Available online at www.sciencedirect.com
ScienceDirect
Remote Sensing of Environment 112 (2008) 613 – 622
Remote Sensing of Environment
www.elsevier.com/locate/rse

A Bayesian network algorithm for retrieving the characterization of land surface vegetation

Yonghua Qu [a,b,*], Jindi Wang [a,b], Huawei Wan [d], Xiaowen Li [a,b], Guoqing Zhou [c]

[a] School of Geography, Beijing Normal University, Beijing, China
[b] State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Beijing Normal University and the Institute of Remote Sensing Applications of Chinese Academy of Sciences, Beijing, China
[c] Department of Civil Engineering and Technology, Old Dominion University, Norfolk, VA 23529, United States
[d] Environmental Satellite Center Preparing Office, State Environmental Protection Administration, Beijing, China

Received 15 March 2006; received in revised form 23 March 2007; accepted 31 March 2007

Abstract

A hybrid inversion technique based on Bayesian network is proposed for estimating the biochemical and biophysical parameters of land surface vegetation from remotely sensed data. A Bayesian network is a unified knowledge-inferring process that can incorporate information derived from multiple sources including remote sensing and information derived from a priori knowledge. Using this inversion approach, content of chlorophyll *a* and chlorophyll *b* (Cab) and leaf area index (LAI) of winter wheat were estimated from data derived from simulations as well as field measurements. Estimations from the simulated data proved accurate, with root mean square errors (RMSEs) of 0.54 m^2/m^2 in LAI and 4.5 $\mu g/cm^2$ in Cab. In validating the estimates against field measurements, it was found that prior knowledge of target parameters improved the accuracy of estimates, in terms of RMSEs from 0.73 to 0.22 m^2/m^2 in LAI and 9.6 to 4.0 $\mu g/cm^2$ in Cab. Bayesian inference in this hybrid inversion scheme produces a posterior probability distribution, which can reveal such properties of the inferred results as updated information contained in the inversion result. Using entropy, the revision of posterior information about the parameters of interest was calculated. Including more data may allow more information to be retrieved about parameters in general. Exceptions were also observed where data from some viewing angles slightly reduced the information on the parameters of interest. It was also found that data from these viewing angles were less sensitive to the parameters. The method proposed here was also validated using LandSat ETM+ imagery provided by the BigFoot project. When used for mapping LAI with ETM+ imagery, the proposed method with an RMSE of 0.70 and a correlation of 0.67 produced a slightly better result than that from empirical regression.
© 2007 Elsevier Inc. All rights reserved.

Keywords: Bayesian network; Leaf area index; Chlorophyll concentration; Hybrid inversion; Information entropy

1. Introduction

Estimating such biophysical and biochemical parameters of land surface vegetation as leaf area index (LAI) and leaf chlorophyll content (Cab) is an important application of remote sensing (Koetz et al., 2005; Myneni et al., 1995; Verstraete et al., 1996). LAI, defined as half the total leaf surface area per unit area of horizontal surface, is an important structural variable of vegetation and a key variable in understanding several ecophysiological processes within the vegetation canopy (Gong et al., 2003; Tian et al., 2003). Cab, the sum of the contents of chlorophyll *a* and chlorophyll *b* per unit leaf area, is intimately associated with physiological functions of leaves (Gitelson & Merzlyak, 1997; Sims & Gamon, 2002). Both LAI and Cab are affected when vegetation is exposed to natural and anthropogenic stresses, and non-destructive determination of these parameters from a distance is a good method of studying leaf function, and plant physiological state and stress (Koetz et al., 2005).

LAI and Cab can be estimated either by empirical methods or by inverting a radiative transfer model (Baret & Guyot, 1991;

* Corresponding author. School of Geography, Beijing Normal University, Beijing, China. Tel.: +86 10 58809966; fax: +86 10 58805274.
E-mail address: qyh@bnu.edu.cn (Y. Qu).

0034-4257/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2007.03.031

Verstraete et al., 1996). The empirical method involves constructing empirical formulae that link spectral features (e.g. vegetation index) and parameters of the earth's surface using experimental data (Gong et al., 1992; Walthall et al., 2004). The empirical method is simple and efficient in estimating the parameters but has a few inherent disadvantages. Since the relationship is derived using data at a specific time and place, the empirical formulae are limited temporally and spatially. An alternative is to invert a physical model, a method that has a clear physical basis and the retrieved results can be explained using physical models (Myneni et al., 1995).

However, physical models are often complicated and non-linear; as a result, the inversion process is often ill-posed (Combal et al., 2003). Many mathematical techniques have been developed to handle such problems, such as the regularization method (Doicu et al., 2003, 2004; Fymat, 1979) and artificial neural network techniques (Fang & Liang, 2003). The problem of inversion can also be solved by the knowledge reasoning method rather than by mathematical optimization (Kimes et al., 1991). In applying knowledge reasoning, the ill-posed problem can be regarded as a result of insufficient information during the process of knowledge reasoning. Therefore, additional knowledge is used in the inversion process (Combal et al., 2003; Li et al., 2001).

Recent research has shown that combining the empirical method and physical model inversion into a new hybrid inversion scheme is a promising approach to estimating surface parameters (Fang & Liang, 2003; Liang, 2004). The scheme uses simulated data sets to fit an empirical formula and the fitted equation is then used for estimating land surface parameters. Although non-linear fitting methods such as artificial neural network (ANN) and projection pursuit regression (PPR) have been used in earlier hybrid inversion models, and have shown how inversion efficiency and accuracy can be improved, there is still scope for improvement by incorporating more information in the inversion process. In the earlier hybrid inversion schemes, parameters of the physical model alone were included; other parameters that may influence the parameters of interest were not incorporated into the process of estimation. For example, although it is known that LAI is affected by crop growth stages, current hybrid inversion methods lack a way to introduce such temporal information into the estimations.

