Application of the Expectation Maximization Algorithm to Estimate Missing Values in Gaussian Bayesian Network Modeling for Forest Growth

Yaseen T. Mustafa, *Member, IEEE*, Valentyn A. Tolpekin, *Member, IEEE*, and Alfred Stein

***Abstract*—The leaf area index (LAI) is a biophysical variable related to atmosphere–biosphere exchange of CO_2. One way to obtain LAI value is by the Moderate Resolution Imaging Spectroradiometer (MODIS) biophysical products. In this paper, we use this product to improve the physiological principles predicting growth model within a Gaussian Bayesian network (GBN) setup. The MODIS time series, however, contains gaps caused by persistent clouds, cloud contamination, and other technique problems. We used the Expectation Maximization (EM) algorithm to estimate these missing values. During a period of 26 successive months, the EM algorithm is applied to four different cases: successively and not successively missing values during two different winter seasons, successively and not successively missing values during one spring season, and not successively missing values during the full study. Results show that the maximum value of the averaged absolute error between the original values and those estimated equals 0.16. This low value indicates that the estimated values well represent the original values. Moreover, the root mean square error of the GBN output reduces from 1.57 to 1.49 when performing the EM algorithm to estimate the not successively missing values. We conclude that the EM algorithm within a GBN can adequately handle missing MODIS LAI values and improves the estimation of the LAI.**

***Index Terms*—Expectation maximization (EM) algorithm, Gaussian Bayesian networks (GBNs), leaf area index (LAI), Moderate Resolution Imaging Spectroradiometer (MODIS).**

NOMENCLATURE

BN	Bayesian network.
FPAR	Fraction of absorbed photosynthetically active radiation.
GBN	Gaussian BN.
LAI	Leaf area index.
ML	Maximum likelihood.
NDVI	Normalized difference vegetation index.
PAR	Photosynthetically active radiation.
LAI_{BN}	LAI output of BN.
LAI_M	Derived LAI from NDVI Moderate Resolution Imaging Spectroradiometer (MODIS) product.
LAI_{3PG}	LAI output of physiological principles predicting growth (3-PG) model.
LAI_{FD}	LAI field measurement.
$RMSE_{LAI_{\mathrm{BN}}}$	Root mean square error (*RMSE*) of LAI_{BN} with respect to the LAI_{FD}.
$RE_{LAI_{\mathrm{BN}}}$	Relative error (RE) of LAI_{BN} with respect to the LAI_{FD}.
$RMSE_{LAI_M}$	*RMSE* of LAI_M with respect to the LAI_{FD}.
RE_{LAI_M}	RE of LAI_M with respect to the LAI_{FD}.
AAE_{LAI_M}	Averaged absolute error (AAE) of the estimating LAI_M with respect to the original LAI_M.
μ	Mean.
Σ	Covariance matrix.
σ	Standard deviation.
β_{ij}	Regression coefficients of BN output on its parents.
pa_i	Set of parents of node i.
$\#pa_i$	Number of the set of parents of node i.
ν_i	Conditional variance of node i (LAI_{BN}) given its parents.
θ	Set of parameters (μ, Σ).

I. INTRODUCTION

FORESTS play a major role in the carbon cycle on Earth, by means of CO_2 absorption during photosynthesis. This leads to sequestration of a fraction of CO_2 in the atmosphere [1], thus serving to reduce the CO_2 ratio in the atmosphere and most likely influencing the speed of climate change. For these and other reasons, observing and analyzing forest growth have received wide attention [2].

The LAI is a vegetation biophysical parameter that is widely used to observe forest growth. It is defined as the one-sided area of leaf tissue per unit ground surface area $(\mathrm{m}^2 \cdot \mathrm{m}^{-2})$. The LAI is strongly related to gas-vegetation exchange, modeled as photosynthesis, evaporation and transpiration, rainfall interception, and carbon flux [3]–[5]. Long-term monitoring of LAI can provide an understanding of dynamic changes in forest productivity and climate impacts on forest ecosystems.

Forest growth is commonly estimated using empirical models, for which LAI is an output parameter. Such a model, for example, is the physiological principles predicting growth (3-PG) model. It is a process-based forest growth model, developed by Landsberg and Waring [6] that we will use in this study. It requires parameters, site, and climate data as input and predicts

Manuscript received November 7, 2010; revised April 8, 2011 and June 15, 2011; accepted September 11, 2011. Date of publication October 25, 2011; date of current version April 18, 2012.

The authors are with the Department of Earth Observation Science, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, 7500 AE Enschede, The Netherlands (e-mail: mustafa@itc.nl; tolpekin@itc.nl; stein@itc.nl).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2011.2168823

the time course of stand development on a monthly basis. A full detailed description of this model is provided in [6].

Remote sensing of vegetation provides valuable information about the LAI. In particular, the Moderate Resolution Imaging Spectroradiometer (MODIS) on board Earth Observing System Terra/Aqua platforms provides LAI as a standard product (MOD15) at a 1-km resolution, every eight days [7]–[9].

Statistical methods have been used in the past to estimate and improve forest growth estimates. Patenaude *et al.* [10] calibrated the 3-PG model using remotely sensed data from a small forested region to simulate forest production. They used Bayesian calibration to reduce the uncertainty input. Also, graphical models have been developed in statistics to provide a flexible framework for specification and computation in complex systems [11]. A particular example of such models for estimating forest growth is the use of Bayesian networks (BNs) [12], [13]. A BN combines graphics and probabilities to express mutual relationships between variables. It uses a directed acyclic graph (DAG) to describe this relationship. Each node in the network represents a random variable, and the arc linking the nodes represents the relationship between variables. BNs have been widely applied in medical systems, computer science, image processing, artificial intelligence studies, and spatial data analysis [12], [14], [15]. Mustafa *et al.* [13] used a Gaussian Bayesian network (GBN) to improve LAI estimates by combining the 3-PG model output with MODIS images. The applicability of their approach depends on availability of satellite images and field data. Satellite images, however, often contain gaps (missing values) due to atmospheric circumstances, such as the presence of clouds, or to incomplete track spatial coverage.

For decades, researchers have relied on a variety of techniques to complete data by filling the missing values. These techniques are based on statistical or deterministic, and spatial or temporal approaches. Gao *et al.* [16] presented an algorithm to produce spatially and temporally continuous time series of global LAI fields based on MODIS data. They applied their method in North America and over specific locations containing deciduous broadleaf forest. A major breakthrough came in the 1970s with the advent of maximum likelihood (ML) estimation [17]–[19]. A common method to find ML for missing data is the expectation maximization (EM) algorithm [18]. Handling missing data using the EM algorithm within a GBN has not been fully explored and applied. In this study, the EM algorithm is formulated and applied to handle missing MODIS LAI values by estimating the missing parameters which are needed to execute the GBN approach in [13].

The objective of this work is twofold. First, we design an EM algorithm to be used within a GBN. Second, we estimate missing MODIS LAI values and check the performance of the GBN. The methodology is applied to the Speulderbos study area in The Netherlands.

II. Materials and Methods

A. Bayesian Network

A BN is a graphical model that denotes joint probabilistic distribution among variables of interest based on their probabilistic relationships. The BN is a DAG, where each node represents a variable for which the label either is given or has to be assessed. The relationships between the nodes are expressed in terms of conditional probabilities and shown as arrows between nodes. The arrow starts at the condition event and expresses the probability of the event at its head [20], [21].

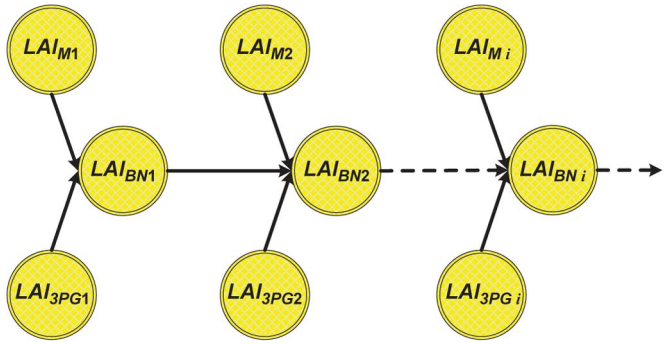

Fig. 1. BN for the first, second, and i^{th} time steps. Each time step consists of three nodes $LAI_{3\mathrm{PG}_i}$, LAI_{BN_i}, and LAI_{M_i} obtained from the 3-PG model, the BN, and a MODIS image, respectively.

BNs have been widely applied in different research domains. Mustafa *et al.* [13] designed a network to improve LAI estimation by combining LAI values derived from MODIS images and from the 3-PG model. Fig. 1 shows the graphical part of the BN. The intermediate node (LAI_{BN}) represents the estimated LAI values of the BN as the variable of interest.

