Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Environmental Sciences 7 (2011) 74–79

Procedia Environmental Sciences

1st Conference on Spatial Statistics 2011 – Mapping Global Change

Application of the EM-algorithm for Bayesian Network Modelling to Improve Forest Growth Estimates

Y. T. Mustafa*, V. Tolpekin, and A. Stein

Faculty of Geo-Information Science and Earth Observation of the University of Twente (ITC), Enschede 7500 AE, The Netherlands.

Abstract

Leaf area index (LAI) is a biophysical variable that is related to atmosphere-biosphere exchange of CO_2. One way to obtain LAI value is by the Moderate Resolution Imaging Spectroradiometer (MODIS) biophysical products (LAI MODIS). The LAI MODIS has been used to improve the physiological principles predicting growth (3-PG) model within a Bayesian Network (BN) set-up. The MODIS time series, however, contains gaps caused by persistent clouds, cloud contamination, and other retrieval problems. We therefore formulated the EM-algorithm to estimate the missing MODIS LAI values. The EM-algorithm is applied to three different cases: successive and not successive two winter seasons, and not successive missing MODIS LAI during the time study of 26 successive months at which the performance of the BN is assessed. Results show that the MODIS LAI is estimated such that the maximum value of the mean absolute error between the original MODIS LAI and the estimated MODIS LAI by EM-algorithm is 0.16. This is a low value, and shows the success of our approach. Moreover, the BN output improves when the EM-algorithm is carried out to estimate the inconsecutive missing MODIS LAI such that the root mean square error reduces from 1.57 to 1.49. We conclude that the EM-algorithm within a BN can handle the missing MODIS LAI values and that it improves estimation of the LAI.

Selection and peer-review under responsibility of Spatial Statistics 2011

Keywords: EM-algorithm; Gaussian Bayesian networks (GBNs); leaf area index (LAI); Moderate Resolution Imaging Spectroradiometer (MODIS).

1. Introduction

Forests play a critical role in carbon sequestration [1], thus affecting the speed of climate change. Therefore, monitoring forest growth has received increasing attention [2]. An interesting parameter in observing forest growth is the leaf area index (LAI), defined as the total one-sided area of leaf tissue per unit ground surface area (m^2m^{-2})[2]. The LAI is estimated using process-based models, such as the Physiological Principles in Predicting Growth (3-PG) model, being a stand-level model of forest growth [3]. Similarly, remote sensing (RS) also provides the LAI estimates. For instance, the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor provides 8-day global data sets of the LAI [4].

Bayesian networks (BNs) have been used to estimate forest growth parameters [5, 6]. A BN is a directed acyclic graph consisting of nodes and arcs, to represent variables and the dependencies between variables, respectively [7]. Gaussian Bayesian network (GBN) has been used to improve LAI estimates by combining the 3-PG model output with MODIS images [6]. This approach relays on availability of satellite images. RS data, however, often contain gaps (missing values) due to atmospheric characteristics. A major development in statistical methods came in the 1970s with the maximum likelihood (ML) estimation [8, 9], and the expectation maximization (EM)-algorithm have been used to find ML [9].

* Corresponding author Y.T. Mustafa, Tel.: +31 68 416 4774; fax: +31 53 487 4335.
E-mail address: Mustafa@itc.nl.

1878-0296
Selection and peer-review under responsibility of Spatial Statistics 2011
doi:10.1016/j.proenv.2011.07.014

The objective of this study is to handle missing data in a GBN using the EM-algorithm. Therefore, the EM-algorithm is formulated and applied to handle the missing MODIS LAI values by estimating the missing parameters which are needed to execute GBN approach in Mustafa et al.[6].

2. Bayesian network

A BN is a probabilistic graphical model that provides a graphical framework of complex domains with lots of inter-related variables. Mustafa et al. [6] designed a network to improve LAI estimation by combining LAI values derived from MODIS images and estimated by the 3-PG model. Fig. 1(a) shows the graphical part of BN. The intermediate node (LAI_{BN}) represents the estimated LAI values of BN. Based on the continuous variation of LAI over time, it has shown in [6] that LAI follow normal distribution where the GBN is applied. A GBN is a BN where the joint probability distribution associated with its variables $\mathbf{LAI} = \{LAI_1, \dots, LAI_n\}$ is the multivariate normal distribution $N(\mu, \Sigma)$, given by $f(\mathbf{LAI}) = (2\pi)^{-n/2}|\Sigma|^{-1/2} exp\left\{-\frac{1}{2}(\mathbf{LAI}-\mu)^T\Sigma^{-1}(\mathbf{LAI}-\mu)\right\}$. 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 normal distribution with density

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

where μ_i is the expectation of LAI_{BN_i} at time i, the β_{ij} are a 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. Further, Σ_i is the unconditional variance of the LAI_{BN_i}, Σ_{ipa_i} are the covariances between LAI_{BN_i} and the variables pa_i, and Σ_{pa_i} is the covariance matrix of pa_i. For more details about a GBN of improving forest growth estimates and its mathematical formulation we refer to [6].