This paper proposes an alternative hybrid inversion method, which uses a Bayesian network to map the simulated reflectance to its corresponding biophysical parameters. As a hierarchical probability model, a Bayesian network can be used not only as a non-parametric regression model but also to deduce information from multi-layer parameters (Marcot et al., 2001). Kalacska et al. (2005) used it for estimating LAI of a tropical dry forest from ETM+ data and obtained better results than those obtained from spectral vegetation indices or ANN. However, for the initial network estimates the authors used data from ground surveys combined with known forest structure, LAI, and satellite reflectance. The method, therefore, was not free of the disadvantages inherent in the empirical method. In our proposed Bayesian network approach, the initial network estimates use the data obtained by simulating a physical model. Therefore, the approach can be used for estimating the biophysical and biochemical parameters of a standing crop over a wider temporal and spatial range than is possible with a limited amount of ground measurements. In our approach, we also focus on incorporating ancillary information extracted from a spectral library, namely the Spectral Library on Typical Land Surface Objects in China (SLTLSOC) (Qu et al., 2003), to support the inversion, which differs from other non-parametric regression methods such as ANN and PPR.

Our study sought to develop a new hybrid inversion scheme supported by the SLTLSOC and was tested against data sets obtained from both simulations and field measurements. The second purpose was to study changes in the values of the parameters of interest after sequentially adding multi-angle observation data to the inversion process.

2. Methods

2.1. Radiative transfer model

A coupled radiative transfer model, PROSPECT+SAIL (PROSAIL), was used to simulate the reflectance of vegetation canopies. PROSPECT can simulate the reflectance and transmittance of leaves using their biochemical and biophysical parameters (Baret & Fourty, 1997; Jacquemoud & Baret, 1990). These parameters include Cab, leaf water content (Cw), dry matter content (Cm), and the leaf mesophyll structural parameter (N). SAIL (Scattering by Arbitrarily Inclined Leaves) is a physics-based radiative transfer model used for simulating the hemispheric reflectance spectra of canopies under different viewing directions (Verhoef, 1984). The version of SAIL used in this study was developed by Kuusk, which included the hot-spot effect in the original SAIL model (Kuusk, 1991). The SAIL model needs seven input parameters: LAI, average leaf angle (ALA), ratio of leaf length to canopy height (SL), leaf hemispheric reflectance (LR), leaf transmittance (LT), soil reflectance, and atmospheric visibility (VIS). LR and LT can be simulated by PROSPECT. The coupled PROSAIL model computes multi-spectral reflectance under different incident and observation directions. We used the simulated BRDF (Bidirectional Reflectance Distribution Function) data from wavelengths of 400–900 nm at intervals of 5 nm, with the illumination-viewing angle of 55°. Because Landsat ETM+ data and field measurements were to be used, the response functions of ETM+ on green (525–605 nm), red (630–690 nm), and near-infrared (775–900 nm) bands were used to resample the simulated reflectance from simulated narrow bands into broad ETM+ bands 2, 3, and 4 respectively.

To reduce the number of model parameters to be retrieved, some parameters can be fixed when simulating the canopy reflectance. PROSPECT computes leaf reflectance and transmittance based on the specific absorption coefficient (SAC), the main factor influencing leaf reflectance, of each component in every band (Jacquemoud & Baret, 1990). The effect of Cw and Cm is mainly in wavelengths longer than 1300 nm and of Cab, in approximately 400–800 nm. Since reflectance in wavelengths less than 900 nm is used to estimate chlorophyll content,

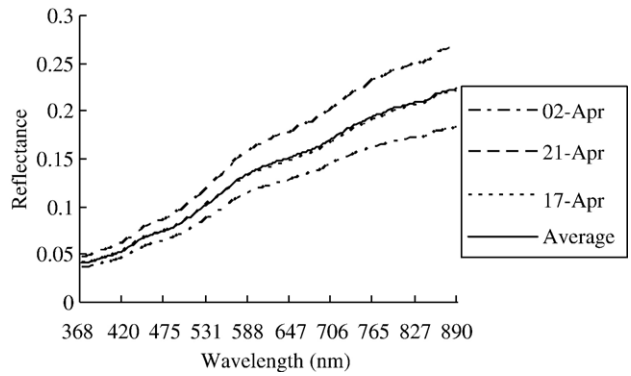

Fig. 1. Soil reflectance of PROSAIL model.

Cm and Cw can be fixed values. The structural parameter N in PROSPECT mainly influences leaf multi-scattering in the near-infrared band, so N is fixed when retrieving Cab. The canopy structural parameter SL in SAIL mainly affects the shape and size of hot spots (Jacquemoud et al., 2000). VIS is employed to calculate the diffuse part of the incoming radiation in the SAIL model. In simulations using PROSAIL, VIS can be a constant value. As a background, soil is assumed to be a Lambertian medium, and determining soil reflectance is a key problem in parameters retrieved through inverting a physical model. It is difficult to account for variations in soil background corresponding with the canopy reflectance (Fang & Liang, 2003). In this paper, soil reflectance was assigned a fixed value determined by averaging field measurements on different days (Fig. 1).

Thus, the model inputs were five fixed parameters and three free variables, namely LAI, Cab, and ALA. We simulated canopy reflectance from five observation angles to the solar principal plane: −55°, −25°, 0°, 25°, and 55°. The solar zenith and azimuth angles were fixed. These input parameters are shown in Table 1. A total of 1300 groups of BRDF were generated, simulating the measurements of three spectral bands from five observation angles.

2.2. Bayesian network model integrated with prior knowledge

The Bayesian theorem describes posterior probability distributions under specific conditions and has been widely applied in the field of remote sensing (Yager, 2006). When applying the Bayesian theorem to the estimation of land surface parameters, the parameters and observed data are regarded as random variables. The Bayesian network is a mathematical model combining graphics and probabilities to express mutual relationships between variables. It uses a directed acyclic graph to describe this relationship. Each node in the network represents a random variable, and the arc linking the nodes represents the relationship between variables. Fig. 2 shows a simplified Bayesian network.