Based on the continuous variation of LAI over time, we showed in [13] that we may assume that the LAI follows the Gaussian distribution. The most common type of a BN containing continuous variables is the GBN [13], [22]–[24]. A GBN is a BN where the joint probability distribution associated with its variables $\mathbf{LAI} = \{LAI_1, \ldots, LAI_n\}$ is the multivariate Gaussian distribution $N(\mu, \Sigma)$, given by

$$f(\mathbf{LAI}) = (2\pi)^{-n/2}|\Sigma|^{-1/2} \times \exp\left\{-\frac{1}{2}(\mathbf{LAI}-\mu)^T\Sigma^{-1}(\mathbf{LAI}-\mu)\right\}. \quad (1)$$

Here, μ is the n-dimensional mean vector, and Σ is the $n \times n$ positive definite covariance matrix with determinant $|\Sigma|$.

The conditional probability distribution of the LAI_i represented by the LAI_{BN_i} as the variable of interest given its parentage is the univariate Gaussian distribution with density

$$f\left(LAI_{\mathrm{BN}_i}|pa_i\right) \sim N\left(\mu_i + \sum_{j=1}^{\#pa_i}\beta_{ij}\left(pa_{ij}-\mu_{pa_{ij}}\right), \nu_i\right) \quad (2)$$

where μ_i is the expectation of LAI_{BN_i} at time i, β_{ij} represents the regression coefficients of LAI_{BN_i} on its parents, $\#pa_i$ is the number of parents of LAI_{BN_i}, and $\nu_i = \Sigma_i - \Sigma_{ipa_i}\Sigma_{pa_i}^{-1}\Sigma_{ipa_i}^T$ is the conditional variance of LAI_{BN_i} given its parents. Furthermore, Σ_i is the unconditional variance of the LAI_{BN_i}, Σ_{ipa_i} represents the covariances between LAI_{BN_i} and the variables pa_i, and Σ_{pa_i} is the covariance matrix of pa_i.

The regression coefficients relating LAI_{BN_i} to MODIS LAI (LAI_{M_i}), previous GBN output $(LAI_{\mathrm{BN}_{i-1}})$, and LAI output

of 3-PG model ($LAI_{3\mathrm{PG}_i}$) can be expressed as $\beta_{LAI_{\mathrm{BN}_i}LAI_{M_i}}$, $\beta_{LAI_{\mathrm{BN}_i}LAI_{\mathrm{BN}_{i-1}}}$, and $\beta_{LAI_{\mathrm{BN}_i}LAI_{3\mathrm{PG}_i}}$, respectively. Moreover, the expectations of LAI_{M_i}, $LAI_{\mathrm{BN}_{i-1}}$, and $LAI_{3\mathrm{PG}_i}$ are also expressed as $\mu_{LAI_{M_i}}$, $\mu_{LAI_{\mathrm{BN}_{i-1}}}$, and $\mu_{LAI_{3\mathrm{PG}_i}}$, respectively. Hence, for $i \geq 2$ and based on the graphical part of GBN in Fig. 1, the conditional distribution of LAI_{BN_i} equals

$$\begin{aligned} LAI_{\mathrm{BN}_i} &\sim N\left(\mu_{LAI_{\mathrm{BN}_i}} + \beta_{LAI_{\mathrm{BN}_i}LAI_{M_i}}\left(LAI_{M_i} - \mu_{LAI_{M_i}}\right)\right. \\ &\quad + \beta_{LAI_{\mathrm{BN}_i}LAI_{\mathrm{BN}_{i-1}}}\left(LAI_{\mathrm{BN}_{i-1}} - \mu_{LAI_{\mathrm{BN}_{i-1}}}\right) \\ &\quad \left. + \beta_{LAI_{\mathrm{BN}_i}LAI_{3\mathrm{PG}_i}}\left(LAI_{3\mathrm{PG}_i} - \mu_{LAI_{3\mathrm{PG}_i}}\right),\right. \\ &\quad \left.\Sigma_{LAI_{\mathrm{BN}_i}}\right). \end{aligned} \tag{3}$$

The posterior probability of LAI_{BN_i} in (3) is conditionally determined by three parents LAI_{M_i}, $LAI_{3\mathrm{PG}_i}$, and $LAI_{\mathrm{BN}_{i-1}}$. These parents are used as prior input information into the GBN, where every LAI_{M_i} consists of 16 pixels of LAI value derived from MODIS images with 250-m spatial resolution. Also, every $LAI_{3\mathrm{PG}_i}$ consists of 16 simulated LAI outputs from the 3-PG model, and every intermediate node LAI_{BN_i} consists of 16 LAI values. Based on the contribution of all of the MODIS images, the 3-PG output, and the previous GBN output, the prior information of the intermediate node is assumed and defined in [13] as

$$\begin{aligned} LAI_{\mathrm{BN}_i} = \rho(1-\tau)LAI_{M_i} + \tau LAI_{3\mathrm{PG}_i} \\ +(1-\rho)(1-\tau)LAI_{\mathrm{BN}_{i-1}} \end{aligned} \tag{4}$$

where τ and ρ are the weighing values, defined as $\tau = |(LAI_{M_i} - LAI_{M_{i-1}})/LAI_{M_{i-1}}|$ and $\rho = |(LAI_{3\mathrm{PG}_i} - LAI_{3PG_{i-1}})/LAI_{3PG_{i-1}}|$. They are proportional to the change in the LAI values obtained from the MODIS images and the 3-PG output. Moreover, this prior information has been tested on Gaussianity by Mustafa *et al.* [13], in terms of the temporal and the spatial resolutions of the input sources into the GBN. They used the Shapiro–Wilk and Lilliefors tests for that purpose. For more details about a GBN of improving forest growth estimates and its mathematical explanation, we refer to [13].

B. EM Algorithm for Estimating Missing Values in a GBN

The study aims to analyze a time series of satellite imageries. A common problem in such a series is the occurrence of incomplete observations (missing data). This is due to factors as persistent cloud cover, elevated aerosol loading, or sparse observations owing to insufficient repeat frequency of the satellite sensor. Mustafa *et al.* [13] used a time series of satellite images of two years to update estimated LAI values of the 3-PG model within a GBN. Their method requires a complete time series of satellite observations as an input source into a GBN.

Several methods have been applied and proposed to estimate missing values [17]–[19], [25]. A well-established technique for estimating parameters of statistical models from incomplete data is the EM algorithm. The origins of the EM algorithm date back to the 1970s [18]. Early EM applications primarily focused on estimating a mean vector and a covariance matrix with missing data. Since then, it has been extended to address a variety of incomplete-data estimation problems [25].

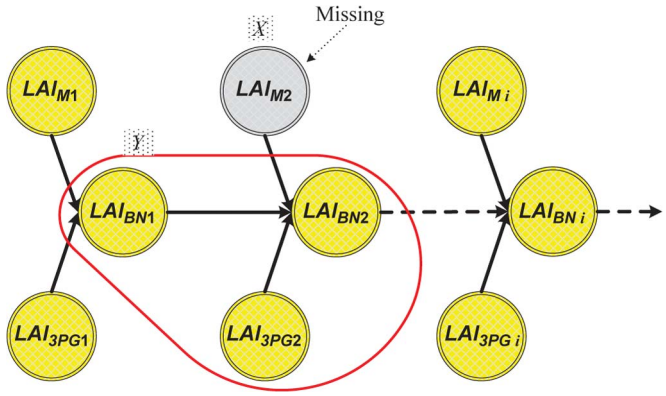

Fig. 2. BN with missing satellite observations. Y represents an observed data set consisting of three nodes LAI_{BN_2}, LAI_{BN_1}, and $LAI_{3\mathrm{PG}_2}$, whereas X represents the variable LAI_{M_2} for which an observation is missing.

The EM algorithm is broadly applicable for maximizing likelihoods and handling incomplete-data problems. The idea of the algorithm is to alternate between computing the expectation of the sample log-likelihood conditional on the previous parameter estimates (the E-step) and maximizing this expectation with respect to the desired parameters to obtain parameter estimates for the next recursion (the M-step). It was shown in [18] that this procedure is guaranteed to produce a monotonically increasing sequence of expected sample log-likelihoods which converges to a local maximum of the likelihood function. Thus, the EM algorithm is formulated and applied to handle the missing satellite data. To do so, it estimates parameters in a GBN implementation on LAI.

C. Derivation and Algorithmic Steps of the EM Algorithm

Consider missing data of satellite images at the i^{th} moment $(i > 1)$ within a GBN, as shown in Fig. 2. The GBN output LAI_{BN_i} conditionally depends on three nodes (variables), i.e., LAI_{M_i}, $LAI_{\mathrm{BN}_{i-1}}$, and $LAI_{3\mathrm{PG}_i}$. In this work, we consider LAI_{M_i} as a missing value.

