A dynamic Bayesian network data fusion algorithm for estimating leaf area index using time-series data from in situ measurement to remote sensing observations

Yonghua Qu\textsuperscript{a,b}, Yuzhen Zhang\textsuperscript{a,b} & Jindi Wang\textsuperscript{a,b}

\textsuperscript{a} State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Beijing Normal University and Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, 100875, China

\textsuperscript{b} Beijing Key Laboratory for Remote Sensing of Environment and Digital Cities, Beijing Normal University, Beijing, 100875, China

Available online: 02 Nov 2011

To cite this article: Yonghua Qu, Yuzhen Zhang & Jindi Wang (2012): A dynamic Bayesian network data fusion algorithm for estimating leaf area index using time-series data from \emph{in situ} measurement to remote sensing observations, International Journal of Remote Sensing, 33:4, 1106-1125

To link to this article: http://dx.doi.org/10.1080/01431161.2010.550642

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings,
A dynamic Bayesian network data fusion algorithm for estimating leaf area index using time-series data from in situ measurement to remote sensing observations

YONGHUA QU*†‡, YUZHEN ZHANG†‡ and JINDI WANG†‡
†State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Beijing Normal University and Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing 100875, China
‡Beijing Key Laboratory for Remote Sensing of Environment and Digital Cities, Beijing Normal University, Beijing 100875, China

(Received 5 November 2009; in final form 17 November 2010)

Leaf area index (LAI) products retrieved from remote sensing observations have been widely used in the fields of ecosphere, atmosphere etc. However, because satellite-observed images are captured instantaneously and sometimes screened by cloud, some current LAI products are inherently discontinuous in time and their accuracy may not meet the needs of users well. To solve these problems, we proposed a dynamic Bayesian network (DBN)-based data fusion algorithm that integrates dynamic crop growth information, a canopy reflectance (CR) model and remote sensing observations from the perspective of Bayesian probability. Using the proposed algorithm, LAI was estimated using data sets from both field measurements for winter wheat in Beijing, China, and MODIS reflectance data at two American flux tower sites. Results showed good agreement between the LAI estimated by the DBN-based data fusion method and the true ground LAI, with a correlation coefficient of \( R \) 0.95 and 0.96, respectively, and a corresponding root mean square error (RMSE) of 0.35 and 0.49, respectively. In addition, the LAI estimated by the DBN-based data fusion method formed a continuous time series and was consistent with the variety law of vegetation growth at both plot and flux tower site scales. It has been demonstrated that the proposed DBN-based data fusion algorithm has the potential to be used to accurately estimate LAI and to fill the temporal gap by integrating information from multiple sources.

1. Introduction

Leaf area index (LAI) is a key characteristic of vegetation structure and surface energy balance. It controls many biological and physical processes in plant canopies. Accurate LAI estimates are required in studies of global change, land surface process and ecological environment (Chen et al. 1997). Algorithms used for estimating LAI are traditionally grouped into two categories: statistical algorithms and physical algorithms (Liang 2007). Statistical algorithms are mainly based on vegetation index (Colombo et al. 2003), and physically based inversion algorithms are built on canopy reflectance (CR) models (Meroni et al. 2004). Researchers devoted to retrieving LAI from remote sensing observations using both algorithms have made some
DBN-based data fusion algorithm for estimating LAI

progress. However, the estimated results are often affected by noise and measurement uncertainty, and the accuracy of estimates leaves much room for improvement. Moreover, satellite observations are discontinuous because of the effect of phase factors and sometimes weather conditions and instrument problems, and remote sensing geo-information gives only a reflection of the instantaneous physical state of the crop community, which cannot reveal crop growth; reliance solely on satellite observations does not allow continuous features of vegetation structural parameters or spatio-temporal evolitional information to be captured completely. However, crop growth functions can provide information complementary to remote sensing observations, so evolitional information as well as continuous features of land surface parameters can be obtained and at the same time the accuracy of estimates can be improved by fusing crop growth functions and remote sensing observations using data fusion algorithms.

Data fusion algorithms are often used to improve the accuracy of estimates and obtain more a complete knowledge of the state. Much work has been done in developing methods of data fusion. The basic approach has been to fuse the information using weighted-average techniques (Punska 1999). It is the most intuitive and simplest data fusion method. Its greatest shortcoming is that weighted values of varying information are empirical and difficult to determine in most cases. The Kalman filter (KF) is another well-known fusion algorithm. It is an efficient recursive filter modelled on a Markov chain built on linear operators; it has been widely used in estimating the state of linear dynamic systems from a series of noisy measurements (Welch and Bishop 1995). Unfortunately, because most processes in real life are not linear, the linear assumption of the basic KF is not appropriate to this study. To overcome this obstacle, various non-linear filter algorithms, termed Extended KF, have been developed (Wan and Van Der Merwe 2000, Han and Li 2008). However, these filters all hold to the assumption that the probability density function (PDF) is Gaussian with the result that they do not work well in situations where the true posterior is known to be non-Gaussian (Han and Li 2008). The dynamic Bayesian network (DBN)-based data fusion algorithm, not limited to the above assumptions, is proposed in this article.