Table 1
Input parameters for the PROSAIL model simulation

Input parameters	Unit	Value range	Step
LAI	m^2/m^2	0.5–5.5	0.2
ALA	Degree	35–75	10
Cab	$\mu g/cm^2$	14–60	5
Cw	cm	0.015	–
Cm	mg/cm^2	0.01	–
N	–	1.5	–
SL	–	0.25	–
VIS	km	20	–
Relative azimuth	Degree	0,180	–
Sun zenith (SZA)	Degree	55	–
View zenith (VZA)	Degree	55, 25, 0, −25, −55	–

In Fig. 2, the joint probability distribution (JPD) of random variables A, B, and C can be computed using

$$\begin{aligned} p(C,A,B) &= p(C,A)*p(B|C,A) \\ &= p(C)*p(A|C)*p(B|C,A). \end{aligned} \tag{1}$$

With the hypothesis of conditional independence in Bayesian network, i.e. for a given A, parameters B and C are independent (Murphy, 1998), we have

$$p(B|C,A) = p(B|A). \tag{2}$$

Eq. (1) can be rewritten as Eq. (3)

$$\begin{aligned} p(C,A,B) &= p(C,A)*p(B|C,A) \\ &= p(C)*p(A|C)*p(B|A). \end{aligned} \tag{3}$$

Based on the principle of retrieving parameters by Bayesian network, the posterior probability density distribution of A can be calculated using the observed data and their ancillary parameters. Eq. (4) can then be deduced as follows

$$p(A|B=b_i, C=c_k) = \frac{p(A|C=c_k)p(B=b_i|A)}{\sum_{\{a_j\}} p(A=a_j|C=c_k)p(B=b_i|A)} \tag{4}$$

where $p(A|C=c_k)$ represents the probability density distribution of the parameters to be derived after obtaining the ancillary information. The quantitative relationship between ancillary information and parameters to be retrieved (i.e. $p(A|C)$) can be obtained from SLTLSOC through a statistical method, a type of information referred to as prior knowledge. $p(B=b_i|A)$ is used to describe the probability density of the discrepancy between observed data and that obtained from simulation. The denominator has no relation with the parameters and serves mainly as a normalization factor. By extending the Bayesian theorem to the Bayesian network using a multi-factor deducing

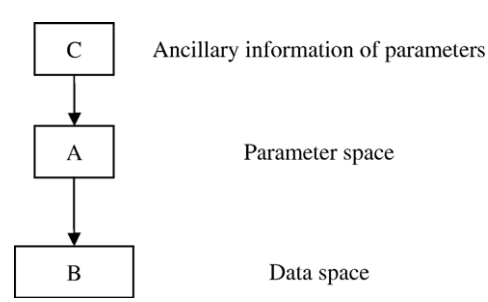

Fig. 2. Conceptual figure of a Bayesian network.

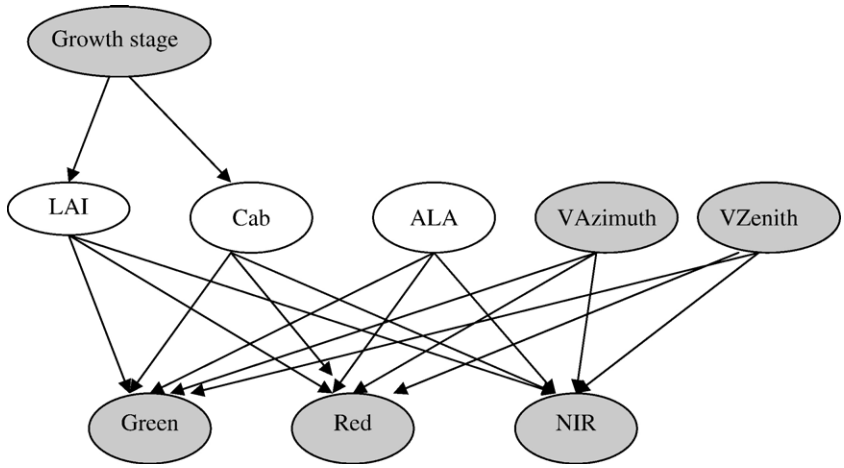

Fig. 3. Bayesian network model used for estimating LAI and Cab (shaded nodes represent the observable nodes and white nodes represent free variables).

method, prior knowledge about land surface parameters can be extracted from accumulated data comprising field observations. Such data sources can be thought of as objective information, reducing the subjective influence from researchers.

2.3. Estimating the parameters

Fig. 3 is a Bayesian network design based on simulated canopy spectral data and land surface parameters. The model in Fig. 3 omits the fixed parameters in PROSAIL, considering the influence only of LAI, Cab, ALA, viewing azimuth (VAzimuth), and viewing zenith (VZenith). The task of estimating the parameters in a Bayesian network involves two procedures: a forward procedure to determine the statistical relation between land surface parameters and simulated reflectance, and an inverse procedure to estimate the parameters through the measured canopy reflectance data and other information (e.g. stage of crop growth). The posterior probability distribution of parameters can be calculated from Eq. (4) in order to estimate the maximum posterior probability value or mean value of the parameters. Compared to other methods (e.g. ANN) the Bayesian network method has the advantage of being a bidirectional inferential mechanism, i.e. it allows information to be obtained in the form of instantiated variable states forward or backward through nodes—the Bayesian network allows both deduction and abductions (Kalacska et al., 2005).

The equation to calculate the posterior probability of the parameters of interest can be derived from Fig. 3, which uses the following abbreviations to designate the nodes of the Bayesian network structure: T is the growth stage, A is ALA, and $V1$ and $V2$ are the relative azimuth angle and view zenith angle, respectively. Other variables are abbreviated to the first letter of the corresponding node. Given the reflectance values G, R, and N for the three bands, the posterior probability of LAI can be calculated using Eq. (5), which is deduced from Eq. (4)

$$p(L|T,V1,V2,G,R,N) = C \times p(L|T) \sum_{\{C_i,A_j\}} p(G,R,N|L,C_i,A_j,V1,V2) \quad (5)$$

where C is a constant factor and $p(L|T)$ is the conditional probability distribution of LAI during a given growth stage, derived from the spectral library as a form of discrete distribution (Fig. 4). This type of knowledge presentation in a discrete distribution is an extension of the traditional assumption that prior knowledge must obey the rule of normal distribution. The last factor represents the ability of the model to fit the observed data under certain conditions. Thus, the Bayesian method combines three sources of information: prior distribution of parameters, the physical model of remote sensing, and measured parameters. It can be seen from Eq. (5) that the estimate is a probability rather than a single value. More information about the estimated parameters can be deduced from their probability, such as the mean value, maximum posterior probability value, and information entropy, which can be used to describe the information content or the degree of uncertainty. For example, the posterior entropy can be calculated using Eq. (6)

$$H = -\sum_i p_i \log(p_i) \quad (6)$$

where p_i is the parameter's posterior probability on the ith state and log() is a logarithm operator to the base of 10 or 2 (Shannon, 1948).