3. EM-algorithm for estimating missing values in a GBN

The Expectation Maximization (EM)-algorithm is a technique for estimating parameters of statistical models from incomplete data. The EM-algorithm is applicable for maximizing likelihoods. The EM-algorithm is formulated and applied in this study to handle the problem of missing satellite data by estimating the missing parameters that are needed to implement a GBN approach in Mustafa et al. [6].

Consider missing data of satellite images at the i^{th} moment ($i > 1$) of the GBN as shown in Fig.1(b). The GBN output, LAI_{BN_i}, conditionally depends on three nodes (variables), i.e., LAI_{M_i}, $LAI_{BN_{i-1}}$, and LAI_{3PG_i}, where LAI_{M_i} is considered as a missing value. Let (X, Y) be the complete data set at the i^{th} moment of GBN, with observed (complete) data $Y = \left\{LAI_{BN_{i-1}}, LAI_{3PG_i}, LAI_{BN_i}\right\}$ and missing data $X = LAI_{M_i}$ (Fig. 1 (b)). For clarity, we re-name the variables in the GBN model as $y = LAI_{BN_i}$, $x = LAI_{M_i}$, $z = LAI_{BN_{i-1}}$, $w = LAI_{3PG_i}$. Hence expression (1) can be reformulated as:

$$f(y|x,z,w) \sim N\left(\mu_y + \beta_{yx}(x-\mu_x) + \beta_{yz}(z-\mu_z) + \beta_{yw}(w-\mu_w), \sigma_y^2\right). \quad (2)$$

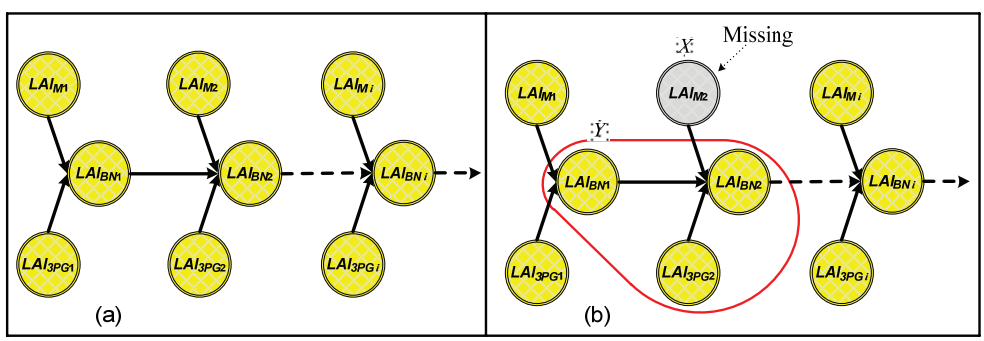

Fig. 1. (a) The BN for i^{th} iterations. Each iteration consists of three nodes LAI_{3PG_i}, $LAI_{BN_i}$and LAI_{M_i}; (b) BN with missing satellite observations. Y represents an observed data set consisting of three nodes LAI_{BN_2} , LAI_{BN_1} and LAI_{3PG_2} while X represents the variable LAI_{M_2} for which an observation is missing.

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

- Choose an initial setting for the parameters θ and name it as θ^{old}. These are guessed based on seasonal changes of LAI values that are obtained from MODIS observations as:

$$\theta^{\text{old}} = \left(\mu^{\text{old}},\ \sigma^{\text{old}}\right) = \begin{cases} \left(\mu_x - \left|\frac{\mu_{x_{i-2}} - \mu_{x_{i-1}}}{\mu_{x_{i-1}}}\right|, \sigma_x - \left|\frac{\sigma_{x_{i-2}} - \sigma_{x_{i-1}}}{\sigma_{x_{i-1}}}\right|\right) & \text{if } \mu_{x_{i-2}} \le \mu_{x_{i-1}} \\ \left(\mu_x + \left|\frac{\mu_{x_{i-2}} - \mu_{x_{i-1}}}{\mu_{x_{i-1}}}\right|, \sigma_x + \left|\frac{\sigma_{x_{i-2}} - \sigma_{x_{i-1}}}{\sigma_{x_{i-1}}}\right|\right) & \text{Otherwise} \end{cases} \quad (3)$$

where μ_x, σ_x are the mean and the standard deviation values of the MODIS LAI, and obtained either for the period from September to February (nongrowing season), or for the period from March to August (growing season). The determination of which period needs to obtain the μ_x, σ_x, is based on the occurrence of missing observation in that period. The $\left|\frac{\mu_{x_{i-2}} - \mu_{x_{i-1}}}{\mu_{x_{i-1}}}\right|$ and $\left|\frac{\sigma_{x_{i-2}} - \sigma_{x_{i-1}}}{\sigma_{x_{i-1}}}\right|$ are the relative changes of the mean and the standard deviation of the previous two MODIS LAI observations. Adding or subtracting these relative changes are based on the condition of an increase or decrease the MODIS LAI during the period of non-growing or growing season.