The DBN-based data fusion algorithm relies on the classical Bayesian paradigm. Both the fusion procedure of data from multiple sources and the uncertainties of information are treated from a perspective of probability, facilitating the gathering of precise land surface parameters. The starting point of the proposed algorithm is the Bayesian perspective. The primary processes of the DBN-based data fusion method can be summarized as prediction and updating (Ihler et al. 2007, Wikle and Berliner 2007). The crucial step of prediction is the Chapman–Kolmogorov integral, which combines prior information with crop growth functions. Bayes’ rule is used to update the prediction when new data become available. The fusion of all available information is completed after two steps and the optimal state of the target parameter is obtained. Better continuous and evolitional information of vegetation structural parameters can then be captured through dynamic iteration.

In this study, we estimated the LAI using the DBN-based data fusion algorithm at both the plot and site scales, and the results were evaluated by field measurements and moderate resolution imaging spectroradiometer (MODIS) LAI data.

2. DBN-based data fusion algorithm

Bayesian networks are directed graphical models of stochastic processes, where nodes represent random variables and the lack of arcs represents conditional independence.
assumptions (Murphy 1998). The use of Bayesian networks to estimate LAI was inves-
tigated by Kalacska et al. (2005) and Qu et al. (2008) and proved to be feasible. DBN is
the expansion of a Bayesian network in time series and describes the update processes
of state variables during that time.

When the dynamic process models and remote sensing observations are used in
combination to estimate LAI using DBN, two main goals need to be achieved: first,
perform inference to get a posterior probability distribution, and then estimate the
parameters from the posterior probability distribution.

2.1 Inference in DBN

The general inference goal in a DBN is to compute the posterior probability
$P(X_t^i|y_{1:τ})$, where $X_t^i$ represents the $i$th hidden variable at time $t$, and $y_{1:τ}$ represents
all evidence up to time $τ$ (Murphy 1998, 2002). Our interest is the special case of
$τ = t$, which represents real-time estimates of the target parameters. The fundamen-
tal problem to be solved is to seek a new estimate of the target state $P(X_t^i|y_{1:t})$ given
the old estimate $P(X_{t-1}^i|y _{1:t-1})$ (Challa and Koks 2004, Wikle and Berliner 2007). The
derivation process is given in detail below.

According to Bayes’ rule,

$$P(A|B) = \frac{p(B|A)p(A)}{p(B)},$$

where $p(A)$ is prior probability, and $p(B|A)$ is the probability of $B$ occurring given that
$A$ has already occurred. It can be converted to a likelihood function when $B$ takes
a specific value. $p(B)$ is a normalization and $p(A|B)$ is the posterior probability of $A$
occurring.

For a DBN, prior probability derives from the dynamic model (Lorenc 1986), so
equation (1) may be expanded as:

$$P(A|B, C) = \frac{P(B|A, C) \times P(A|C)}{P(B|C)}.$$

Assuming time-series observations have already been obtained, namely
$Y_T = (y_1, y_2, y_3, \ldots, y_{T-1}, y_T)$, and estimation of vegetation parameters $x_T$ at
time $T$ is $P(x_T|y_1, y_2, y_3, \ldots, y_{T-1}, y_T)$, or $P(x_T|Y_{T-1}, y_T)$, as is shown in figure 1.

In figure 1, the grey filled nodes represent observed variables and the white nodes
represent hidden variables (including variables to be estimated). From equation (2),
the states of the variable $x_T$ in the time series can be estimated from

$$P(x_T|y_T, Y_{T-1}) = \frac{P(y_T|x_T, Y_{T-1}) \times P(x_T|Y_{T-1})}{P(y_T|Y_{T-1})}.$$
The three terms in this equation are the likelihood term \( P(y_T|x_T, Y_{T-1}) \), the predicted term \( P(x_T|Y_{T-1}) \) and the normalization term \( P(y_T|Y_{T-1}) \).

The likelihood term deals with the probability of an observation \( y_T \). Assume that the noise is ‘whiteness’, so that the latest observation does not depend on previous observations but is determined only by the true state at the current time. In this case, the likelihood can be simplified as

\[
P(y_T|x_T, Y_{T-1}) = P(y_T|x_T). \tag{4}\]

We will also assume that the dynamic system obeys a first-order Markov evolution (Dowd 2007), implying that its future state depends directly on its current state and is independent of its previous state. Using the Chapman–Kolmogorov equation (Wikle and Berliner 2007), the predicted term is deduced as follows:

\[
P(x_T|Y_{T-1}) = \int P(x_T|x_{T-1})P(x_{T-1}|Y_{T-1})dx_{T-1}. \tag{5}\]

Lastly, the normalization can be expanded after Chapman–Kolmogorov, using the simplified likelihood and the predicted term:

\[
P(y_T|Y_{T-1}) = \int P(y_T|x_T)P(x_T|Y_{T-1})dx_T. \tag{6}\]

Equation (3) relates \( P(x_T|Y_T) \) to \( P(x_{T-1}|Y_{T-1}) \) via equations (4)–(6). Our fundamental problem is solved. Synthesize further and we will get the following equation:

\[
P(x_T|Y_T) = \frac{P(y_T|x_T) \int P(x_T|x_{T-1})P(x_{T-1}|Y_{T-1})dx_{T-1}}{\int P(y_T|x_T)P(x_T|Y_{T-1})dx_T}. \tag{7}\]

Equation (7) shows that the DBN-based data fusion method integrates multi-source information – for example, in this case, data from remote sensing observations \( Y_T \), dynamic process models \( P(x_T|x_{T-1}) \) and observation models \( P(y_T|x_T) \).