3. Analysis of results

3.1. Extracting prior knowledge from the spectral library

In the approach suggested by Li et al. (2001), there are a number of ways to use prior knowledge in estimating the values of land surface parameters. The probability density distribution under different conditions can be regarded as one kind of prior knowledge. In the Bayesian network, extracting prior knowledge is a matter of constructing a conditional probability table (CPT) of the network nodes. Here the conditional probabilities are $P(\mathrm{LAI}|T)$ and $P(\mathrm{Cab}|T)$. They represent the probability density distribution of LAI and Cab at different stages of crop growth. Time series on LAI and Cab can be selected as the data

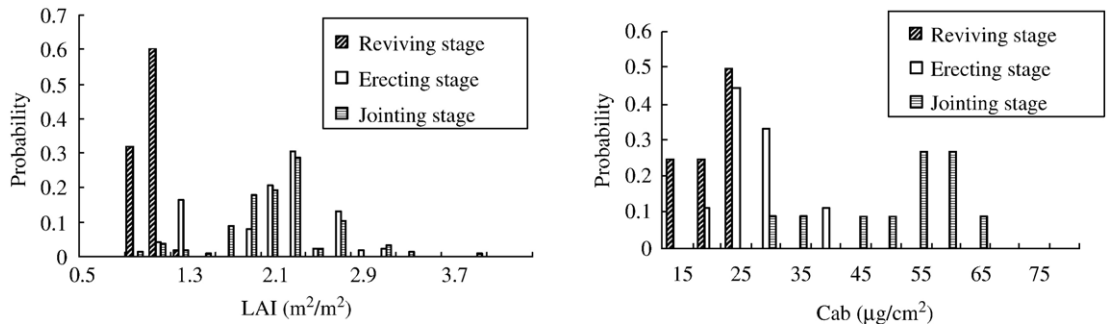

Fig. 4. Conditional probability distribution of LAI and Cab at different growth stages of winter wheat.

source to calculate the above conditional probabilities. In the SLTLSOC, we have collected time series data on biophysical and biochemical parameters of plants, which can be used as a data pool to calculate the CPT of LAI and Cab. As an example, Fig. 4 shows the statistical probability distribution of LAI during the reviving, erecting and jointing stages of winter wheat northern part of China.

3.2. Estimating LAI and Cab using simulated data

The simulated data set was divided into two subsets, one to train the Bayesian network and the other to test the trained network. A subset comprising 30 groups of data was selected from the simulated data in order to test the trained Bayesian network. It should be noted that since the simulated data set was used for both training and testing, information on different stages of crop growth was not introduced.

To test the Bayesian network, the posterior probability distributions of LAI and Cab were calculated and the values with maximum posterior probability were selected as the estimated values. With the values of LAI and Cab that served as inputs to the PROSAIL model as reference values, the estimated and true values of LAI and Cab were compared. The results are plotted in Fig. 5, which also shows the RMSE (root mean square error).

As can be seen from Fig. 5, the proposed method predicted LAI and Cab accurately. The results indicate that the Bayesian network can learn the relationship between the parameter space and the data space, and has inference capability from the data space to the parameter space simultaneously. It is also evident that greater values of both LAI (>3.5) and Cab (>45) lowered the accuracy: as LAI and Cab increase, they were less sensitive to bidirectional reflectance from the canopy. This result is in agreement with other studies on validation of the PROSAIL model (Jacquemoud et al., 1995) and also suggests that whenever LAI and Cab exceed a given value, more information (i.e. prior knowledge) may increase accuracy.

3.3. Estimating LAI and Cab using data from field measurements

The hybrid inversion method was also validated using data from field measurements. Field experiments were carried out from 26 March to 19 May 2001 in Shunyi District, Beijing, China. To compare spectral data and some parameters under different conditions, the experimental area was divided into different regions based on location (the north-western district was region NW, and so on). Each region was subdivided into five observation plots, numbered from 1 to 5. This section refers to plot NW4, the 4th plot in the north-western region (40°11′40.1″–40°11′51.4″N, 116°34′32.7″–116°34′49.4″E), in which measurements were carried out on 2, 12, 17, and 21 April. The data covered three growth stages of winter wheat (reviving, erecting, and jointing). The instrument used for measuring the BRDF of canopies was an SE590 FieldSpec hand-held spectrometer with a spectral response ranging from 400–1100 nm. The field of view was fixed at 25°, and the instrument was 1.5 m above the ground level.

Spectral information on the canopies and on biophysical parameters of leaves, e.g. LAI, was collected on clear days. To calculate LAI, leaf samples from an area of 0.6 m × 0.6 m were

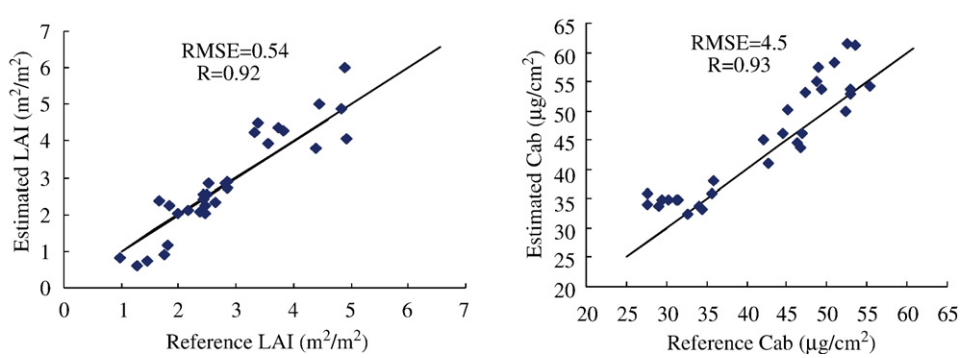

Fig. 5. Estimated LAI and Cab using simulated data.