Let (Y, X) be the complete data set at the i^{th} moment of GBN, with observed (complete) data $Y = \{LAI_{\mathrm{BN}_{i-1}}, LAI_{3\mathrm{PG}_i}, LAI_{\mathrm{BN}_i}\}$ and missing data $X = LAI_{M_i}$ (Fig. 2). For clarity, we abbreviate the GBN variables as $y = LAI_{\mathrm{BN}_i}$, $x = LAI_{M_i}$, $z = LAI_{\mathrm{BN}_{i-1}}$, and $w = LAI_{3\mathrm{PG}_i}$ such that the model parameters of the GBN are named μ_k, σ_k, and β_{yk} for $k \in \{x, y, z, w\}$.

Hence, (3) can be reformulated as

$$\begin{aligned} y \sim N\left(\mu_y + \beta_{yx}(x - \mu_x) + \beta_{yz}(z - \mu_z)\right. \\ \left. +\beta_{yw}(w - \mu_w), \sigma_y^2\right). \end{aligned} \tag{5}$$

The EM steps to find new ML estimates for the parameters $\theta = (\mu_x, \Sigma_x)$ are as follows.

1) Choose an initial setting for the parameters θ, and name it as $\theta^{\mathrm{old}} = (\mu^{\mathrm{old}}, \sigma^{\mathrm{old}})$. Initial parameter values are

guessed based on seasonal changes of LAI values obtained from MODIS observations

$$\theta^{\text{old}} = \begin{cases} \left(\mu_x - \left|\frac{\mu_{x_{i-2}} - \mu_{x_{i-1}}}{\mu_{x_{i-1}}}\right|, \right. \\ \quad \left. \sigma_x - \left|\frac{\sigma_{x_{i-2}} - \sigma_{x_{i-1}}}{\sigma_{x_{i-1}}}\right|\right) & \text{if } \mu_{x_{i-2}} \le \mu_{x_{i-1}} \\ \left(\mu_x + \left|\frac{\mu_{x_{i-2}} - \mu_{x_{i-1}}}{\mu_{x_{i-1}}}\right|, \right. \\ \quad \left. \sigma_x + \left|\frac{\sigma_{x_{i-2}} - \sigma_{x_{i-1}}}{\sigma_{x_{i-1}}}\right|\right) & \text{Otherwise} \end{cases} \quad (6)$$

where μ_x and σ_x are the mean and the standard deviation values of the MODIS LAI, respectively. They are obtained either for the period from September to February (the nongrowing season) or for the period from March to August (the growing season). The determination of the period for which we need to obtain the μ_x and σ_x is based on the occurrence of missing observations in that period.

In (6), the $|(\mu_{x_{i-2}} - \mu_{x_{i-1}})/\mu_{x_{i-1}}|$ and $|(\sigma_{x_{i-2}} - \sigma_{x_{i-1}})/\sigma_{x_{i-1}}|$ are the relative changes of the mean and the standard deviation of the previous two MODIS LAIs. Adding and subtracting these relative change values are based on the condition of increase or decrease the MODIS LAI values during the period of nongrowing or growing season.

2) E-step: Compute the expectation (with respect to the X data) of the likelihood function of the model parameters by including the missing variables as they were observed

$$\begin{aligned} Q(\theta, \theta^{\text{old}}) &= E_X\left[\log f(Y, X|\theta)|Y, \theta^{\text{old}}\right] \\ &= \int \log f(Y, X|\theta) f(X|Y, \theta^{\text{old}}) dX \\ &= \int \log f(x, y, z, w|\theta) f(x|y, z, w, \theta^{\text{old}}) dx \end{aligned} \quad (7)$$

where $\log f(x, y, z, w|\theta) = \log f(y|x, z, w, \theta) f(x|\theta) f(z|\theta) f(w|\theta)$ and $f(y|x, z, w, \theta)$ is the conditional distribution of y given its parents x, z, and w, such that $f(y|x, z, w, \theta) \sim N(\mu_y + \beta_{yx}(x - \mu_x) + \beta_{yz}(z - \mu_z) + \beta_{yw}(w - \mu_w), \sigma_y^2)$, $f(x|\theta) \sim N(\mu_x, \sigma_x^2)$, $f(z|\theta) \sim N(\mu_z, \sigma_z^2)$, and $f(w|\theta) \sim N(\mu_w, \sigma_w^2)$.

Therefore, the $\log f(x, y, z, w|\theta)$ can be expressed as in (8), shown at the bottom of the page.

Based on the graphical representation of the GBN model, the second factor of (7) can be found as

$$f(x|y, z, w, \theta^{\text{old}}) = \frac{f(x, y, z, w|\theta^{\text{old}})}{\int f(x, y, z, w|\theta^{\text{old}}) dx} \quad (9)$$

which can be written as $Q = \int_{-\infty}^{\infty} V(-fx^2 + gx - h) e^{-ax^2 + bx - c} dx$, where $V = \sqrt{\sigma_y^2 + \beta_{yx}^2 (\sigma^{\text{old}})^2}/(\sqrt{2\pi}\sqrt{(\sigma^{\text{old}})^2 \sigma_y^2})$, $f = (1/2)((1/\sigma_x^2) + (\beta_{yx}^2/\sigma_y^2))$, $g = (((y - \mu_y + \beta_{yx}\mu_x - \beta_{yz}(z - \mu_z) - \beta_{yw}(w - \mu_w))\beta_{yx}/\sigma_y^2) + (\mu_x/\sigma_x^2))$, $h = (1/2)(((y - \mu_y + \beta_{yx}\mu_x - \beta_{yz}(z - \mu_z) - \beta_{yw}(w - \mu_w))^2/\sigma_y^2) + ((z - \mu_z)^2/\sigma_z^2) + ((w - \mu_w)^2/\sigma_w^2) + (\mu_x^2/\sigma_x^2)) - \log(4\pi^2 \sigma_y \sigma_x \sigma_z \sigma_w)$, $a = (1/2)((1/(\sigma^{\text{old}})^2) + (\beta_{yx}^2/\sigma_y^2))$, $b = ((y - \mu_y + \beta_{yx}\mu^{\text{old}} - \beta_{yz}(z - \mu_z) - \beta_{yw}(w - \mu_w))\beta_{yx}/\sigma_y^2) + (\mu^{\text{old}}/(\sigma^{\text{old}})^2)$, and $c = b^2/4a$.

Here, μ^{old} and σ^{old} refer to the guessed mean and standard deviation values of x variable (LAI_M) obtained using (6). It can further be shown that the $Q(\theta, \theta^{\text{old}})$ equals

$$Q(\theta, \theta^{\text{old}}) = \Omega\left(-\frac{f}{2a} + \frac{bg}{2a} - \frac{fb^2}{(2a)^2} - h\right) \quad (10)$$

where $\Omega = V\sqrt{\pi/a} e^{(-c + (b^2/4a))}$.

3) M-step: Compute the ML estimates of the parameters θ by maximizing the expected likelihood obtained during the E-step, i.e., $\theta^{\text{new}} = \arg\max_{\theta} Q(\theta, \theta^{\text{old}})$.

Hence, by differentiating $Q(\theta, \theta^{\text{old}})$ with respect to θ and solving the equations for $\theta^{\text{new}} = (\mu_x^{\text{new}}, \sigma_x^{\text{new}})$, the maximum values are found as

$$\mu_x^{\text{new}} = \frac{\sqrt[3]{\phi + \Psi}}{6\lambda} + \frac{2(-3\delta\lambda + \alpha^2)}{3\lambda\sqrt[3]{\phi + \Psi}} + \frac{\alpha}{3\lambda} \quad (11)$$

$$\sigma_x^{\text{new}} = \sqrt{(\mu_x^{\text{new}})^2 - \frac{2C}{B}\mu_x^{\text{new}} + \frac{E}{B}}. \quad (12)$$