### 2.2 Estimation in DBN

Once the posterior probability has been obtained, it will be used to estimate the state of a target parameter. An intuitive approach is to find the most likely values on the basis of probability according to some criterion (Punska 1999). The following are the most frequently used estimates.

- Maximum a posteriori (MAP) estimator:
  \[
  \hat{x}_{MAP} = \text{arg max } P(x_i|Y_t). \tag{8}\]

- Minimum mean square error (MMSE) estimator:
  \[
  \hat{x}_{MMSE} = \text{arg min } \mathbb{E}_{P(x_i|Y_t)} \left\{ (\hat{x}_i - x_i)(\hat{x}_i - x_i)^T \right\}. \tag{9}\]

In this article, equation (8) is selected to estimate parameters from posterior probability for simplicity.
After the above two steps, the target parameter at a certain time slice is obtained. Through such dynamic iteration, the DBN-based data fusion method has completed the inversion of parameters in a time series.

3. Data and method

3.1 Simulated data sets using canopy reflectance model

The coupled PROSPECT leaf optical properties model and scattering by arbitrarily inclined leaves (SAIL) canopy reflectance model, also referred to PROSAIL, was used to simulate the reflectance of vegetation canopies (Jacquemoud et al. 2009). PROSPECT simulates the leaf’s hemispherical transmittance and reflectance using its biochemical and biophysical parameters, namely the leaf structure parameter $N$, the leaf chlorophyll a + b concentration $C_{ab}$, water content $C_w$ and dry matter content $C_m$ (Jacquemoud and Baret 1990, Jacquemoud et al. 2000). SAILH is a physically based model widely used for simulating the CR at different viewing directions (Verhoef 1984, Kuusk 1985). It needs several input parameters: LAI, average leaf angle (ALA), leaf hemispheric reflectance (LR), leaf transmittance (LT), soil reflectance (SR), atmospheric visibility (VIS), hot-spot parameter (SL), solar zenith angle (SZA), view zenith angle (VZA) and relative azimuth angle (RAZ); LR and LT simulation data are obtained from the PROSPECT model.

Reasonable error of input parameters in the CR model is permitted and does not lead to loss of accuracy in the inversion (Goel and Strebel 1983). For the sake of simplicity, some parameters in the PROSAIL model can be fixed when simulating the CR. Input parameters in the PROSPECT model are taken from the Leaf Optical Properties Experiment 93 (LOPEX’93) database (Hosgood et al. 1995). VIS is related to the diffuse part of incoming radiation and it also can be given a fixed value in the simulation. SR is determined on the basis of the simulated results of 25 different soils in algorithms of the MODIS LAI product (Knyazikhin et al. 1999). The SR simulations are mainly affected by moisture level, soil roughness and percentage of clay, sand and peat in the soil in the study area (Jacquemoud et al. 1992, Shepherd and Walsh 2002, Brown 2007). Through this information along with existing simulated results, a fixed input value for SR is obtained. The detailed inputs in the PROSAIL model are shown in Table 1.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Units</th>
<th>Range (or value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAI</td>
<td>m² m⁻²</td>
<td>0–8.0</td>
</tr>
<tr>
<td>ALA</td>
<td>degree</td>
<td>15–60</td>
</tr>
<tr>
<td>$C_{ab}$</td>
<td>μg cm⁻²</td>
<td>20–60</td>
</tr>
<tr>
<td>$C_m$</td>
<td>g cm⁻²</td>
<td>0.01</td>
</tr>
<tr>
<td>$C_w$</td>
<td>cm</td>
<td>0.02</td>
</tr>
<tr>
<td>$N$</td>
<td>–</td>
<td>1.5</td>
</tr>
<tr>
<td>SL</td>
<td>–</td>
<td>0.25</td>
</tr>
<tr>
<td>VIS</td>
<td>km</td>
<td>20</td>
</tr>
<tr>
<td>SZA</td>
<td>degree</td>
<td>25–55</td>
</tr>
<tr>
<td>VZA</td>
<td>degree</td>
<td>5–55</td>
</tr>
<tr>
<td>RAZ</td>
<td>degree</td>
<td>125</td>
</tr>
</tbody>
</table>
Based on the parameters in table 1, we generated a look-up table (LUT) comprising 172,800 combinations of LAI (80 classes), ALA (10 classes), Cab (9 classes), SZA (4 classes) and VZA (6 classes).

### 3.2 In situ data

*In situ* data were collected from a standard spectrum library of typical land surface objects at http://spl.bnu.edu.cn. The experimental field was located in Xiaotangshan district (116° 44' E, 40° 18' N) in Beijing, China. The average annual temperature in this experimental field was 11.8°C and annual precipitation was 644 mm. The field data set mainly involves canopy bidirectional reflectance under different viewing directions: leaf reflectance ($\rho_l$), transmittance ($\tau_l$), SR ($\rho_s$) and vegetation structural parameters LAI and leaf angle distribution (LAD) for the year 2004. LAI was measured using an LAI-2000 plant canopy analyser (UNIVERSAL) (Li-Cor, Inc., Lincoln, NE, USA); canopy bidirectional reflectance was measured using an analytical spectral device. Reflectance employed in this study was in the near-infrared region (wavelength 850 nm) and viewing zeniths 0°, 15°, 30°, 45° and 60°. The observation dates were the day of year (DOY) 92, 108, 124, 132, 141, 151 and 161 (table 2). During this period, the corresponding vegetation type is winter wheat. The DOY and their corresponding LAI were then fitted to the curve of the crop growth function at the plot scale.

