parameter 'err'.

WOFOST is a one-dimensional model and its simulation results are 'point scale', however, the remote sensing inversion results are 'pixel scale', being considered as 'plane scale' in this paper. Mismatch of 'point scale' and 'plane scale' affects the remote sensed data assimilated into WOFOST model. To improve the assimilation accuracy, err is participated into the assimilation test to reduce the influence of different spatial scales.

Based on higher-resolution image,  $err$  defined as the bias for the  $LAI$  retrieved from up-scaling images, is shown as follows.

$$err = LAI_p - LAI_{mean} \quad (8)$$

where  $LAI_p$  is the  $LAI$  retrieved from lower-resolution images;  $LAI_{mean}$  is the average of  $LAI$  retrieved from higher-resolution images (which are taken as relative truth); and  $err$  is the difference of  $LAI_p$  and  $LAI_{mean}$ .

The purpose here is to obtain scale information of different-resolution remote sensing data, so the semi-empirical conversion formula between  $LAI$  and  $NDVI$  is established as follows (Baret and Guyot, 1991; Garrigues *et al.*, 2006).

$$LAI = f(NDVI) = \frac{-1}{k_{NDVI}} \log\left(\frac{NDVI - NDVI_{\infty}}{NDVI_s - NDVI_{\infty}}\right) \quad (9)$$

$$= b_1 \times \log(b_2 \times NDVI + b_3)$$

where the extinction coefficient  $k_{NDVI}$  and  $NDVI_{\infty}$  are calculated by the SAIL model, and  $k_{NDVI}$  is close to 0.5 (Zhang *et al.*, 1999);  $NDVI_{\infty}$  is often close to the pure vegetation  $NDVI$  when the leaf angle distribution is spherical;  $NDVI_s$  is the  $NDVI$  of bare soil;  $b_1$ ,  $b_2$ , and  $b_3$  are constants. The equation will be applied to higher-resolution remote sensing images.

Ignoring the non-linear formula of  $NDVI$ ,

$$LAI_p \approx f\left(\frac{1}{n} \sum_{i=1}^n NDVI_i\right) = f(NDVI_{mean}) \quad (10)$$

$$LAI_{mean} = \frac{1}{n} \sum_{i=1}^n f(NDVI_i) \quad (11)$$

where  $NDVI_i$  is the  $NDVI$  value of the  $i$ th pixel;  $n$  is the total number of pixel.

Then, the Taylor Mean Value Theorem is applied to designing the polynomial expansion of ' $LAI_{mean}$ ' around ( $NDVI_i = NDVI_{mean}$ ).

$$LAI_{mean} = \frac{1}{n} \sum_{i=1}^n \left[ f(NDVI_{mean}) + f'(NDVI_{mean}) \right. \\ \left. (NDVI_i - NDVI_{mean}) + \frac{f''(NDVI_{mean})(NDVI_i - NDVI_{mean})^2}{2} + R \right] \\ \approx f(NDVI_{mean}) + \frac{f''(NDVI_{mean})}{2} \times \\ \frac{1}{n} \sum_{i=1}^n (NDVI_i - NDVI_{mean})^2 \quad (12)$$

Finally,  $err$  equation is shown as below, with more details available in Zhu *et al.* (2010; 2012).

$$err = LAI_p - LAI_{mean} \\ = f(NDVI_{mean}) - \frac{1}{n} \sum_{i=1}^n f(NDVI_i) \\ \approx -\frac{1}{2} \sigma_{NDVI} \times f''(NDVI_{mean}) \\ = -\frac{1}{2} \sigma_{NDVI} \times (b_1 \times \log(b_2 \times NDVI + b_3))'' \Big|_{NDVI = NDVI_{mean}} \\ = \frac{\log(e)}{2} \times \frac{b_1 b_2^2 \times \sigma_{NDVI}}{(b_2 \times NDVI_{mean} + b_3)^2} \quad (13)$$

where  $NDVI_{mean}$ ,  $\sigma_{NDVI}$  are the mean value and variance of  $NDVI$  in higher-resolution image, respectively.