Here

$$\phi = -36\delta\alpha\lambda + 108\eta\lambda^2 + 8\alpha^3$$

$$\Psi = 12\sqrt{3}\sqrt{4\delta^3\lambda - \delta^2\alpha^2 - 18\delta\alpha\lambda\eta + 27\eta^2\lambda^2 + 4\eta\alpha^3}\lambda$$

$$\lambda = AB$$

$$\alpha = 2AC + DB$$

$$\delta = AE + B^2 + 2DC$$

$$\eta = CB + dE$$

$$\begin{aligned} \log f(x, y, z, w|\theta) = &-\frac{1}{2}\left(\frac{1}{\sigma_x^2} + \frac{\beta_{yx}^2}{\sigma_y^2}\right)x^2 + \left(\frac{(y - \mu_y + \beta_{yx}\mu_x - \beta_{yz}(z - \mu_z) - \beta_{yw}(w - \mu_w))\,\beta_{yx}}{\sigma_y^2} + \frac{\mu_x}{\sigma_x^2}\right)x \\ &-\frac{1}{2}\left(\frac{(y - \mu_y + \beta_{yx}\mu_x - \beta_{yz}(z - \mu_z) - \beta_{yw}(w - \mu_w))^2}{\sigma_y^2} + \frac{(z - \mu_z)^2}{\sigma_z^2} + \frac{(w - \mu_w)^2}{\sigma_w^2} + \frac{\mu_x^2}{\sigma_x^2}\right) - \log(4\pi^2\sigma_y\sigma_x\sigma_z\sigma_w) \end{aligned} \quad (8)$$

where

$$A = 4\frac{a^2\Omega\beta_{yx}^2}{\sigma_y^2}$$

$$B = 4a^2$$

$$C = 2ab\Omega$$

$$D = -4\frac{a^2\Omega\left(y-\mu_y-\beta_{yz}(z-\mu_z)-\beta_{yw}(w-\mu_w)\right)\beta_{yx}}{\sigma_y^2} + 2\frac{ab\Omega\beta_{yx}^2}{\sigma_y^2}$$

$$E = 2a\Omega + b^2\Omega.$$

4) Using the iterative EM algorithm requires to check the convergence of the θ^{new} values. If $|\theta^{\text{new}} - \theta^{\text{old}}| > \varepsilon$ for a small value ε, then $\theta^{\text{old}} \leftarrow \theta^{\text{new}}$, and the algorithm returns to step 2 (Algorithm 1).

Algorithm 1: EM algorithm in GBN

Data: LAI_{BN_i}, $LAI_{\text{BN}_{i-1}}$, LAI_{3PG_i}, and observed LAI_{M_i} values
Result: Estimated parameter $\theta^{\text{new}} = (\mu^{\text{new}}, \sigma^{\text{new}})$ of the missing LAI_{M_i}
Initial set
Choose initial θ^{old} values as in the expression (6);
while $|\theta^{\text{new}} - \theta^{\text{old}}| > \varepsilon$ **do**
E-Step
Compute $\log f(Y, X|\theta)$ using (8);
Compute $f(X|Y, \theta^{\text{old}})$ using (9);
Compute $Q(\theta, \theta^{\text{old}})$ using (7);
M-Step
Formulate $Q(\theta, \theta^{\text{old}})$ with respect to the θ;
Solve $Q(\theta, \theta^{\text{old}})$ to find θ^{new};
Find a solution which maximizes $Q(\theta, \theta^{\text{old}})$ to get (11) and (12);
end

The convergence criterion ε has been set in this study equal to 10^{-5} after examining the EM algorithm for several cases of missing LAI_M values. It ensures that the estimated θ differs by less than 0.1%. Fig. 3 shows the changes of parameter estimates and the differences in those during consecutive iterations for one run of the EM algorithm.

III. Application

The EM algorithm, as described in Section II, is applied to the Speulderbos forest in The Netherlands where the LAI is observed and is available as a time series from July 2007 to September 2009. A brief description of the study area and data used is given hereinafter.

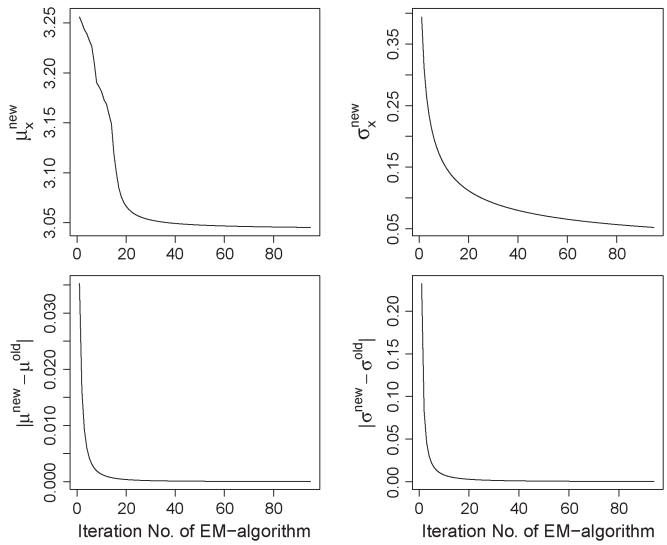

Fig. 3. Estimated parameters $(\mu_x^{\text{new}}, \sigma_x^{\text{new}})$ and the differences of two consecutive iterations in the EM algorithm. This is an example of a missing LAI_M value at day 289 of 2008.

A. Study Area

The Speulderbos forest reserves are located at 52°15′08″ N, 05°41′25″ E, covering 2390 ha, near the village in Garderen, The Netherlands. A climate station is placed within a dense 2.5-ha Douglas fir (*Pseudotsuga menziesii*) stand planted in 1962. The tree density is 785 trees ha^{-1}, and the tree height in 2006 was approximately 32 m. Dominant species in the neighborhood of the Douglas fir stand are the Japanese larch (*Larix kaempferi*), beech (*Fagus sylvatica*), Scotch pine (*Pinus sylvestris*), and hemlock (*Tsuga spp*) [26]. The single-sided LAI varies between 8 and 11 throughout the year [27]. The stand is surrounded by a larger forested area of approximately 50 km^2; the nearest edge is at a distance of 1.5 km southeast from the site. The study area of this work has been assumed a homogeneous area of 1 km^2 around the climate station in the Speulderbos forest as it has been considered in [13].

B. Data Description

The data used in this study are the MODIS Terra land satellite collection 5 data and ground measurements.

1) Ground Data: Ground data are collected at the observation tower of the Speulderbos forest, which is equipped with a weather station and various scientific instruments. The required data for the 3-PG model are the 16-day mean of climate data, site factors, initial conditions, and 3-PG parameters. The parameter values used in this work as well as in [13] are Douglas fir parameters. The values of these parameters are obtained from [28] and [29].

LAI was measured at the ground (LAI_{FD}) to validate the estimated LAI_{BN} values. LAI is indirectly measured from canopy transmission by inversion of measurements of photosynthetically active radiation (PAR) above and below the canopy. The PAR data are acquired from the Speulderbos forest using sensors placed at the tower above and below the canopy. PAR data are recorded every minute during daytime. Calculation of

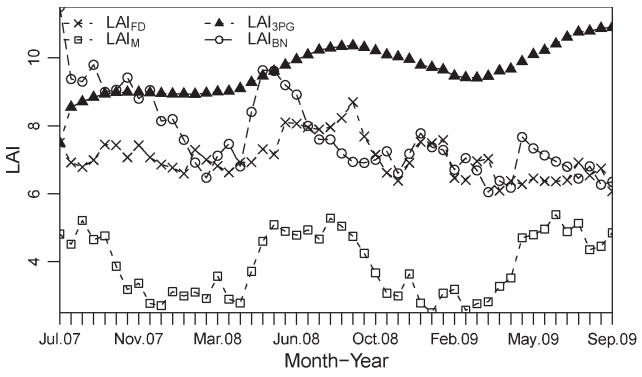

Fig. 4. LAI values of the Speulderbos forest obtained from four sources: (LAI_{FD}) field data, $(LAI_{3\mathrm{PG}})$ the 3-PG model, (LAI_M) MODIS images, and (LAI_{BN}) GBN, during the period from July 2007 to September 2009.

LAI from PAR data is based on the relationship between leaf area and light transmittance, described by the Beer–Lambert model [30]. The daily average LAI_{FD} has been computed from the recorded PAR data after calculating LAI at every minute of solar noon. The LAI_{FD} values are adjusted by using a clumping factor. For more calculation details of LAI_{FD}, we refer the reader to [13].

2) MODIS Data: MODIS product MOD15A2 provides LAI values at a 1-km spatial resolution, every eight days [7]–[9]. In this study, LAI values are derived from the normalized difference vegetation index (NDVI) MODIS product (MOD13Q1) every 16 days at a 250-m spatial resolution, thus matching the temporal frequency of the 3-PG model and ground data. This is processed by establishing the relation between the NDVI and the FPAR. A linear relationship between the NDVI and the FPAR is present with $R^2 = 0.71$ [13]. The LAI_M of 250 m is calculated as $LAI = -\ln(1 - \mathrm{FPAR})/k$, where FPAR is the fraction of absorbed PAR at 250-m spatial resolution derived from NDVI of 250 m based on the relationship equation between them and k is the extinction coefficient. For more details of LAI_M, we refer to [13].

C. Implementation

The time period contains two winter seasons (October–March; 2008 and 2009), two summer seasons (May–August; 2008 and 2009), and two spring seasons (end of March–middle of May; 2008 and 2009) (Fig. 4). We have a full set of observed MODIS LAI images during the monitoring period. To implement the approach of this work, synthetic gaps are introduced. This is done by removing some of the MODIS LAI observations successively and not successively, following the common occurrence of missing satellite observations as explained hereinafter. The missing values are estimated using the EM algorithm and saved as a test set to compare with the original MODIS LAI.