A variety of growth curves have been developed to model general biological growth. Most successful predictive models are based on extended forms of the classical Verhulst logistic growth equation (Tsoularis and Wallace 2002). The crop growth function adopted in this study is the modified Verhulst logistic equation (Lin *et al.* 2003):

$$x = \frac{d}{1 + \exp(at^2 + bt + c)},$$

where $x$ is LAI, $t$ is DOY and $a$, $b$, $c$ and $d$ are undetermined coefficients that can be fitted using available field data. Only time factors are included and applicable to the plot scale where spatial heterogeneity can be ignored. In practice, LAI dynamics are mainly governed by the phenology (Weiss *et al.* 2001), but the effect of phenology is constant for any given small plot.

### Table 2. Field data set used in the estimation of LAI in the Xiaotangshan experimental field.

<table>
<thead>
<tr>
<th>DOY</th>
<th>LAI (m² m⁻²)</th>
<th>$\rho_l$</th>
<th>$\tau_l$</th>
<th>$\rho_s$</th>
<th>ALA (degree)</th>
<th>VIS (km)</th>
<th>SZA (degree)</th>
</tr>
</thead>
<tbody>
<tr>
<td>92</td>
<td>1.06</td>
<td>0.50</td>
<td>0.46</td>
<td>0.29</td>
<td>53.7</td>
<td>20</td>
<td>39.00</td>
</tr>
<tr>
<td>108</td>
<td>4.01</td>
<td>0.49</td>
<td>0.47</td>
<td>0.24</td>
<td>31.1</td>
<td>20</td>
<td>41.88</td>
</tr>
<tr>
<td>124</td>
<td>4.29</td>
<td>0.45</td>
<td>0.52</td>
<td>0.14</td>
<td>51.8</td>
<td>20</td>
<td>35.07</td>
</tr>
<tr>
<td>132</td>
<td>3.85</td>
<td>0.45</td>
<td>0.50</td>
<td>0.26</td>
<td>53.4</td>
<td>20</td>
<td>27.93</td>
</tr>
<tr>
<td>141</td>
<td>2.98</td>
<td>0.49</td>
<td>0.47</td>
<td>0.28</td>
<td>38.9</td>
<td>20</td>
<td>38.14</td>
</tr>
<tr>
<td>151</td>
<td>1.30</td>
<td>0.49</td>
<td>0.48</td>
<td>0.26</td>
<td>41.1</td>
<td>20</td>
<td>40.17</td>
</tr>
<tr>
<td>161</td>
<td>0.84</td>
<td>0.47</td>
<td>0.39</td>
<td>0.23</td>
<td>52.4</td>
<td>20</td>
<td>21.50</td>
</tr>
</tbody>
</table>
3.3 Observational data from micrometeorological tower sites

On a larger scale, the phenology factor should be considered to represent the spatial heterogeneity in a crop growth function. Studies have been carried out combining phenology models with land-surface biology growth models to provide dynamic, large-scale predictions of interactions between vegetation and climate (Foley et al. 1998). In this study, the phenology models are mainly driven by meteorological data derived from micrometeorological tower sites, namely, the Rosemount G19 Alternative Management Corn Soybean Rotation (G19) site and the Rosemount G21 Conventional Management Corn Soybean Rotation (G21) site (Botta et al. 2000) (table 3). Both sites belong to the AmeriFlux network (http://public.ornl.gov/ameriflux).

Among numerous meteorological data sets, air temperature, vapour pressure deficit and global radiation have been chosen to build a simple growing season index (GSI) phenology model (Jolly et al. 2005, Stockli et al. 2008). In the model, vegetation growth is controlled by environmental factors (temperature, light and humidity). If soil and other conditions are suitable for vegetation growth, assuming that temperature, sunshine and water affect the spatial variations in vegetation, the prognostic phenology model can simulate dynamic LAI changes on a regional scale. The analytic equation for GSI is simply the product of three factors:

\[ GSI = f(\bar{T}_m) f(\bar{R}_g) f(\overline{VPD}) \]  

where \( \bar{T}_m \) is minimum daily temperature, \( \bar{R}_g \) is mean daily global radiation and \( \overline{VPD} \) is mean daily vapour pressure deficit. \( \bar{T}_m \), \( \bar{R}_g \) and \( \overline{VPD} \) are 20-day moving average of daily \( T_m \), \( R_g \) and \( VPD \).

\[ f(\bar{T}_m) = \frac{\bar{T}_m - T_{m\min}}{T_{m\max} - T_{m\min}}, \]  

\[ f(\bar{R}_g) = \frac{\bar{R}_g - R_{g\min}}{R_{g\max} - R_{g\min}}, \]  

\[ f(\overline{VPD}) = 1 - \frac{\overline{VPD} - VPD_{\min}}{VPD_{\max} - VPD_{\min}}, \]  

where the empirical climate parameters \( T_{m\max}, T_{m\min}, R_{g\max}, R_{g\min}, VPD_{\max} \) and \( VPD_{\min} \) are maximum and minimum \( \bar{T}_m \), \( \bar{R}_g \) and \( \overline{VPD} \) values, and they were set according to the assimilated results of croplands reported by Stockli et al. (2008) (table 4). \( f(\bar{T}_m) \), \( f(\bar{R}_g) \) and \( f(\overline{VPD}) \) vary linearly between the constraining limits of 0 and 1, and thus regulate vegetation activity.