### 2.3.6 EnKF assimilation process based on spatial knowledge

Ensemble Kalman Filter, combining ensemble prediction and Kalman Filter (KF), is a four-dimensional assimilation method by using Monte Carlo short-term ensemble forecasting methods to estimate the prediction error covariance (Evensen, 1994). The basic idea of EnKF is that the covariance of state variables and observed variables is calculated by the results of ensemble forecast, then, analysis ensemble is updated by the covariance and observational data, making the prediction continue to move forward. The EnKF algorithm, overcoming the shortages of KF algorithm and not limiting to linear model and observation operator, has been widely used to assimilate remote sensing data into land surface models.

In the study, observation data are MOD09A1 with a resolution of 500 m, while the MOD09Q1 is used during the crop model optimization process and its resolution is 250 m. Therefore, the spatial knowledge extracted by Taylor series expansion is introduced into the EnKF assimilation algorithm. The operation process is as follows.

(1)  $LAI_{t,j}$  is the WOFOST model state variable of sub-pixel  $j$  at time  $t$ ;  $j$  (1, 2, ...,  $l$ ) is the sub-pixel of MOD09A1, that is the pixel of MOD09Q1. Monte Carlo method is used to generate ensemble members  $i$  at time  $t$ , and the number of ensemble members is  $n$  ( $n = 50$  in this paper).

$$LAI_{i,t,j} = (LAI_{1,t,1}, LAI_{2,t,2}, \dots, LAI_{n,t,l}) \quad (14)$$

Ignoring the flag of  $j$ , then, the ensemble members are:

$$LAI_{i,t} = (LAI_{1,t}, LAI_{2,t}, \dots, LAI_{ln,t}) \quad (15)$$

where  $LAI_{i,t,j}$  is the ensemble members at time  $t$  of sub-pixel  $j$ ;  $j$  is sub pixel of pixel  $J$ ;  $LAI_{i,t}$  is new ensemble members at time  $t$  of pixel  $J$ ;  $ln$  is the new number of ensemble members.

(2) Model moves forward and continues to forecast, predictive value  $LAI_{i,t+1}^f$  of time  $t+1$  is given as:

$$LAI_{i,t+1}^f = WOFOST(LAI_{i,t}^a) + \omega_i \quad (16)$$

where  $LAI_{i,t}^a$  is the analytical value of the random state variable  $i$  at time  $t$ ;  $\omega_i \in N(0, Q)$ ;  $Q$  is model error.

(3) Correcting the space heterogeneity effect and giving observation noise to generate observation ensemble, the sample size is  $m$  ( $m = ln$ ).

$$LAI^{e,c} = LAI^c + err \quad (17)$$

$$\begin{aligned} LAI_{i,t+1}^{e,c} &= LAI_{i,t+1}^{e,c} + v_i \\ &= (LAI_{1,t+1}^{e,c}, LAI_{2,t+1}^{e,c}, \dots, LAI_{m,t+1}^{e,c}) \end{aligned} \quad (18)$$

where  $LAI^c$  is the ACRM inversed value and corrected by Equation (8);  $LAI^{e,c}$  is the result corrected by  $err$ ,  $LAI_{i,t+1}^{e,c}$  are the observation data;  $v_i \in N(0, R)$ ,  $R$  is the observation error.

(4) Calculating Kalman gain matrix  $K$  at time  $t+1$ :

$$K_{t+1} = P_{t+1}^f H^T (HP_{t+1}^f H^T + R_{t+1})^{-1} \quad (19)$$

$$P_{t+1}^f = \frac{1}{m-1} LAI_{t+1}^{i,f} [LAI_{t+1}^{i,f}]^T \quad (20)$$

$$\begin{aligned} LAI_{t+1}^{i,f} &= \{LAI_{1,t+1}^{i,f}, LAI_{2,t+1}^{i,f}, \dots, LAI_{m,t+1}^{i,f}\} \\ &= \{(LAI_{1,t+1}^{i,f} - \overline{LAI_{t+1}^{i,f}}), \\ &\quad (LAI_{2,t+1}^{i,f} - \overline{LAI_{t+1}^{i,f}}), \dots, \\ &\quad (LAI_{m,t+1}^{i,f} - \overline{LAI_{t+1}^{i,f}})\} \end{aligned} \quad (21)$$

where  $\overline{LAI_{t+1}^{i,f}}$ ,  $LAI_{t+1}^{i,f}$  are the mean and deviation matrix of the predicted value of the state variables at time  $t+1$ ;  $H$  (\*) is the observation operator;  $P^f$  is the covariance matrix of forecast error at according time.