Missing satellite imageries mainly occur during the winter season, due to the atmospheric conditions such as the presence of clouds and of aerosols. Moreover, satellite images may not be available in other seasons due to the incomplete track spatial coverage. The EM algorithm is applied to estimate missing MODIS LAI in four cases. The first case considers successively and not successively missing MODIS LAI values during the winter of 2008. The second case similarly considers the winter of 2009. The third case considers successively and not successively missing MODIS LAI values during the spring of 2009. The fourth case concerns not successively missing MODIS LAI values during the period from July 2007 to September 2009.

In each winter season, the EM algorithm is carried out to estimate successively missing LAI_M values during the first half (first two and a half months), the second half (second two and a half months), and the whole time period of each season. Moreover, the EM algorithm is applied to estimate not successively missing LAI_M values during each season. The performance of the GBN is assessed when performing the EM algorithm. The method was implemented in C++ code. For clarity, the figures only include the LAI_M, LAI_{BN}, and LAI_{FD} as the values of interest.

The accuracy of LAI_{BN} is assessed using the *RMSE* $(RMSE_{LAI_{\mathrm{BN}}})$ and the RE $(RE_{LAI_{\mathrm{BN}}})$ with respect to the LAI_{FD} before and after performing the EM algorithm. The accuracy of LAI_M is assessed as well using the $RMSE_{LAI_M}$ and RE_{LAI_M} with respect to the LAI_{FD} before and after performing the EM algorithm. The AAE (AAE_{LAI_M}) of the estimating LAI_M with respect to the original LAI_M is also calculated. The *RMSE*, RE, and AAE are computed as follows: $RMSE_{LAI_{\mathrm{BN}}} = \sqrt{\sum_{i=1}^{n}(LAI_{\mathrm{BN}_i} - LAI_{FD_i})^2/n}$, $RMSE_{LAI_M} = \sqrt{\sum_{i=1}^{n}(LAI_{M_i} - LAI_{FD_i})^2/n}$, $RE_{LAI_{\mathrm{BN}}} = |(LAI_{\mathrm{BN}} - LAI_{\mathrm{FD}})/LAI_{\mathrm{FD}}|\ \ 100\%$, $RE_{LAI_M} = |(LAI_M - LAI_{\mathrm{FD}})/LAI_{\mathrm{FD}}|\ \ 100\%$, and $AAE_{LAI_M} = (1/n)\sum_{i=1}^{n}|\ Origin\ LAI_{M_i} - Estimated\ LAI_{M_i}|$.

IV. Experimental Results

A. Estimating LAI_M During the Winter of 2008

Estimation of missing LAI_M values is carried out during the first half, the second half, and the whole time period of the season [Fig. 5(a)–(d)]. Estimated values are close to the original LAI_M with AAE_{LAI_M} values less than 0.1. The deviation between the LAI_{BN} values after estimating the missing LAI_M values and the LAI_{FD} becomes significantly lower than the deviation between the LAI_{BN} values with the original LAI_M and the LAI_{FD}. Fig. 5(a) shows LAI values after performing the EM algorithm of estimating five successively missing LAI_M values during the first half of the season. We found an $RMSE_{LAI_{\mathrm{BN}}}$ of 1.53 and an $RE_{LAI_{\mathrm{BN}}}$ of 13.2%. Fig. 5(b) shows five successively missing LAI_M values estimated during the second half of the season. The $RMSE_{LAI_{\mathrm{BN}}}$ and the $RE_{LAI_{\mathrm{BN}}}$ in this case are 1.55 and 14.2%, respectively, whereas in Fig. 5(d), five not successively missing LAI_M values during the whole time period of the season are estimated with an $RMSE_{LAI_{\mathrm{BN}}}$ of 1.51 and an $RE_{LAI_{\mathrm{BN}}}$ of 13.3%. The deviation between LAI_{BN} and LAI_{FD} increases, however, after performing the EM algorithm [Fig. 5(c)].

The eight successively missing LAI_M values during the whole time period of the season are estimated using the EM algorithm. The $RMSE_{LAI_{\mathrm{BN}}}$ and the $RE_{LAI_{\mathrm{BN}}}$ after and before performing the EM algorithm equal 1.68 against 1.57 and 17.6% against 14.7%, respectively.

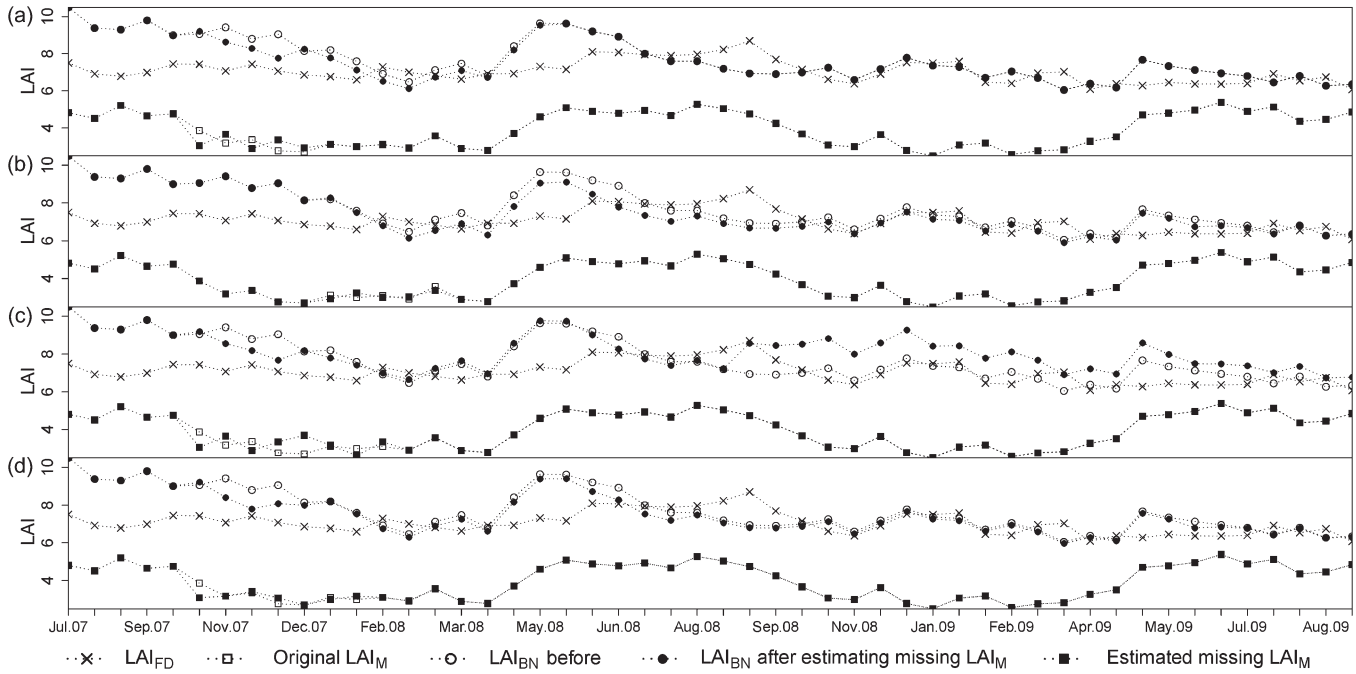

Fig. 5. LAI_{BN} and LAI_M values of the Speulderbos forest obtained before and after performing the EM algorithm during the winter of 2008. (a) Five successively missing LAI_M values during the first half of the season. (b) Five successively missing LAI_M values during the second half of the season. (c) Eight successively missing LAI_M values during the full season. (d) Five not successively missing LAI_M values during the full season.

TABLE I
RMSE AND THE RE OF LAI_M AND LAI_{BN}, AND THE AAE OF LAI_M. THEY ARE OBTAINED BEFORE AND AFTER APPLYING THE EM ALGORITHM DURING THE WINTER OF 2008

Cases	$RMSE_{LAI_M}$	$RMSE_{LAI_{BN}}$	RE_{LAI_M}	$RE_{LAI_{BN}}$	AAE_{LAI_M}
without missing (original LAI_M)	3.26	1.57	44.1%	14.7%	
Five successively missing LAI_M estimated during the first half	3.27	1.53	44.0%	13.2%	0.05
Five successively missing LAI_M estimated during the second half	3.26	1.55	44.0%	14.2%	0.02
Eight successively missing LAI_M estimated during the whole season period	3.23	1.68	43.6%	17.6%	0.16
Five not successively missing LAI_M estimated during the whole season period	3.27	1.51	44.1%	13.3%	0.02

Estimating LAI_M using the EM algorithm represents the original LAI_M particularly in the not successively missing case with an AAE_{LAI_M} of 0.02 (Table I). This indicates that the missing LAI_M is estimated successfully.