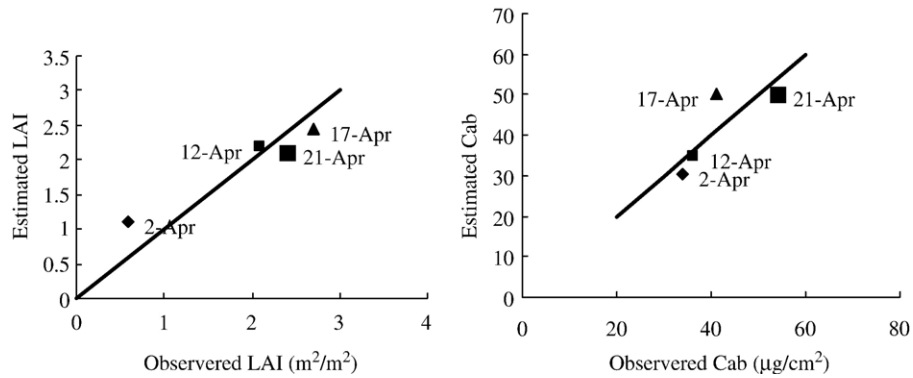

Fig. 6. Estimated results using field measurements.

collected and their total area (one side only) measured. The LAI can be obtained using Eq. (7)

$$\mathrm{LAI} = \frac{A_L}{0.36}(\mathrm{m}^2/\mathrm{m}^2) \tag{7}$$

where A_L is the total single side area of all collected leaves.

The biochemical parameters (e.g. chlorophyll content) of the leaf samples were determined in the laboratory. For Cab, a total of 50 leaves were sampled from each of the 10 sample sites, from each of the 5 plots in each region. From the central part of the leaves, 5 cm lengths were cut, mixed, and ground, and the total content of chlorophyll *a* (Cha) and chlorophyll *b* (Chb) was determined using a spectrophotometer. Cab was calculated as the total weight of Cha and Chb per unit leaf area.

In a Bayesian network, the value of a parameter is inferred from the posterior distribution of the parameter at a given growth stage and reflectance. The prior distribution is extracted from the Bayesian network by specifying the time period—in this case, from the beginning to the end of April. The measurement days covered the three growth stages in our study area: 2 April represented the reviving stage, 12 April the erecting stage, 17 April and 21 April both represented the jointing stage. Here, prior information on LAI and Cab was extracted as a form of probability distribution (Fig. 4). After the growth stage was specified in the Bayesian network, spectral data on green, red, and near-infrared bands resampled from hyper-spectral data were taken as the input parameters to estimate LAI and Cab.

The estimated value of LAI was close to the field measurement value (RMSE of 0.22; Fig. 6); the error in estimating Cab was equally obvious (RMSE of 4.0; Fig. 6). The maximum error in Cab was on 17 April and 21 April, when the absolute errors were respectively 9 μg/cm^2 and 4 μg/cm^2. As found for simulated data, when Cab was higher, the results were less reliable.

To illustrate the role of prior knowledge in estimating LAI and Cab, values calculated without the benefit of information on the growth stages are shown in Fig. 7. The RMSEs of LAI and Cab were 0.73 m^2/m^2 and 9.6 μg/cm^2, respectively. It can be seen from Fig. 7 that prior knowledge of the stage of growth improved the accuracy significantly. Since the introduced prior knowledge is derived from temporal information, the proposed method can adjust the behavior of retrieved results according to the growth trajectory of target parameters. Other research has revealed a similar result when temporal information (a canopy structure dynamic model) was used to adjust the retrieved parameters (Koetz et al., 2005).

3.4. Change in the posterior information in the inversion process

As stated in Section 2.3, the output of a Bayesian network is not just one value but a probability distribution of the parameter under study, which can be used to investigate changes in posterior information. In information theory, entropy is often employed to describe information content and its changes in a dynamic system (Maselli et al., 1994; Wang et al., 2001). In general, when the entropy is reduced, it indicates that the retrieved result has incorporated more information and its uncertainty is reduced, and vice versa. This section deals with changes in posterior information on LAI and Cab.

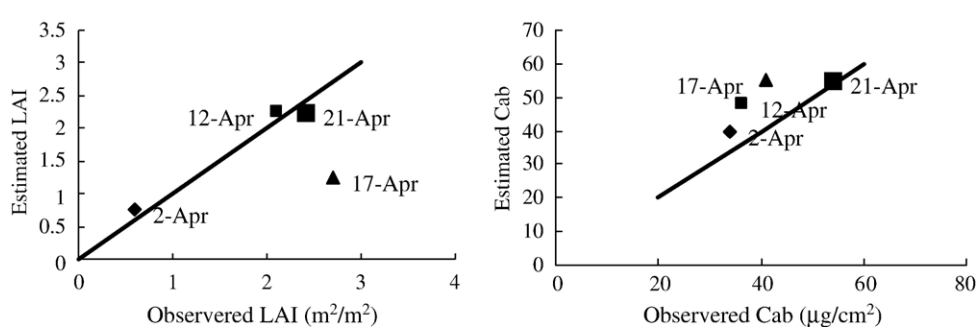

Fig. 7. Estimated results without prior knowledge.

Table 2
Configuration of input data, negative angles represent forward observations

Symbol	Combination of view zenith
D1	55°
D2	55°, 25°
D3	55°, 25°, 0
D4	55°, 25°, 0, −25
D5	55°, 25°, 0°, −25°, −55°

The posterior probability and the corresponding posterior entropy can be calculated using Eq. (5) and the definition of entropy in Eq. (6). It can be proved that this method of updating the posterior probability is independent of the sequence of data input (Chan & Darwiche, 2005). Based on this point, we studied the change in posterior information as a result of inputting the data on different viewing angles in sequence. The viewing zenith of the input data is shown in Table 2.

The posterior entropy of retrieved parameters was calculated (Fig. 8). There was a general trend toward reduction in posterior entropy; however, the rate of reduction slowed down as more multi-angular data were added and, occasionally, the entropy even increased with addition of data. In the LAI inversion, the observed data, which reduced entropy to the minimum, were all from the backward direction. The largest slowing down was generally near the hot spot (backward 25°). In the Cab inversion, we found the same trend for two days (2 and 12 April). On those days, after adding data from two backward view angles (55° and 25°), the posterior entropy was nearly at its minimum. With the addition of forward-looking data for LAI and Cab, the rate of entropy reduction declined. This was more obvious in the Cab inversion. The entropy increased on 12 April in the case of LAI and on 2 April in the case of Cab. The magnitude of increase was larger for LAI, but its maximum posterior probability value remained unchanged despite the increase in entropy (Fig. 9). The maximum posterior probability value of Cab changed from 24 μg/cm^2 to 34 μg/cm^2 compared to the measured value of 30 μg/cm^2.