(5) Calculating the average value of the state variables at time  $t+1$  and the background error covariance matrix  $P_{t+1,j}^a$  of sub pixel  $j$ :

$$LAI_{i,t+1}^a = LAI_{i,t+1}^f + K_{t+1} (LAI_{i,t+1}^{e,c} - H(LAI_{i,t+1}^f)) \quad (22)$$

$$\overline{LAI_{t+1}^a} = \frac{1}{m} \sum_{i=1}^m LAI_{i,t+1}^a \quad (23)$$

$$P_{t+1,j}^a = \frac{1}{n-1} LAI_{i,t+1,j}^a [LAI_{i,t+1,j}^a]^T \quad (24)$$

$$\begin{aligned} LAI_{t+1,j}^a &= \{LAI_{1,t+1,j}^a, LAI_{2,t+1,j}^a, \dots, LAI_{n,t+1,j}^a\} \\ &= \{(LAI_{1,t+1,j}^a - \overline{LAI_{t+1,j}^a}), \\ &\quad (LAI_{2,t+1,j}^a - \overline{LAI_{t+1,j}^a}), \dots, \\ &\quad (LAI_{n,t+1,j}^a - \overline{LAI_{t+1,j}^a})\} \end{aligned} \quad (25)$$

$$\overline{LAI_{t+1,j}^a} = \frac{1}{n} \sum_{i=1}^n LAI_{i,t+1,j}^a \quad (26)$$

where  $\overline{LAI_{t+1}^a}$  is the average value of  $LAI_{i,t+1}^a$  ( $i = 1, 2, \dots, m$ ), which is considered as the optimal estimation by EnKF (Burgers and Concha, 1998).

(6)  $t = t+1$ , return to step (2), Loop calculating until the end.

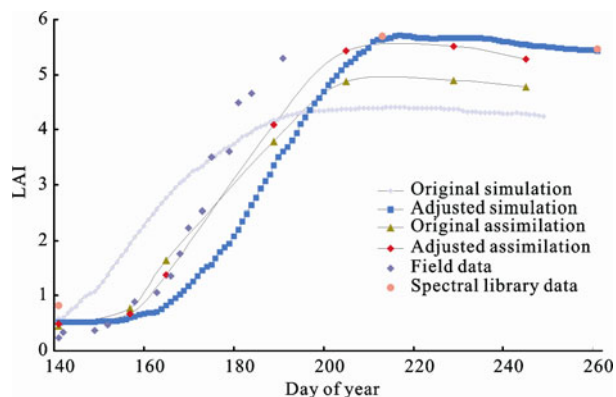
### 3 Results and Analyses

The assimilation method mentioned above is firstly applied in point assimilation test of pixel P (38.86°N, 100.43°E) which is shown in Fig. 2. The results calculated by original simulation, adjusted simulation, original assimilation and assimilation based on temporal-spatial knowledge are compared for analyzing the methods proposed in this paper. In addition, the assimilation method based on prior knowledge is applied in whole study area, and then the estimation results up-scaling for comparative analysis with MODIS LAI product, further evaluating the effectiveness of the method.

MOD09A1 on 129, 137, 153, 161, 185, 201, 225 and 241 day of year (DOY) in 2008 were used in the assimilation experiment, basically covering the growth period of maize and wheat (maize is sowed in April or May and matured in September or October; spring wheat is sowed in March or April and matured in July or August). The observation dates of different regions are not the same due to that the MOD09 product data is 8-day synthesis (Sun et al., 2009), so the middle date is chosen as the image observation date in this study. Due



to lacking of observation data during crop growth later stage, the ground truth data including spectral library data (<http://spl.bnu.edu.cn/index.asp>) are used together for analyzing the assimilation results. The verification results are shown in Fig. 3.