Since phenology together with climate factors contribute to LAI, the general form of the crop growth function at site scales can be depicted as:

\[ LAI = f(t) * g(GSI) + \text{error}. \]  

First-order Taylor series expansion of \( g(GSI) \) gives

\[ x = \frac{dGSI + e}{1 + \exp(at^2 + bt + c)}. \]
<table>
<thead>
<tr>
<th>CEIP ID</th>
<th>Site name</th>
<th>Vegetation type</th>
<th>Location</th>
<th>Latitude (+/-N)</th>
<th>Longitude (+/-E)</th>
<th>Elevation (m)</th>
<th>Years covered by the analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>US-Ro3</td>
<td>Rosemount G19</td>
<td>Alternative Management Corn Soybean Rotation</td>
<td>Minnesota, USA</td>
<td>44.7217</td>
<td>-93.0893</td>
<td>259.7</td>
<td>2004–2006</td>
</tr>
<tr>
<td>US-Ro1</td>
<td>Rosemount G21</td>
<td>Conventional Management Corn Soybean Rotation</td>
<td>Minnesota, USA</td>
<td>44.7143</td>
<td>-93.0898</td>
<td>259.7</td>
<td>2004–2006</td>
</tr>
</tbody>
</table>

Table 3. Location and description of micrometeorological tower sites.

Table 4. Maximum and minimum values of the meteorological factors.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Units</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{m\text{max}}$</td>
<td>°C</td>
<td>15.0</td>
</tr>
<tr>
<td>$T_{m\text{min}}$</td>
<td>°C</td>
<td>0.0</td>
</tr>
<tr>
<td>$R_{g\text{max}}$</td>
<td>W m$^{-2}$</td>
<td>252.0</td>
</tr>
<tr>
<td>$R_{g\text{min}}$</td>
<td>W m$^{-2}$</td>
<td>152.0</td>
</tr>
<tr>
<td>$VPD_{\text{max}}$</td>
<td>kPa</td>
<td>3.5</td>
</tr>
<tr>
<td>$VPD_{\text{min}}$</td>
<td>kPa</td>
<td>2.1</td>
</tr>
</tbody>
</table>

where $x$ is LAI, $t$ is DOY and $a$, $b$, $c$, $d$ and $e$ are undetermined coefficients. Equation (16) is the crop growth function used to generate dynamic estimates of LAI for a site.

The crop growth function for a site is based on the DOY and GSI calculated from the meteorological data and the corresponding LAI measured in the footprint of a tower site. LAI measurement protocols were developed for the AmeriFlux tower sites in order to avoid the inconsistency of the measured LAI caused by its various definitions, instruments etc. and to obtain the necessary variables for computing the LAI. For example, measure LAI using LAI-2000 at all sites if possible, otherwise measure at selected solar zenith angles using the Tracing Radiation and Architecture of Canopies (TRAC) instrument (3rd Wave Engineering, Nepean, ON, Canada). The detailed information such as sampling and measurement protocols of LAI is described by Law et al. (2008).

3.4 MODIS surface reflectance products and LAI products

The MODIS surface reflectance product MOD09A1 (Collection 5) serves as remote sensing observations in time series. It is a seven-band product with a spatial resolution of 500 m provided by a MODIS land science team. Red band (620–670 nm) and near-infrared (841–876 nm) band reflectance observations in the footprint of a flux tower were required for this study. In addition, the MODIS LAI product MOD15A2 (Collection 5) was used to compare with the estimated LAI derived from the DBN-based data fusion method.

3.5 Retrieving LAI using the DBN-based data fusion method

Using the DBN-based data fusion algorithm (see §2) to retrieve LAI from the above all data sets, we need to establish the following quantities:

Likelihood $P(y_T|x_T)$: This is derived from the observation model (CR model) from

$$y_T = H(x_T) + \eta,$$  \hspace{1cm} (17)

where $y_T$ is CR at the time $T$, $H$ is an observation operator and $\eta$ is observational error.

Predicted density $P(x_T|Y_{T-1})$: From equation (5), transition density $P(x_T|x_{T-1})$ and estimates of the previous time slice $P(x_{T-1}|Y_{T-1})$ are required. Transition density results from crop growth functions which can be simplified as:
where $x_T$ is LAI at the time $T$, $F$ is a model operator and $\varepsilon$ is the modelling error.

Normalization $P(y_T|Y_{T-1})$: This is an integral over quantities of likelihood and predicted terms already dealt with. From equations (7), (17) and (18),

$$P(x_T|Y_T) = \frac{P(y_T - H(x_T)) \int P(x_T - F(x_{T-1}))P(x_{T-1}|Y_{T-1})dx_{T-1}}{\int P(y_T - H(x_T))P(x_T|Y_{T-1})dx_T}$$

$$\propto P(y_T - H(x_T)) \int P(x_T - F(x_{T-1}))P(x_{T-1}|Y_{T-1})dx_{T-1}. \tag{19}$$

Equation (19) is the core of the DBN-based data fusion method, and it describes how to integrate the multi-source information from crop growth functions, satellite observes, meteorological factors and so on. The main procedure can be summarized as: it predicts the state variables with the dynamic system through one-dimensional integration over the previous state, then updates the integration with new observations.