The abnormal behavior of the posterior information change in the case of LAI and Cab posed a problem. The increasing entropy showed a conflict between adding information to the new data and adding it to the previously used data when using multi-angle observation data in the inversion—it actually increased the degree of uncertainty in the estimates. However, when the posterior entropy increased, the maximum posterior probability value simultaneously moved closer to the "true" value. We cannot adequately explain the change in the maximum posterior probability value with the increase in entropy, although the current experiment allows a preliminary conclusion. A change in entropy cannot be the only index of the degree of uncertainty in the values of parameters and was insufficient to express the entire information. Alternative methods to fully describe the change in information from inverting non-linear models used in remote sensing should be further investigated. Earlier efforts to solve this problem focused only on inverting linear models, which can retrieve the parameters using analytical methods (Yang et al., 2003).

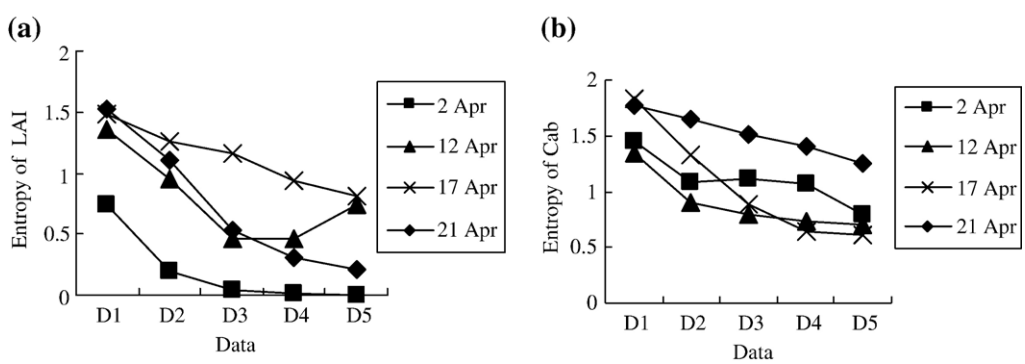

Fig. 8. Posterior entropy of LAI (a) and Cab (b) updated by adding data.

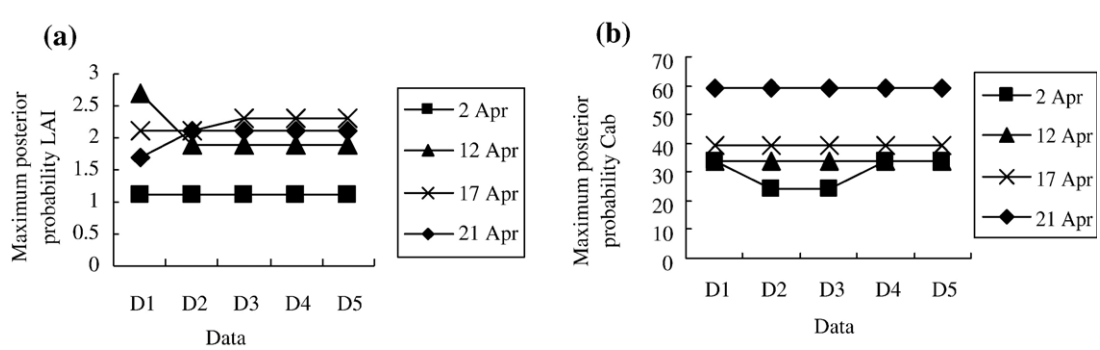

Fig. 9. Maximum posterior probability value of LAI (a) and Cab (b).

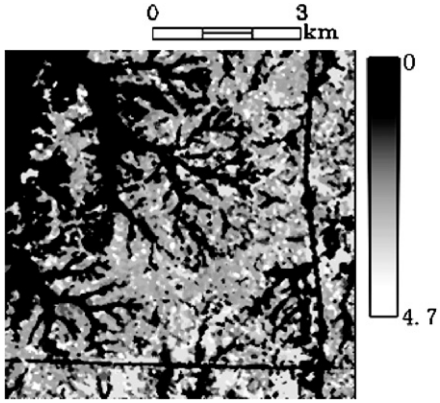

Fig. 10. Mapping LAI using ETM+ data.

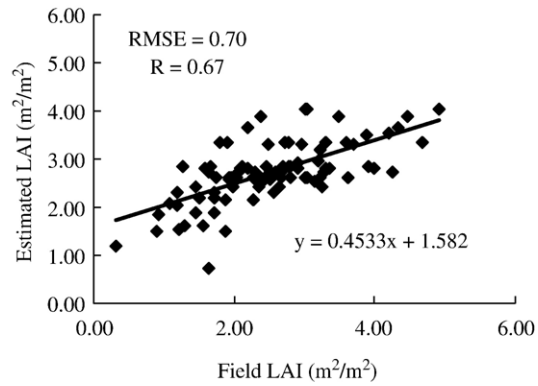

Fig. 11. Comparison and linear regression of measured and estimated LAI values.

3.5. Mapping LAI

This section describes how the proposed method was used for mapping LAI using satellite imagery. We used data from the BigFoot Project (Cohen et al., 2003b), which was designed to link parameters measured in situ and estimates of these parameters from satellite remote sensing. The BigFoot project collects field-based data over 5 km × 5 km domains and uses Landsat ETM+ image data and ecosystem process models to characterize 7 km × 7 km areas around each field measurement site. We used data collected from one of the BigFoot validation sites, KONZ, which is a tall-grass prairie in central Kansas, US. Although the study site also includes areas under forest and croplands, we investigated only the grassland: the reflectance values were therefore those of grassland extracted from ETM+ imagery according to the map of the land cover. The ETM+ imagery used was as on 13 August 2001. The image had a resolution of 25 m and covered an area 7 km × 7 km (Cohen et al., 2006).

It should be noted that although prior knowledge of the study target was required for estimating the LAI, it was not necessary to have such prior knowledge of every pixel covered by the satellite imagery; prior knowledge to enable a rough estimate of the target parameter was often adequate. So long as the land cover in the entire study area was uniform, although LAI values may have varied spatially, it was justifiable to assume the prior knowledge obeyed the same probability distribution for every cell in the ETM+ imagery. MOD15A2 is one of MODIS products that can capture the seasonal variation in LAI at a resolution of 1 km. In this study, prior knowledge was extracted from the MOD15A2 trajectory. Cohen et al. (2003b) has presented the LAI trajectory of MOD15A2 in the KONZ site, and we used the data as prior knowledge of LAI for this site. From the trajectory of 2001, LAI on 13 August (DOY is 225) had a mean value of 3.4 and standard variance of 1.2, which is what we inputted into the Bayesian network (as in Fig. 3) as a prior distribution before estimating the LAI of the BigFoot measurement plots from ETM+ and canopy reflectance. After establishing a model of the spectral data and LAI, we inputted the reflectance pixel-by-pixel to map the LAI (Fig. 10). The statistical properties of inferred LAI are given in Table 3. The statistic values of LAI retrieved from the Bigfoot Project (Cohen et al., 2006), which uses an orthogonal regression method called reduced major axis (Cohen et al., 2003a) also are presented in Table 3.