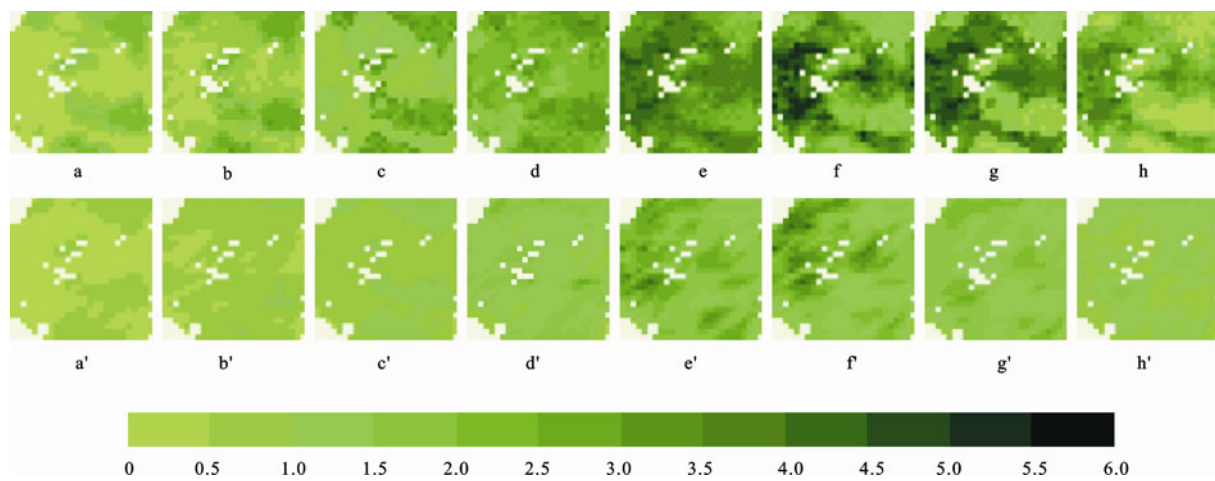
**Fig. 3** Assimilation results compared with field data, spectral library data and WOFOST model simulation data

In Fig. 3, assimilation results refer to the result of LAI which is calculated by assimilating remote sensing data into WOFOST model, and simulation results refer to the result of LAI simulated by WOFOST model. As seen from the Fig. 3, although the crop emergence date and initial LAI are reasonable, due to the absence of remote sensing observations, the simulation curve started to deviate from the true growth situation and was faster into stable period. Meanwhile, without optimization of model sensitive parameters, the lower temperature and specific leaf area coefficient make the simulation of growing period to be shorter and the maximum LAI to be lower than normal. However, the assimilation results are based on remote sensing observations, whether it is single assimilation or temporal-spatial knowledge-based assimilation, the calculated LAI value is closer to the ground truth value. The root mean square error (RMSE) and average error (AE) of LAI estimated by knowledge-based assimilation method are 0.3795 and 0.0265, while those simulated by model are 0.9185 and 0.3563. It indicates that model parameters optimized by phenology information make the WOFOST model localizable and simulation results are closer to the growth trend of study area. At the same time, by integrating spatial knowledge into the assimilation process, the LAI value is corrected to be more reasonable and closer to the ground truth data and spectral library data

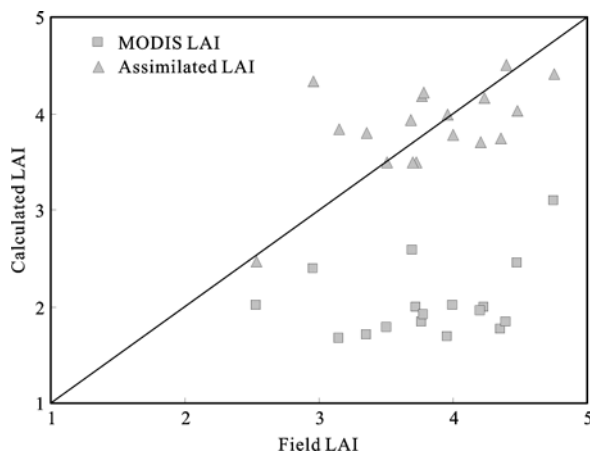
during the early growth stage. During the early growth stage (DOY: 161–201), the crop grows rapidly and makes the surface vegetation cover change dramatically leading to high spatial heterogeneity.