When the nodes in figure 1 represent continuous variables, equation (19) is well suited for estimating LAI; when they represent discrete variables, the equation should be written in the form:

$$P(x_T|Y_T) = \frac{P(y_T|x_T) \sum_{x_{T-1}} P(x_T|x_{T-1}) \times P(x_{T-1}|Y_{T-1})}{\sum_{x_T} P(y_T|x_T) \times P(x_T|Y_{T-1})}. \tag{20}$$

A detailed flowchart of the DBN-based data fusion algorithm is shown in figure 2. At a plot scale, spatial heterogeneity is not obvious, so the crop growth function includes only the time factor DOY represented by the grey shaded box in figure 2. Studies on a larger scale, however, cannot ignore spatial heterogeneity. In this case, the
crop growth function is driven by meteorological factors in addition to time factors (the growth stage). Based on crop growth functions and estimates of LAI at time slice $T-1$, the predicted probability distribution at time slice $T$ is obtained. Then, through computing the likelihood probability of observed reflectance via the CR model, the predicted probability distribution at time slice $T$ will be updated and the posterior probability distribution acquired. The updated probability is used as input into the next time slice $T+1$. This process is then repeated.

4. Results

In this study, LAI was estimated at plot and site scales using the DBN-based data fusion method. At the plot scale, for the results using in situ data the bias of estimated versus referenced LAI values was evaluated using the cross-validation method (Efron and Gong 1983, Schneider 1997, Yang et al. 2007). At the site scale, the in situ measured LAI and MODIS LAI product were both used.

4.1 Estimated results using in situ data at the plot scale

In the study area, crop growth function (equation (10)) was determined empirically. The indeterminate coefficients need to be obtained from field-measured LAI values. Because there were only seven measurements of LAI at the growth stage of the winter wheat, we used the leave-one-out cross-validation method to evaluate the estimated results using the DBN-based data fusion method. In this method the key was to evaluate the integrals, which was straightforward for Gaussian models. The non-linear observation and evolution model in this article were handled during the process of data fusion through linearization of the model and evolution operators, and the detailed formula derivation is given in Challa and Koks (2004). For non-Gaussian error structures, Monte Carlo simulation methods were used to evaluate integrals.

To assess the performance of the DBN-based method, we estimated the LAI in the Xiaotangshan experimental crop by using both the DBN-based data fusion algorithm and inversion of the traditional physical PROSAIL model (figure 3). A cross-validation approach was then used on the estimated results: one point was selected, and the other points were fitted to the crop growth function during an inversion process, giving seven estimates for each DOY. The maximum and minimum DBN-data fusion LAI values at the growth stage are shown in figure 3. The cross-validation results compared well with the estimated LAI values using the DBN-based data fusion algorithm and field-measured LAI (root mean square error (RMSE) = 0.35, correlation coefficient ($R$) = 0.95) (figure 4), but the PROSAIL model underestimated LAI values (figure 3) possibly due to systematic offset of measurements in the near-infrared region for simulations using a radiative transfer model (RTM). In addition, LAI estimated by PROSAIL model inversion were only as a set of time-discrete points corresponding to the specific times at which reflectance was observed. The resulting discontinuous curve did not conform with LAI dynamics. However, introduction of the crop growth function modified the estimated results to produce a continuous curve reflecting the dynamic LAI data for the growth season.

4.2 Estimated results at the site scale

At the G19 and G21 sites, we estimated the LAIs for which corresponding field measurements were available during 2004–2006. All parameters were viewed as discrete
Figure 3. Estimated leaf area index (LAI) at the Xiaotangshan experimental field using various methods.
Note: DBN, dynamic Bayesian network; SAIL, scattering by arbitrarily inclined leaves; DOY, day of year.

Figure 4. Comparison between estimated leaf area index (LAI) using dynamic Bayesian network-based data fusion algorithm and field-measured LAI in the Xiaotangshan experimental field.
Note: RMSE, root mean square error.

variables during the process of DBN-based data fusion. At each site for each year, five points were set aside for validation of the estimated values (i.e. 30 points in all); all other points were used for fitting the curve to the crop growth equation (16). Having obtained the underdetermined coefficients for equation (16), computation of the
conditional transition probability $P(x_T \mid x_{T-1})$ was straightforward. Combining this value with the posterior probability for the previous time slice, we obtained the predicted term $P(x_T \mid y_{T-1}) \sum_{x_{T-1}} P(x_T \mid x_{T-1}) \times P(x_{T-1} \mid y_{T-1})$ in equation (20). Finally, the new incoming MODIS reflectance observations together with the LUT (generated in §3.1) were used to compute the likelihood term $P(y_T \mid x_T)$ in equation (20). After normalization, the posterior probability distribution of the LAI at the current time slice $T$ was obtained. The current LAI was then obtained from equation (8).

The DBN-based data fusion method was used to calculate the 2004–2006 LAIs for the G19 and G21 sites. These were then evaluated using the measured LAI and MODIS LAI data.