Table 3
Comparison of the statistical properties of the retrieved LAI map

Method	RMSE	R	Min.	Max.	Mean	S.D.	Median
Bayesian network	0.70	0.67	1.2	4.8	2.9	0.8	2.8
RMA	0.96	0.50	0.0	9.0	2.6	1.3	2.5

The results from both methods (Fig. 10, Table 3) were in reasonable agreement; the differences between mean and median were less than 0.4 m^2/m^2. However, in the result derived from the proposed method, the LAI value of grass at the KONZ site was spatially stable across the entire extent with a mean value of 2.93 and standard variance of 0.80.

Field measurements of LAI values from 82 plots were used to test the mapping accuracy of LAI (Fig. 11). In this study, the RMSE between the measured and the estimated LAI was 0.70 and the correlation between the measured and estimated LAI values was 0.67.

In general, the proposed method may have produced a better result; it improved the correlation coefficient from 0.50 to 0.67 and reduced RMSE from 0.96 to 0.70. However, where the LAI measured in the field was <2.0, the Bayesian network method often overestimated the value. When LAI was low, it is likely that the canopy reflectance value was more derived from the background soil than from the leaves. It is also possible that the PROSAIL model used did not differentiate between vegetation and vegetation mixed with soil. The BigFoot result takes into account the brightness, wetness, and greenness extracted from ETM+ data to establish a regression equation between the optical index and the LAI from field measurements. This type of empirical regression method may produce a better result than that of a model-based method in this case.

4. Discussion and conclusions

Based on a Bayesian network, a hybrid inversion scheme integrating observed data and new information into a unified

knowledge-inferring framework was proposed. This new scheme can incorporate information, in addition to physical model parameters, into the inversion process to estimate biophysical and biochemical parameters. Information for the inversion model may come from accumulated historical data, thus, the process of extracting knowledge from historical databases and using it for retrieving information from remotely sensed data is the accumulation of prior knowledge.

The Bayesian network was used to estimate the Cab and calculate LAI. The proposed method was tested against data from simulations, field measurements, and remote sensing. From experimental results, the following conclusions can be drawn.

(1) By training a Bayesian network using simulated data from the PROSAIL model, the mapping relationship between the parameter space and the data space can be established, which demonstrates the learning ability of the Bayesian network.
(2) The difference between the Bayesian network method and traditional Bayesian or ANN methods is that the former can consider the input parameters of physical models as well as other parameters not included in the model. In our approach, the prior knowledge can be statistically extracted from historical data and incorporated into the inversion process. This provides a possible way to introduce more parameters, such as crop growth stages, to assist in estimating different parameters of the land surface.
(3) Using simulated data, the inversion process estimates LAI and Cab more accurately, as seen from the RMSEs of 0.54 m^2/m^2 for LAI and 4.5 $\mu g/cm^2$ for Cab. Validation using data from field measurements shows that prior knowledge improved the estimates, especially as the parameters were less sensitive to canopy reflectance. The proposed method, when used for mapping LAI using ETM+ imagery, produced a slightly more accurate result than the empirical regression method, with an RMSE of 0.70 and a correlation of 0.67.
(4) In calculating the posterior probability of parameters, the entropy of parameters decreased as progressively more prior knowledge from new data was incorporated. Therefore, generally, as posterior information accumulated, the degree of uncertainty decreased in parameter estimations.
(5) Unlike the existing iterative algorithms based on optimization theory, the proposed method did not need initial parameter values. Prior information could be determined from the remote sensing database from the conditional probability distribution of the Bayesian network.

However, the proposed hybrid inversion method to calculate LAI and Cab considered the growth stage alone, assuming a certain relationship between the growth stage and LAI or Cab. In practice, however, such factors as planting conditions, soil conditions, water, and application of fertilizers also influence the growth stages of crops. Thus, integrating more agricultural knowledge into the estimation of land surface parameters may be essential for further remote sensing research.

The accuracy of inversion was influenced by sensitivity of the parameter. For example, if Cab value was sufficiently large, the parameter became less sensitive, making them less reliable. Improving the accuracy of the inversion process when the parameters are less sensitive is an area for future studies.

Acknowledgements

This work is supported by the Natural Science Foundation of China Project(40601059, 40571107), the Program for Changjiang Scholars and Innovative Research Team in University, the Open Fund of State Key Laboratory(LRSS0608) and the Special Fund for Doctor Education. We thank Y.Q. Xiang of the Institute of Geographical Science and Natural Resources Research at the Chinese Academy of Sciences, who provided us with field-measured data, and K.P. Murphy, who provided the source code of the Bayesian network toolbox. We also give thanks to P. Gong, H.L. Fang and Jack Teng for editing an earlier version of this paper.

References

Baret, F., & Fourty, T. (1997). Estimation of leaf water content and specific leaf weight from reflectance and transmittance measurements. *Agronomie*, *17*, 455−464.

Baret, F., & Guyot, G. (1991). Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sensing of Environment*, *35*, 161−173.

Chan, H., & Darwiche, A. (2005). On the revision of probabilistic beliefs using uncertain evidence. *Artificial Intelligence*, *163*, 67−90.

Cohen, W. B., Maiersperger, T. K., Gower, S. T., & Turner, D. P. (2003a). An improved strategy for regression of biophysical variables and landsat ETM+ data. *Remote Sensing of Environment*, *84*, 561−571.

Cohen, W. B., Maiersperger, T. K., & Pflugmacher, D. (2006). *Bigfoot leaf area index surfaces for north and south American sites, 2000–2003.* Data set. Available on-line [http://www.daac.ornl.gov] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A.

Cohen, W. B., Maiersperger, T. K., Yang, Z., Gower, S. T., Turner, D. P., Ritts, W. D., et al. (2003b). Comparisons of land cover and LAI estimates derived from ETM+ and MODIS for four sites in north America: A quality assessment of 2000/2001 provisional MODIS products. *Remote Sensing of Environment*, *88*, 233−255.