The LAI images calculated by assimilation method based on the whole study area are up-scaling to 1 km and compared with MODIS LAI products (Fig. 4). It can be seen from comparison in Fig. 4, spring wheat and maize LAI changes more in line with the seasonal changes in crop growing period. Around early June (Fig. 4c, 4d), the spring wheat LAI should generally reach peak value. The spring wheat LAI assimilation value is between 3.5 and 4.5, closing to the field data value of 3.81 (the field data collection time is from 164 to 166); while MODIS LAI value is generally lower than 2.0. More detail comparative analysis between assimilation LAI (Fig. 4d) and MODIS LAI (Fig. 4d') based on the wheat measurement data on 164–166 is shown in Fig. 5. The measurement point distribution is shown in Fig. 2. The statistical result points out that RMSE of verification results was reduced from 1.9257 to 0.4409, and the average value of assimilation LAI is 3.87, which is 50.26% higher than that of MODIS LAI and more conform to ground situation.

Analysis from spectral library data and field data, during early to mid-May (Fig. 4a), the wheat LAI should be greater than 1.0 but MODIS LAI value did not reached up to it, while assimilation results vary between 1.0 and 3.0. It means that the assimilation results are more in line with the crop growth characteristics in the study area. MODIS LAI value is generally low, especially in maize LAI. During maize flowering period, the LAI reaches the maximum, and then it decreases slowly and keeps at a high level until the end of mature period. Statistics from the ground truth data, maize LAI at this period should be about 5.0 (the measure time is 184–191, lacking of late measured data), and the spectral library prior statistics is between 5.46 and 5.70. Seen from the assimilation results, LAI value in this period is between 5.0–6.0 (Fig. 4f, 4g, 4h), similar to that from the ground measurement and spectral library, reasonably indicating the LAI changes in maize growing period. While maximum LAIs of MODIS product mainly concentrate in the 2.5–4.0 and then descend to 1.0–2.0, deviating from a reasonable trend. As a whole, LAI calculated by knowledge-based assimilation method is more reasonable than MODIS LAI and more consistent with the field data.



**Fig. 4** Assimilation results (a–h) based on temporal-spatial knowledge comparing with MODIS LAI product (a'–h') on day of year 129, 137, 153, 161, 185, 201, 225, 241, respectively



**Fig. 5** Comparison between assimilated LAI and MODIS LAI on day of year from 164 to 166

## 4 Discussion

Traditional LAI inversion researches tend to focus on algorithms. In consciousness to the importance of prior knowledge and the synergism advantage of multiple remote sensing data, scholars paid less attention to excavate and utilize prior knowledge to improve the accuracy estimation of LAI.

In this study, based on the assimilation technology, we combined the time-series remote sensing observations and crop growth model, so that the extracted LAI not only considers the remote sensing information but also depends on the simulation information from crop growth model. The research tried to excavate and use the prior knowledge in two aspects. During the crop model assimilating MODIS data, the process would be

affected by scale restriction between 'point-plane'. We use the multi-resolution characteristic of MODIS data to extract the spatial information between different resolution data as prior knowledge which expanding the connotation of prior knowledge. As seen from Equation (9), the inversion result of low resolution data was adjusted by the NDVI variance of higher resolution data and second order of relationship function. Through such processing, low resolution data inversion result used as observation operator reduced the surface heterogeneity influence and improve the accuracy of observation information. During the assimilation process, the higher resolution of 'point' model simulating scale was combined with the low resolution of 'plane' inversion results, realizing the union of information between 'point' and 'plane'. During the process of crop model optimization, the phenological information extracted from MOD09Q1 was used as temporal information, participating in the optimization by cost function Equation (7), improving the simulation accuracy.

The experimental results showed that, in point test, the assimilation results based on temporal and spatial knowledge were more reasonable than direct simulation results. The RMSE was reduced by 58.7% and average error reduced by 92.6%. It indicated that the model simulation curve would show larger bias by unreasonable parameters. At the same time, the crop model parameters are selectively optimized by SCE-UA, improving 'localization' efficiency of model. After the adjustment, the simulated maximum LAI and crop growing circle were more suitable for the study area. In addi-

tion, the estimated LAI by assimilating remote sensing data into adjusted crop model was more accuracy and the RMSE reduced from 0.6770 to 0.3795. It showed that the assimilation method based on prior knowledge could obviously improve the estimation accuracy of crop LAI. Further comparing with the MODIS LAI, it can also be found that the changing curve of assimilated LAI is more conformed to the crop growth condition. This indicated that, in addition to the remote sensed observation information, the assimilated LAI also contains the simulation information of crop growth model.