At the AmeriFlux sites, time-series reflectance (MOD09A1) data with a spatial resolution of 500 m were selected as observation data, so estimated results using such data are in the form of a $14 \times 14$-pixel image with a spatial resolution of 500 m in the footprint of a flux tower (7 km $\times$ 7 km). Because MODIS LAI products have a spatial resolution of 1 km, the estimated LAI using the DBN-based data fusion method were averaged through a $2 \times 2$ moving window in order to be consistent with MOD15A2 resolution.

Because field-measured LAI points were located at the central pixel of the footprint of a flux tower site, we picked out the corresponding pixel at each site to compare the estimated LAI using the DBN-based data fusion method, the MODIS LAI and the field-measured LAI (including the LAIs used for fitting the crop growth function and those set aside for validation) during 2004–2006 (figure 5). It is evident from figure 5 that the estimated LAI using DBN-based data fusion algorithm agree more closely with the measured LAI than do the MODIS LAI, at both the G19 and G21 sites. The MODIS LAI at the G19 site for 2004 (figure 5, G19_2004) fluctuated within the range of 1.2–2.0 between DOY 190 and DOY 270. This is not compatible with either the growth laws for soybeans or the LAI dynamics that the field-measured LAI revealed at this stage. Although figure 5 (G21_2004) shows that the MODIS LAI tended to agree with the growth laws at the G21 site in 2004, compared to the measured LAI, the MODIS LAI continued to underestimate the values. The same phenomenon is apparent for 2005 and 2006.

To identify this problem clearly, we selected four representative pixels from the footprints of two flux tower sites. Figure 6(a) and (b) shows the only two pixels at which the MODIS LAI was not underestimated superficially in the footprint of the G19 site in 2004. It was clear, however, that the curves connecting scattered points for the MODIS LAI were intermittent, with sudden falls at the DOY 209, all of which is inconsistent with the variety laws of LAI dynamics. According to the quality control (QC) flags provided by the MODIS LAI product, the quality of the LAI value at the DOY 209 was ‘good’. They were retrieved using the main algorithm indicating that the reflectance is not saturation and without clouds cover. On the contrary, the qualities of LAI around the DOY 209 (including DOYs 193, 201, 217, 225 and 233) were in fact relatively poor. In figure 6(a), for example, LAI values at DOYs 193 and 225 were estimated using a backup algorithm; the LAI at DOYs 201, 217 and 233 could not be retrieved, possibly because of bad L1B data and unusable MODIS daily aggregated surface reflectance product (MODAGAGG) data. In figure 6(c) and (d), the MODIS LAI results indicated ‘good’ qualities according to the QC flags throughout the whole growth stage, but the time-series tendencies of LAI dynamics on each pixel were relatively flat and exhibited almost no change over time, even during the period of rapid crop growth. This is especially evident in figure 6(d).
However, significant improvement was noticed after the introduction of crop growth information (figure 5). Estimated LAI values using the DBN-based data fusion algorithm at the G19 and G21 sites for 2004, 2005 and 2006 closely matched the measured LAI, and in addition, are continuous in time series and agree with the laws of LAI dynamics. As mentioned above, in each of the 3 years, five of the LAIs measured at the growth stage were set aside for validating the results estimated by the DBN-based data fusion method (shown as diamond symbols in figure 5). A comparison of the estimated LAI using DBN-based data fusion method and LAI used for validation (reference LAI) are shown in figure 7, and they are closely correlated (RMSE = 0.49, R = 0.96).

Figure 5. Estimated leaf area index (LAI) at the G19 and G21 sites in 2004–2006. DBN LAI represents LAI estimated using DBN-based data fusion method, MODIS LAI means LAI extracted from MODIS LAI product, Measured LAI is the part of measurements employed in the crop growth equation and Reference LAI is the field measured LAI left for the validation.

Note: DOY, day of year; MODIS, moderate resolution imaging spectroradiometer.
Figure 6. Representative pixels in the footprints of the G19 and G21 sites.
Note: RMSE, root mean square error; DBN, dynamic Bayesian network; MODIS, moderate resolution imaging spectroradiometer.

Figure 7. Results of validation of leaf area index (LAI) derived from dynamic Bayesian network-based data fusion method at the G19 and G21 sites in 2004–2006.
Note: RMSE, root mean square error.
5. Conclusions