B. Estimating LAI_M During the Winter of 2009

After applying the EM algorithm to estimate missing MODIS LAI at the same time steps in Section IV-A, we notice that the LAI_{BN} is close to the LAI_{FD} [Fig. 6(a), (b), and (d)]. Fig. 6(a) shows LAI_{BN} and LAI_M values after estimating five successively missing LAI_M values during the first half of the season. The $RMSE_{LAI_{\mathrm{BN}}}$ equals 1.59, and the $RE_{LAI_{\mathrm{BN}}}$ equals 14.7%. Furthermore, the five successively missing LAI_M values during the second half of the season are shown in Fig. 6(b). The $RMSE_{LAI_{\mathrm{BN}}}$ and the $RE_{LAI_{\mathrm{BN}}}$ in this case are 1.58 and 14.6%, respectively. Five not successively missing LAI_M values during the whole time period of the season are estimated with an $RMSE_{LAI_{\mathrm{BN}}}$ of 1.51 and an $RE_{LAI_{\mathrm{BN}}}$ of 14.4% [Fig. 6(d)]. Moreover, the estimated missing LAI_M values are close to the original LAI_M values with $AAE_{LAI_{\mathrm{BN}}}$ values below 0.08.

The strong deviation between LAI_{BN} and LAI_{FD} has occurred after applying the EM algorithm to estimate eight successively missing LAI_M values during the whole time period of the season [Fig. 6(c)]. The $RMSE_{LAI_{\mathrm{BN}}}$ and $RE_{LAI_{\mathrm{BN}}}$ after and before performing the EM algorithm equal 1.69 against 1.57 and 17.0% against 14.4%, respectively. The quantitative results are shown in Table II.

C. Estimating LAI_M During the Spring of 2009

Missing LAI_M values are estimated during the second spring season only (the spring of 2009) as an extra test of the EM algorithm. Fig. 7(a) shows LAI values of estimating five successively missing LAI_M values during the season. There is a noticeable deviation with AAE_{LAI_M} of 0.09 between the estimated missing LAI_M and original LAI_M. Moreover, the LAI_{BN} deviates little from LAI_{FD} after performing the EM algorithm than before applying the EM algorithm, but it remains close to the LAI_{FD}. We found an $RMSE_{LAI_{\mathrm{BN}}}$ of 1.62 and an $RE_{LAI_{\mathrm{BN}}}$ of 15.8%.

The EM algorithm is applied to estimate three not successively missing LAI_M values during the season as well [Fig. 7(b)]. At this case, a small deviation between LAI_{BN} and LAI_{FD} is observed with an $RMSE_{LAI_{\mathrm{BN}}}$, an $RE_{LAI_{\mathrm{BN}}}$, and an AAE_{LAI_M} equal to 1.56, 14.5%, and 0.06, respectively (Table III).

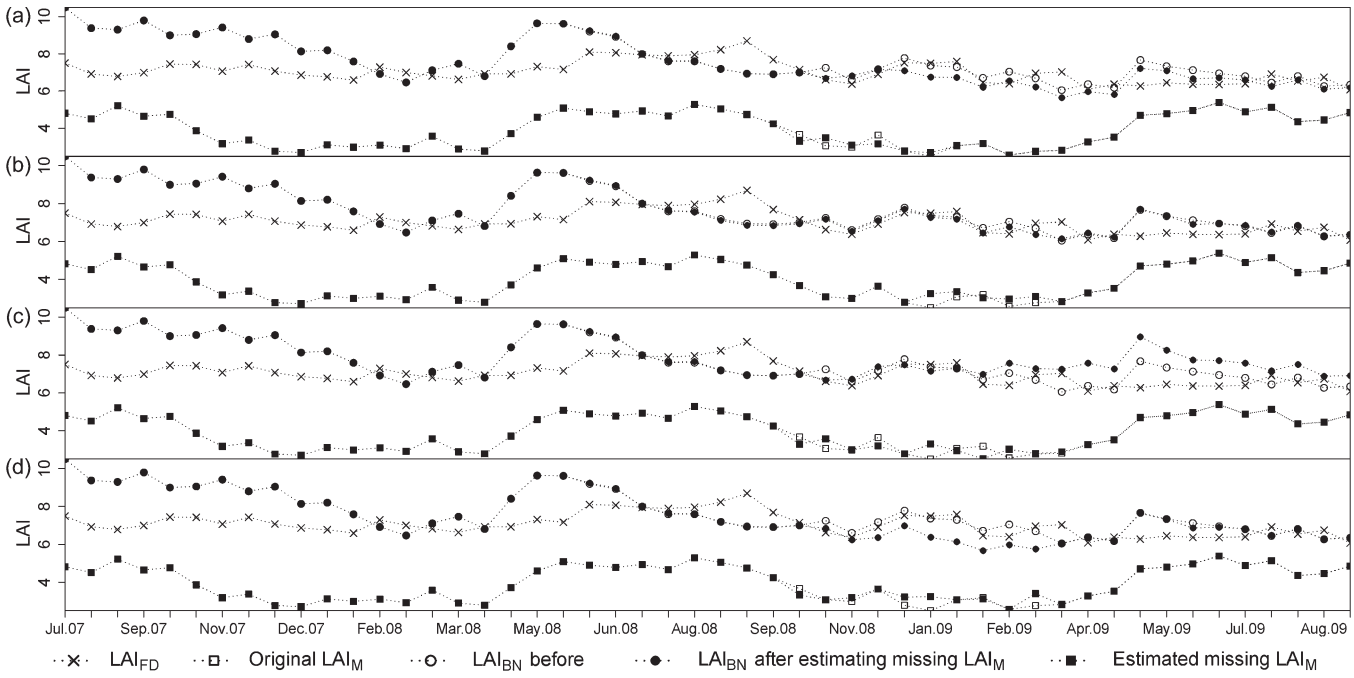

Fig. 6. LAI_{BN} and LAI_M values of the Speulderbos forest obtained before and after performing the EM algorithm during the winter of 2009. (a) Five successively missing LAI_M values during the first half of the season. (b) Five successively missing LAI_M values during second half of the season. (c) Eight successively missing LAI_M values during the whole time period of the season. (d) Five not successively missing LAI_M values during the whole time period of the season.

TABLE II
RMSE AND THE RE OF LAI_M AND LAI_{BN}, AND THE AAE OF LAI_M. THEY ARE OBTAINED BEFORE AND AFTER APPLYING THE EM ALGORITHM DURING THE WINTER OF 2009

Cases	$RMSE_{LAI_M}$	$RMSE_{LAI_{BN}}$	RE_{LAI_M}	$RE_{LAI_{BN}}$	AAE_{LAI_M}
without missing (original LAI_M)	3.26	1.57	44.1%	14.7%	
Five successively missing during LAI_M estimated the first half	3.25	1.59	44.0%	14.7%	0.04
Five successively missing LAI_M estimated during the second half	3.23	1.58	43.6%	14.6%	0.03
Eight successively missing LAI_M estimated during the whole season period	3.24	1.69	43.8%	17.0%	0.08
Five not successively missing LAI_M estimated during the whole season period	3.22	1.51	43.6%	14.4%	0.04

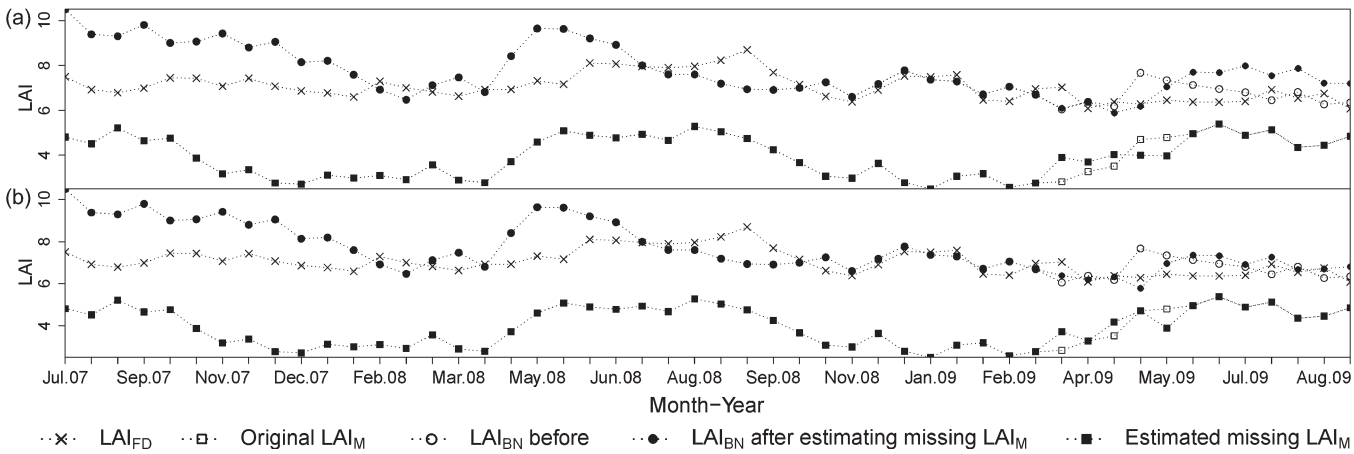

Fig. 7. LAI_{BN} and LAI_M values of the Speulderbos forest obtained before and after performing the EM algorithm during the spring of 2009. (a) Five successively missing LAI_M values during the season. (b) Three not successively missing LAI_M values during the full period of the season.