Combal, B., Baret, F., Weiss, M., Trubuil, A., Mace, D., Pragnere, A., et al. (2003). Retrieval of canopy biophysical variables from bidirectional reflectance: Using prior information to solve the ill-posed inverse problem. *Remote Sensing of Environment*, *84*, 1−15.

Doicu, A., Schreier, F., & Hess, M. (2003). Iteratively regularized Gauss–Newton method for bound-constraint problems in atmospheric remote sensing. *Computer Physics Communications*, *153*, 59−65.

Doicu, A., Schreier, F., & Hess, M. (2004). Iterative regularization methods for atmospheric remote sensing. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *83*, 47−61.

Fang, H. L., & Liang, S. L. (2003). Retrieving leaf area index with a neural network method: Simulation and validation. *IEEE Transactions on Geoscience and Remote Sensing*, *41*, 2052−2062.

Fymat, A. L. (1979). A generalization of Cooke's integral inversion formula with application to remote-sensing theory. *Applied Mathematics and Computation*, *5*, 23−39.

Gitelson, A. A., & Merzlyak, M. N. (1997). Remote estimation of chlorophyll content in higher plant leaves. *International Journal of Remote Sensing*, *18*, 2691−2697.

Gong, P., Pu, R., Biging, G. S., & Larrieu, M. (2003). Estimation of forest leaf area index using vegetation indices derived from hyperion hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, *41*, 1355−1362.

Gong, P., Pu, R., & Miller, J. R. (1992). Correlating leaf area index of ponderosa pine with hyperspectral casi data. *Canadian Journal of Remote Sensing*, *18*, 275−282.

Jacquemoud, S., Bacour, C., Poilve, H., & Frangi, J. P. (2000). Comparison of four radiative transfer models to simulate plant canopies reflectance: Direct and inverse mode. *Remote Sensing of Environment*, 471−481.

Jacquemoud, S., & Baret, F. (1990). Prospect: A model of leaf optical properties. *Remote Sensing of Environment*, *34*, 75−91.

Jacquemoud, S., Baret, F., Andrieu, B., Danson, F. M., & Jaggard, K. (1995). Extraction of vegetation biophysical parameters by inversion of the PROSPECT+ SAIL models on Sugar beet canopy reflectance data. Application to TM and AVIRIS sensors. *Remote Sensing of Environment*, *52*, 163−172.

Kalacska, M., Sanchez-Azofeifa, G. A., Caelli, T., Rivard, B., & Boerlage, B. (2005). Estimating leaf area index from satellite imagery using Bayesian networks. *IEEE Transactions on Geoscience and Remote Sensing*, *43*, 1866−1873.

Kimes, D. S., Harrison, P. R., & Ratcliffe, P. A. (1991). A knowledge-based expert system for inferring vegetation characteristics. *International Journal of Remote Sensing*, *12*, 1987−2020.

Koetz, B., Baret, F., Poilve, H., & Hill, J. (2005). Use of coupled canopy structure dynamic and radiative transfer models to estimate biophysical canopy characteristics. *Remote Sensing of Environment*, *95*, 115−124.

Kuusk, A. (1991). Determination of vegetation canopy parameters from optical measurements. *Remote Sensing of Environment*, *37*, 207−218.

Li, X., Gao, F., Wang, J., & Strahler, A. (2001). A priori knowledge accumulation and its application to linear brdf model inversion. *Journal of Geophysical Research*, *106*, 11,925−11,935.

Liang, S. (2004). *Quantitative remote sensing of land surfaces.* New York: John Wiley and Sons, Inc.

Marcot, B. G., Holthausen, R. S., Raphael, M. G., Rowland, M., & Wisdom, M. (2001). Using Bayesian belief networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement. *Forest ecology and management*, *153*, 29−42.

Maselli, F., Conese, C., & Petkov, L. (1994). Use of probability entropy for the estimation and graphical representation of the accuracy of maximum likelihood classifications. *ISPRS Journal of Photogrammetry and Remote Sensing*, *49*, 13−20.

Murphy, K. (1998). *A brief introduction to graphical models and Bayesian networks.* BNet software. Available on-line [http://www.cs.ubc.ca/~murphyk/Software/BNT].

Myneni, R. B., Maggion, S., & Laquinta, J. (1995). Optical remote sensing of vegetation: Modeling, caveats, and algorithms. *Remote Sensing of Environment*, *51*, 169−188.

Qu, Y. H., Liu, S. H., & Wang, J. D. (2003). The construction of j2ee-based spectrum knowledge base system for typical object in china. *Proceedings of International Geoscience and Remote Sensing Symposium, IGARSS'03*, *Vol. 4*. (pp. 3787−3789).

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 379−423.

Sims, D. A., & Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment*, *81*, 337−354.

Tian, Y., Wang, Y., Zhang, Y., Knyazikhin, Y., Bogaert, J., & Myneni, R. B. (2003). Radiative transfer based scaling of LAI retrievals from reflectance data of different resolutions. *Remote Sensing of Environment*, *84*, 143−159.

Verhoef, W. (1984). Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. *Remote Sensing of Environment*, *16*, 125−141.

Verstraete, M. M., Pinty, B., & Myneni, R. B. (1996). Potential and limitations of information extraction on the terrestrial biosphere from satellite Remote Sensing. *Remote Sensing of Environment*, *58*, 201−214.

Walthall, C., Dulaney, W., Anderson, M., Norman, J., Fang, H., & Liang, S. (2004). A comparison of empirical and neural network approaches for estimating corn and soybean leaf area index from LandSat ETM+ imagery. *Remote Sensing of Environment*, *92*, 465−474.

Wang, G., Gertner, G., Parysow, P., & Anderson, A. (2001). Spatial prediction and uncertainty assessment of topographic factor for revised universal soil loss equation using digital elevation models. *ISPRS Journal of Photogrammetry and Remote Sensing*, *56*, 65−80.

Yager, R. R. (2006). An extension of the naive bayesian classifier. *Information Sciences*, *176*, 577−588.

Yang, H., Xu, W., & Zhao, H. (2003). Information stream in regularization inversion for quantitative remote sensing and its control. *China Science (Series D)*, *33*, 799−808.