## 5 Conclusions

In this paper, a methodology of assimilating MODIS product with WOFOST crop growth model based on temporal and spatial knowledge was proposed to estimate LAI in two directions. The assimilation method is tested in Yingke Oasis in Gansu Province, a typical ecological area.

The experimental results showed that: 1) during the parameters optimization process, the cost function used for SCE-UA was established based on phenology information which improved the accuracy of the crop model simulation. 2) Spatial knowledge involved in the assimilation process, improved the match between the simulation model with coarse resolution remote sensing data, and reduced the effect of spatial heterogeneity in the assimilation process. 3) More than one direction of the remote sensing observations was selected to participate in assimilation and further improved estimation accuracy of the crop LAI.

In this study, spatial knowledge was involved in the assimilation process, but we only focus on the spatial heterogeneity within the coarse resolution pixel, and in the future we will consider the spatial information between pixel and its adjacent pixels. In addition, the research program is an assimilation process based on the combination of 'point simulation' and 'plane inversion'. By considering of the temporal resolution, we use the MOD09 to calculate the phenology parameters, so that the spatial resolution of remote sensing data involved in the assimilation process is a little larger. In future research, the 16-day composite product will be taken into account, so that we can use higher resolution remote sensing data to participate in assimilation, such as Hyperion, TM/ETM+, and ASTER.

## References

- Bateer Bake, 2005. *Studies On Modeling Crop Production in Fen River Irrigation District and Water Balance Model of the Watershed*. Beijing: China Agricultural University. (in Chinese)
- Baret F, Guyot G, 1991. Potentials and limits of vegetation indices for LAI and FAPAR assessment. *Remote Sensing of Environment*, 35(2–3): 161–173. doi: 10.1016/0034-4257(91)90009-U
- Boogaard H L, van Diepen C A, Rotter R P *et al.*, 1998. *User's Guide for the WOFOST 7. 1 Crop Growth Simulation Model and WOFOST Control Center*. DLO Wageningen: Winand Staring Centre, 1–40.
- Burgers R, Concha F, 1998. Mathematical model and numerical simulation of the setting of flocculated suspensions. *International Journal of Multiphase Flow*, 24(6): 1005–1023. doi: 10.1016/S0301-9322(98)00026-3
- Doraiswamy P C, Moulin S, Cook P W *et al.*, 2003. Crop yield assessment from remote sensing. *Photogrammetric Engineering and Remote Sensing*, 69(6): 665–674. doi: 10.1080/01431160410001698870
- Dorigo W A, Zurita-Milla R, de Wit A J W *et al.*, 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International Journal of Applied Earth Observation and Geoinformation*, 9(2): 165–193. doi: 10.1016/j.jag.2006.05.003
- Duan Qingyun, Gupta V K, Sorooshian S, 1993. A shuffled complex evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and Applications*, 76(3): 501–521. doi: 10.1007/bf00939380
- Duan Qingyun, Sorooshian S, Gupta V K, 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. *Journal of Hydrology*, 158(3–4): 265–284. doi: 10.1016/0022-1694(94)90057-4
- Evensen G, 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(5): 10143–10162. doi: 10.1029/94jc00572
- Fang Hongliang, Liang Shunlin, Gerrit H, 2010. Integration of MODIS LAI and vegetation index products with the CSM-CERES-Maize model for corn yield estimation. *International Journal of Remote Sensing*, 32(4): 1039–1065. doi: 10.1080/01431160903505310
- Garrigues S, Allard B D, Baret F, 2006. Influence of landscape spatial heterogeneity on the non-linear estimation of leaf area index from moderate spatial resolution remote sensing data. *Remote Sensing of Environment*, 105(4): 286–298. doi: 10.1016/j.rse.2006.07.013
- Hazarika M K, Yasuoka Y, Ito A *et al.*, 2005. Estimation of net primary productivity by integrating remote sensing data with an ecosystem model. *Remote Sensing of Environment*, 94(3): 298–310. doi: 10.1016/j.rse.2004.10.004
- Jacquemoud S, Baret F, 1990. PROSPECT: A model of leaf optical properties spectra. *Remote Sensing of Environment*, 34(2): 75–91. doi: 10.1016/0034-4257(90)90100-Z