D. Estimating LAI_M During the Full Period

Finally, the EM algorithm is carried out to estimate the missing LAI_M during the whole time study period (16 not successively missing LAI_M values). Performance of the GBN is assessed during the test.

The LAI_{BN} after performing the EM algorithm for estimating missing LAI_M values remains close to the LAI_{FD}. The deviation between LAI_{BN} and LAI_{FD} reduces after applying the EM algorithm [Fig. 8]. We found an $RMSE_{LAI_{\mathrm{BN}}}$ value of 1.49 against 1.57 and an $RE_{LAI_{\mathrm{BN}}}$ value of 14.0% against 14.7%. The $RMSE_{LAI_M}$ equals 3.27, and the RE_{LAI_M} equals

TABLE III
RMSE AND THE RE OF LAI_M AND LAI_{BN}, AND THE AAE OF LAI_M. THEY ARE OBTAINED BEFORE AND AFTER APPLYING THE EM ALGORITHM DURING THE SPRING OF 2009

Cases	$RMSE_{LAI_M}$	$RMSE_{LAI_{BN}}$	RE_{LAI_M}	$RE_{LAI_{BN}}$	AAE_{LAI_M}
without missing (original LAI_M)	3.26	1.57	44.1%	14.7%	
Five successively missing during LAI_M estimated	3.25	1.62	44.2%	15.8%	0.09
Three not successively missing LAI_M estimated during the whole season period	3.25	1.56	43.8%	14.5%	0.06

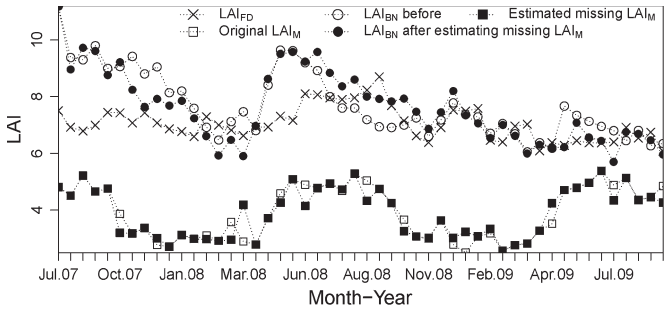

Fig. 8. LAI_{BN} and LAI_M values at the Speulderbos forest obtained before and after performing the EM algorithm for not successively missing LAI_M values during the period from July 2007 to September 2009.

44.4%. Moreover, the AAE_{LAI_M} resulting after comparing the estimated values with the original values equals 0.16. This value is low with respect to the 16 missing LAI_M values, thus confirming the success of formulating and implementing the EM algorithm within a GBN to estimate the missing LAI_M and carrying out the GBN approach.

V. DISCUSSION

In this study, the EM algorithm is formulated within a GBN, and missing LAI_M values are estimated.

Our results show that the missing LAI_M values are estimated successfully, i.e., such that they well represent the original MODIS LAI trend. The GBN output with estimated missing LAI_M values remains closer to the LAI field data than the LAI output of the 3-PG model. Overall, the deviation between the GBN output and the LAI field data after estimating missing LAI_M is lower than the deviation between the 3-PG model and the LAI field data. Moreover, a smaller deviation occurred between GBN output and LAI field data in the case of estimating not successively missing LAI_M during the time study (Section IV-D). This emphasizes that the EM algorithm within a GBN serves as a good approach to estimate missing MODIS LAI values.

The strength of the represented work lies in applying the EM algorithm in a GBN to estimate the missing input source LAI_M of the GBN. Analytically, the most difficult portion of the EM algorithm is the E-step. The reason for this is that the expectation must be computed over all LAI values. This may slow down the convergence of the EM algorithm in some cases, requiring up to 800 iterations. Although other algorithms offer faster computation [19], [25], [31], we selected the EM algorithm because it guarantees convergence.

A well-known drawback of the EM algorithm is that the convergence can be quite slow [25]. In order to save computing time, it is essential to start with good initial parameters. For this reason, we re-sorted (6) such that we can identify the initial values to be close to the estimated LAI_M values. The initial guess in (6) has been formulated based on the seasonal change of LAI values that are derived from MODIS. However, the relative changes of LAI_M are introduced in order to account the LAI change within a relatively short period of time which may show a change in the LAI trend.

From the results of performing the EM algorithm to estimate the missing LAI_M, we observed that even a small difference between the estimated missing LAI_M and the original LAI_M has an impact on the resulting output of the GBN. This is due to the fact that a GBN is sensitive to LAI_M variation and gives more weight to LAI_M (4), as proposed in [13], since GBN needs prior information (data) to propagate the network and to get posterior probability (GBN output). Expression (4) is re-sorted [13] to include prior information of the LAI_{BN} variable at each moment. Expression (4) has weighing parameters τ and ρ which rely on available information of LAI_M and LAI_{3PG}, respectively. Hence, an increase of the $RMSE_{LAI_M}$ (Table I) means that the τ value in (4) decreases and leads to a decrease of the prior value of the intermediate variable LAI_{BN} to be used in the next iteration (observation). This results in a decrease of the posterior probability of LAI_{BN} and a decrease of the differences between LAI_{BN} and LAI_{FD}. This only happens in the case of successively missing LAI_M for which the values of τ are calculated based on two or more estimations of successively missing LAI_M values. This does not happen for not successively missing LAI_M; since then, the τ value decreases because of the decreasing LAI_M, resulting in a decrease of the prior value of LAI_{BN} for the next iteration (Tables II and III). The reason is that values of τ are calculated based on the original LAI_M rather than on those estimated by the EM algorithm, hence resulting in a decrease of GBN output.

The performance of the LAI_{BN} is better in the case of estimating not successively missing LAI_M than in the case of estimating successively missing LAI_M. The poorest LAI_{BN} performance was in the case of estimating successively missing images. Moreover, a noticeable deviation of LAI_{BN} is seen after applying the EM algorithm during the spring season with successively and not successively missing LAI_M values. The reason is that the spring season is the most critical time for forest growth, when the LAI values start changing and increasing. Overall, the results of this study show that the applicability of the EM algorithm within a GBN is relatively limited for a long (more than five observations) series of successively missing LAI_M values.

Some issues require further work. For instance, other methods could be included into the EM algorithm to accelerate the speed of convergence [31]. Such improvements are, e.g., based on the application to the EM algorithm of optimization theory techniques such as Aitkin's acceleration [32]. These methods require some further preparatory analytical work and

may increase the complexity of the implementation. Finally, ground observations LAI_{FD} may be partly or entirely absent as some areas are difficult to reach or the instruments for field survey are too expensive for the national forest survey institutes. The EM algorithm within a GBN, however, might be useful to deal with this kind of problem. This could be achieved by modifying the network such as including another node (variable) that represents the ground data and applying the EM algorithm to the added variable.

VI. Conclusion

This study formulated the EM algorithm within a GBN to estimate missing MODIS LAI values. Performance of the network is evaluated by comparing its output with LAI field measurements before and after performing the EM algorithm. The approach is examined for four cases: successively and not successively missing LAI_M during various time periods of the winter of 2008, successively and not successively missing LAI_M during various time periods of the winter of 2009, successively and not successively missing LAI_M during the spring of 2009, and not successively missing LAI_M during the full study period. We conclude that the missing LAI_M values are estimated successfully using the EM algorithm with 0.16 of the maximum value of the AAE between the original values and those estimated. The presence of more than five successively missing LAI_M values influences the output of the network such that LAI_{BN} does not match the LAI_{FD} but that it approximates. We conclude, in particular, that the LAI_{BN} improves after performing the EM algorithm in the case of estimating not successively missing LAI_M values during the full study period.

Acknowledgment

The authors would like to thank anonymous reviewers for their insightful comments and suggestions.

References

[1] G. W. W. Wamelink, H. J. J. Wieggers, G. J. Reinds, J. Kros, J. P. Mol-Dijkstra, M. van Oijen, and W. de Vries, "Modelling impacts of changes in carbon dioxide concentration, climate and nitrogen deposition on carbon sequestration by European forests and forest soils," *Forest Ecol. Manage.*, vol. 258, no. 8, pp. 1794–1805, Sep. 2009.