Cross validation using measurements at the plot scale for 2004 and preliminary validation of estimated LAI over the two AmeriFlux sites from 2004 to 2006 showed a promising result. Estimated LAIs using only a physically based inversion algorithm and reflectance data are far from true values and do not agree with the variety law of crop growth. This may have been caused by data quality, such as the systematic offset of observed reflectance for simulations using a RTM, or by algorithms, such as the limitation of iterative optimization algorithms associated with the possibility of converging to a local minimum that is not necessarily close to the true solution at the plot scale. Although this limitation was overcome by basing the estimation of MODIS LAI from RTM inversion on an LUT approach, the MODIS LAI was nevertheless unsatisfactory at both tower sites. Underestimation was evident at both sites, and there was no obvious agreement with LAI dynamics, nor even evidence of any irregularity, during the period. Some primary validation programmes have indicated MODIS LAI to be overestimated by approximately 2–15% in the semi-arid grasslands of West Africa (Fensholt et al. 2004), and other researchers have indicated that the MODIS LAI algorithm markedly underestimated LAI values (down to 2–3 m² m⁻²) for the crop vegetation type (Yang et al. 2007). This study also observed that in the footprint of the flux tower the overwhelming majority of pixels of the MODIS LAI product varied within the range 0.8–2.5 and they were underestimated when compared to the ground-measured LAI. This can be partly explained by the effects of atmospheric cloud or aerosol, both of which decrease the value of the MODIS LAI in general. This explanation is also supported by the QC files of MOD15A2, which indicate that some pixels are affected by clouds, aerosol and so on. Another possible reason for the underestimation is that when the LUT algorithm fails, the triggered back-up algorithm of MODIS LAI product is semi-empirical. This, although perhaps applicable on a global scale, does not make allowance for errors caused by atmospheric correction, calibration etc. when applied to a specific region, thus leading to a low level of accuracy. Moreover, even if the QC flags of a pixel equalled 0, indicating that the quality of the corresponding MODIS LAI was supposed to be ‘good’, the estimated MODIS LAI values still did not match the true state. A possible reason is that the MODIS LAI is based on a biome-specific algorithm involving constants such as LAD, canopy heterogeneity, as well as soil and leaf optical properties which may not be appropriate for all vegetation anywhere for a certain biome type. Thanks to the information introduced by the crop growth function, the DBN-based data fusion method obtains reasonable LAI in both the magnitude and time trajectories. It not only improved the accuracy of LAI estimation but captured the features of continuity and evolitional information of land surface parameters when used in conjunction with a crop growth function (figures 3 and 5). The DBN-based data fusion method was demonstrated to be a promising way to retrieve time-series land surface parameters.

6. Discussion

The current LAI product is sometimes spatially and temporally discontinuous because of cloud cover, instrument problems and so on, which limits its application in biophysical models, land surface process simulation and global change research. For the purpose of improving the performance of remotely sensed LAI product, the DBN-based data fusion method is proposed in this article. It integrates remotely sensed observations, a crop growth function and a CR model within the framework of
DBN from the perspective of Bayesian probability, where the observations and model parameters are viewed as random variables. The proposed DBN-based data fusion method is an iterative algorithm at multiple time slices. Taking a certain time slice (e.g., $T$) as an example, the method requires the following five steps to be performed sequentially:

(i) Input the posterior probability distribution at time slice $T - 1$ as the initial information for the crop growth functions at time slice $T$.
(ii) Drive the crop growth functions and compute the probability of state transition time series.
(iii) Compute the likelihood probability of observed reflectance based on the CR model.
(iv) Fuse the results of (ii) and (iii) according to the Bayesian theorem.
(v) Transfer the output to the next time slice $T + 1$.

Using this method, LAI was obtained for the Xiaotangshan experimental field and the AmeriFlux tower sites. Results showed a good fit between the estimated LAI values ($RMSE = 0.35, R = 0.95$) and field-measured LAI values ($RMSE = 0.49, R = 0.96$). The proposed algorithm filled the temporal gaps and improved bad-quality values caused by contaminated remote sensing observations, or even missing remotely sensed data. Most of the improvement was attributed to the introduction of prior information of LAI dynamics. Through employing the temporal information of LAI dynamics, not only the accuracy of estimated LAI is improved but the phenomenon that curves of MODIS LAI may diverge seriously from the true state was corrected.

The crop growth functions used in this study were extended forms of the classical Verhulst logistic growth equation and were constructed through curve fitting. At the plot scale, only the growth stage factor was taken into account, and at the AmeriFlux sites, meteorological factors were incorporated to bridge the gap between the plot scale and regional scale and represent spatial heterogeneity. This facilitates the ability for LAI to be forecast on a larger scale. It should be noted that in a sophisticated crop growth model, soil state parameters, for example soil moisture, may be incorporated to facilitate simulation of the growing crop (Richter et al. 2008). Because most micrometeorological stations do not provide such parameters, crop growth simulation is carried out using an empirical statistic function in this study. The simplified formula means that some uncertainty arises for the simulated results, so a greater focus on investigating the operational data fusion method using a sophisticated crop growth model is required in the further work direction.

In this study, we have tested the generality of the DBN-based data fusion method at only two ecosystem stations and judged its performance against reference LAI values over the period 2004–2006. Results indicate that the method is feasible for improving the accuracy of land-surface parameter inversion using a combination of remote sensing observations and observations from ecological sites. By integrating observations from many more ecological sites, the proposed algorithm could be expanded to a regional scale. We believe that this approach has the ability to map LAI and provide LAI with a space-time continuum for the fields of environment, land process etc. on a regional scale. When the proposed method is applied to regional scales, two factors should be considered. To put it more specifically, (i) more observations relating meteorological factor to LAI are needed and (ii) spatial correlations of both meteorological factors and surface parameters should be taken into account comprehensively. Incorporating crop dynamic information and spatial
autocorrelation of pixels into the DBN-based data fusion procedure is one of the attractive topics that need to be further developed.

Acknowledgement
This work was supported by the National Basic Research Program of China (No. 2007CB714407), National High Technology Research and Development Program 863 (2009AA122101, 2008AA121103-1) and National Natural Science Foundation of China (No. 40601059).

References


Punska, O., 1999, Bayesian approaches to multi-sensor data fusion. MPhil thesis, Cambridge University, Cambridge, UK.