- Kuusik A, 1985. The hot spot effect of a uniform vegetative cover. *Soviet Journal of Remote Sensing*, 3(4): 645–658.
- Kuusik A, 2001. A two-layer canopy reflectance model. *Journal of Quantitative Spectroscopy & Radiative Transfer*, 71(1): 1–9. doi: 10.1016/S0022-4073(01)00007-3
- Li Xiaowen, Gao Feng, Wang Jingdi et al., 1997. Uncertainty and sensitivity matrix of parameters in inversion of physical BRDF model. *Journal of Remote Sensing*, 1(1): 5–14. (in Chinese)
- Per J, Eklundhc L, 2004. TIMESAT—A program for analyzing time-series of satellite sensor data. *Computers and Geosciences*, 30(8): 833–845. doi: 10.1016/j.cageo.2004.05.006
- Price J C, 1990. On the information content of soil reflectance spectra. *Remote Sensing of Environment*, 33(2): 113–121. doi: 10.1016/0034-4257(90)90037-M
- Song Xingyuan, Shu Quanying, Wang Haibo et al., 2009. Comparison and application of SCE-UA, genetic algorithm and simplex method. *Engineering Journal of Wuhan University*, 4(10): 6–15. (in Chinese)
- Sun Huasheng, Huang Jingfeng, Peng Dialing, 2009. Detecting major growth stages of paddy rice using MODIS data. *Journal of Remote Sensing*, 13(6): 1130–1137. (in Chinese)
- Tian Y H, Woodcock C E, Wang Y J, 2002. Multi-scale analysis and validation of the MODIS LAI product, I. Uncertainty assessment. *Remote Sensing of Environment*, 83(3): 414–430. doi: 10.1016/S0034-4257(02)00047-0
- Verhoef W, 1984. Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. *Remote Sensing of Environment*, 16(2): 125–141. doi: 10.1016/0034-4257(84)90057-9
- Wang Dongwei, Wang Jindi, Liang Shunlin, 2010. Retrieving crop leaf area index by assimilation of MODIS data into crop growth model. *Science China Earth Science*, 53(5): 721–730. (in Chinese)
- Wang Tao, Lu Change, Yu Bohua, 2010. Assessing the potential productivity of winter wheat using WOFOST in the Beijing-Tianjin-Hebei Region. *Journal of Natural Resources*, 25(3): 475–487. (in Chinese)
- Wu Dingrong, Ou Yangzhu, Zhao Xiaomin et al., 2003. The applicability research of WOFOST model in North China Plain. *Acta Phytocologica Sinica*, 27(5): 594–602. (in Chinese)
- Xiao Zhiqiang, Liang Shunlin, Wang Jindi et al., 2011. Real-time retrieval of Leaf Area Index from MODIS time series data. *Remote Sensing of Environment*, 115(1): 97–106. doi: 10.1016/j.rse.2010.08.009
- Xie Wenxia, Wang Guanghuo, Zhang Qichun, 2006. Development of WOFOST (World Food Studies) and its Application. *Chinese Journal of Soil Science*, 37(1): 154–158. (in Chinese)
- Yannick C, Allard J.W, Duveiller G et al., 2011. Potential performances of remotely sensed LAI assimilation in WOFOST model based on an OSS Expeniment. *Agricultural and Forest Meteorology*, 151(12): 1843–1855. doi: 10.1016/j.agrformet.2011.08.002
- Zhang Renhua, Sun Xiaomin, Su Hongbo et al., 1999. Remote sensing and scale transferring of levity parameters on earth surface. *Remote Sensing for Land & Resources*, 3(2): 51–59. (in Chinese)
- Zhang X Y, Friedl M A, Schaaf C B et al., 2003. Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, 84: 471–475. doi: 10.1016/S0034-4257(02)00135-9
- Zhang Xiuying, Jiang Hong, Han Ying, 2010. Lamd data assimilation system and its application in global change research. *Remote Sensing Information*, (4): 135–143. (in Chinese)
- Zhu Xiaohua, Feng Xiaoming, Zhao Yingshi et al., 2010. Scale effect and error analysis of crop LAI inversion. *Journal of Remote Sensing*, 14(3): 586–599. (in Chinese)
- Zhu Xiaohua, Feng Xiaoming, Zhao Yingshi, 2012. Multi-scale MSDT inversion based on LAI spatial knowledge. *Science China Earth Science*, 55(8): 1297–1305.