[2] G. B. Bonan, "Importance of leaf-area index and forest type when estimating photosynthesis in boreal forests," *Remote Sens. Environ.*, vol. 43, no. 3, pp. 303–314, Mar. 1993.

[3] H. A. Cleugh, R. Leuning, Q. Mu, and S. W. Running, "Regional evaporation estimates from flux tower and MODIS satellite data," *Remote Sens. Environ.*, vol. 106, no. 3, pp. 285–304, Feb. 2007.

[4] J. Dietz, D. Holscher, C. Leuschner, and Hendrayanto, "Rainfall partitioning in relation to forest structure in differently managed montane forest stands in Central Sulawesi, Indonesia," *Forest Ecol. Manage.*, vol. 237, no. 1–3, pp. 170–178, Dec. 2006.

[5] R. Leuning, H. A. Cleugh, S. J. Zegelin, and D. Hughes, "Carbon and water fluxes over a temperate *Eucalyptus* forest and a tropical wet/dry savanna in Australia: Measurements and comparison with MODIS remote sensing estimates," *Agriculture Forest Meteorol.*, vol. 129, no. 3/4, pp. 151–173, Apr. 2005.

[6] J. J. Landsberg and R. H. Waring, "A generalised model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning," *Forest Ecol. Manag.*, vol. 95, no. 3, pp. 209–228, Aug. 1997.

[7] W. Z. Yang, D. Huang, B. Tan, J. C. Stroeve, N. V. Shabanov, Y. Knyazikhin, R. R. Nemani, and R. B. Myneni, "Analysis of leaf area index and fraction of PAR absorbed by vegetation products from the Terra MODIS sensor: 2000–2005," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 7, pp. 1829–1842, Jul. 2006.

[8] W. Z. Yang, B. Tan, D. Huang, M. Rautiainen, N. V. Shabanov, Y. Wang, J. L. Privette, K. F. Huemmrich, R. Fensholt, I. Sandholt, M. Weiss, D. E. Ahl, S. T. Gower, R. R. Nemani, Y. Knyazikhin, and R. B. Myneni, "MODIS leaf area index products: From validation to algorithm improvement," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 7, pp. 1885–1898, Jul. 2006.

[9] P. Yang, R. Shibasaki, W. Wu, Q. Zhou, Z. Chen, Y. Zha, Y. Shi, and H. Tang, "Evaluation of MODIS land cover and LAI products in cropland of North China plain using in situ measurements and Landsat TM images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3087–3097, Oct. 2007.

[10] G. Patenaude, R. Milne, M. Van Oijen, C. S. Rowland, and R. A. Hill, "Integrating remote sensing datasets into ecological modelling: A Bayesian approach," *Int. J. Remote Sens.*, vol. 29, no. 5, pp. 1295–1315, Mar. 2008.

[11] S. Lauritzen, "Propagation of probabilities, means, and variances in mixed graphical association models," *J. Amer. Statist. Assoc.*, vol. 87, no. 420, pp. 1098–1108, Dec. 1992.

[12] M. Kalacska, A. Sanchez-Azofeifa, T. Caelli, B. Rivard, and B. Boerlage, "Estimating leaf area index from satellite imagery using Bayesian networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 8, pp. 1866–1873, Aug. 2005.

[13] Y. T. Mustafa, P. E. Van Laake, and A. Stein, "Bayesian network modeling for improving forest growth estimates," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 639–649, Feb. 2011.

[14] H. Chan, "Sensitivity analysis of Bayesian networks and its application for service engineering," in *Proc. Int. Conf. Soft Comput. Pattern Recognit.*, Dec. 2009, pp. 551–556, IEEE Comput. Soc.

[15] O. Pourret, P. Naïm, and B. Marcot, *Bayesian Networks: A Practical Guide to Applications*. Hoboken, NJ: Wiley, 2008, ser. Statistics in Practice.

[16] F. Gao, J. T. Morisette, R. E. Wolfe, G. Ederer, J. Pedelty, E. Masuoka, R. Myneni, B. Tan, and J. Nightingale, "An algorithm to produce temporally and spatially continuous MODIS-LAI time series," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 1, pp. 60–64, Jan. 2008.

[17] E. M. L. Beale and R. J. A. Little, "Missing values in multivariate analysis," *J. R. Statist. Soc. Ser. B*, vol. 37, no. 1, pp. 129–145, 1975.

[18] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.

[19] R. J. A. Little and D. B. Rubin, *Statistical Analysis With Missing Data*, 2nd ed. Hoboken, NJ: Wiley, Sep. 2008, ser. Probability and Statistics.

[20] J. Pearl, *Causality: Models, Reasoning, and Inference*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[21] F. V. Jensen and T. D. Nielsen, *Bayesian Networks and Decision Graphs*, 2nd ed. New York: Springer-Verlag, 2007, ser. Inf. Sci. Statist.

[22] D. Geiger and D. Heckerman, "Learning Gaussian networks," in *Proc. 10th Annu. Conf. Uncertainty Artif. Intell.*, R. L. S. de Mántaras and D. Poole, Eds., Washington, DC, 1994, Morgan Kaufmann.

[23] R. D. Shachter and C. R. Kenley, "Gaussian influence diagrams," *Manage. Sci.*, vol. 35, no. 5, pp. 527–550, May 1989.

[24] D. Heckerman, D. Geiger, and D. M. Chickering, "Learning Bayesian networks: The combination of knowledge and statistical-data," *Mach. Learn.*, vol. 20, no. 3, pp. 197–243, 1995.

[25] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, 2nd ed. Hoboken, NJ: Wiley-Interscience, 2008, Probability and Statistics.

[26] M. T. Van Wijk, S. C. Dekker, W. Bouten, W. Kohsiek, and G. M. Mohren, "Simulation of carbon and water budgets of a Douglas-fir forest," *For. Ecol. Manage.*, vol. 145, no. 3, pp. 229–241, May 2001.

[27] E. Steingröver and W. Jans, "Physiology of forest-grown Douglas-fir trees, effects of air pollution and drought," Wageningen Univ. Res. Centre, Wageningen, The Netherlands, 1995, Tech. Rep.

[28] R. Waring and N. McDowell, "Use of a physiological process model with forestry yield tables to set limits on annual carbon balances," *Tree Physiol.*, vol. 22, no. 2/3, pp. 179–188, Feb. 2002.

[29] N. C. Coops and R. H. Waring, "Estimating forest productivity in the eastern Siskiyou Mountains of Southwestern Oregon using a satellite driven process model, 3-PGS," *Can. J. Forest Res.*, vol. 31, no. 1, pp. 143–154, Jan. 2001.

[30] J. M. Maass, J. M. Vose, W. T. Swank, and A. Martinez-Yrizar, "Seasonal changes of leaf area index (LAI) in a tropical deciduous forest in west Mexico," *Forest Ecol. Manage.*, vol. 74, no. 1/3, pp. 171–180, Jun. 1995.

[31] M. Kuroda and M. Sakakihara, "Accelerating the convergence of the EM algorithm using the vector ε algorithm," *Comput. Statist. Data Anal.*, vol. 51, no. 3, pp. 1549–1561, Dec. 2006.
[32] I. Meilijson, "A fast improvement to the EM algorithm on its own terms," *J. R. Statist. Soc. Ser. B*, vol. 51, no. 1, pp. 127–138, 1989.

Yaseen T. Mustafa (M'10) received the B.Sc. degree in mathematics science from the University of Mosul, Mosul, Iraq, in 2000 and the M.Sc. degree in mathematics science from the University of Duhok, Duhok, Iraq, in 2005. He is currently working toward the Ph.D. degree in the Department of Earth Observation Science, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands.

His current research interests include the area of remote sensing, including mathematical and statistical tools, such as Bayesian networks.

Valentyn A. Tolpekin (M'08) received the M.Sc. degree in theoretical physics from Odessa State University, Odessa, Ukraine, in 1997, the M.Sc. degree in computer science from South Ukrainian Pedagogical University, Odessa, in 1998, and the Ph.D. degree in physics from the University of Twente, Enschede, The Netherlands, in 2004.

He is currently an Assistant Professor with the Department of Earth Observation Science, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente. His research interests include mathematical and statistical tools for remote sensing image analysis, including Bayesian methods, wavelets, and neural networks.

Alfred Stein received the M.S. degree in mathematics and information science, with a specialization in applied statistics, from the Eindhoven University of Technology, Eindhoven, The Netherlands, and the Ph.D. degree in agricultural and environmental sciences from the Wageningen University and Research Centre, Wageningen, The Netherlands.

He is currently a Professor of mathematical and statistical methods for geodata with the Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands, where he is also the Chairman of the Department of Earth Observation Science. He is the Chief Editor of the *International Journal of Applied Earth Observation and Geoinformation*. His main interest includes spatial statistics and image mining, including spatial data quality. Applications emerge from a range of agricultural, urban, and environmental fields.