- 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,

$$\begin{aligned} Q\left(\theta, \theta^{\text{old}}\right) &= E_X\left[\log f(Y, X|\theta)|Y, \theta^{\text{old}}\right] \\ &= \int \log f(Y, X|\theta)\, f\left(X|Y, \theta^{\text{old}}\right) dX = \int \log f(x, y, z, w|\theta)\, f\left(x|y, z, w, \theta^{\text{old}}\right) dx, \end{aligned} \quad (4)$$

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*. Therefore, $\log f(x, y, z, w|\theta)$ can be expressed as:

$$\begin{aligned} \log f(x, y, z, w|\theta) = &-\frac{1}{2}\left(\frac{1}{\sigma_x^2} + \frac{\beta_{yx}^2}{\sigma_y^2}\right)x^2 + \left(\frac{\left(y - \mu_y + \beta_{yx}\mu_x - \beta_{yz}(z - \mu_z) - \beta_{yw}(w - \mu_w)\right)\beta_{yx}}{\sigma_y^2} + \frac{\mu_x}{\sigma_x^2}\right)x \\ &-\frac{1}{2}\left(\frac{\left(y - \mu_y + \beta_{yx}\mu_x - \beta_{yz}(z - \mu_z) - \beta_{yw}(w - \mu_w)\right)^2}{\sigma_y^2} + \frac{(z - \mu_z)^2}{\sigma_z^2} + \frac{(w - \mu_w)^2}{\sigma_w^2} + \frac{\mu_x^2}{\sigma_x^2}\right) - \log\left(4\pi^2\sigma_y\sigma_x\sigma_z\sigma_w\right). \end{aligned} \quad (5)$$

Based on the graphical representation of the GBN model, $f\left(x\middle|y, z, w, \theta^{\text{old}}\right) = \frac{f\left(x, y, z, w|\theta^{\text{old}}\right)}{\int f\left(x, y, z, w|\theta^{\text{old}}\right)dx}$, therefore (4) after some simplification can be written as: $Q = \int_{-\infty}^{\infty} V\,(-fx^2 + gx - h)e^{-ax^2 + bx - c}dx$,

where $V = \frac{\sqrt{\sigma_y^2 + \beta_{yx}^2\left(\sigma^{\text{old}}\right)^2}}{\sqrt{2\pi}\sqrt{\sigma_y^2\left(\sigma^{\text{old}}\right)^2}}$, $f = \frac{1}{2}\left(\frac{1}{\sigma_x^2} + \frac{\beta_{yx}^2}{\sigma_y^2}\right)$, $g = \left(\frac{\left(y - \mu_y + \beta_{yx}\mu_x - \beta_{yz}(z - \mu_z) - \beta_{yw}(w - \mu_w)\right)\beta_{yx}}{\sigma_y^2} + \frac{\mu_x}{\sigma_x^2}\right)$,

$h = \frac{1}{2}\left(\frac{\left(y - \mu_y + \beta_{yx}\mu_x - \beta_{yz}(z - \mu_z) - \beta_{yw}(w - \mu_w)\right)^2}{\sigma_y^2} + \frac{(z - \mu_z)^2}{\sigma_z^2} + \frac{(w - \mu_w)^2}{\sigma_w^2} + \frac{\mu_x^2}{\sigma_x^2}\right) - \log\left(4\pi^2\sigma_y\sigma_x\sigma_z\sigma_w\right)$, $a = \frac{1}{2}\left(\frac{1}{\left(\sigma^{\text{old}}\right)^2} + \frac{\beta_{yx}^2}{\sigma_y^2}\right)$,

$b = \left(\frac{\left(y - \mu_y + \beta_{yx}\mu^{\text{old}} - \beta_{yz}(z - \mu_z) - \beta_{yw}(w - \mu_w)\right)\beta_{yx}}{\sigma_y^2} + \frac{\mu^{\text{old}}}{\left(\sigma^{\text{old}}\right)^2}\right)$ and $c = \frac{b^2}{4a}$.

Here μ^{old} and σ^{old} refer the guessed mean and standard deviation of *x* obtained using (3). The $Q(\theta, \theta^{\text{old}})$ after calculate the integral is: $Q\left(\theta, \theta^{\text{old}}\right) = \Omega\left(-\frac{f}{2a} + \frac{bg}{2a} - \frac{fb^2}{(2a)^2} - h\right)$, where $\Omega = V\sqrt{\frac{\pi}{a}}e^{\left(-c + \frac{b^2}{4a}\right)}$.

- M-step: compute the ML estimates of the parameters θ by maximizing the expected likelihood found during the E-step i.e., $\theta^{\text{new}} = \arg\max_\theta Q\left(\theta, \theta^{\text{old}}\right)$. Hence, by differentiation $Q\left(\theta, \theta^{\text{old}}\right)$ with respect to θ, and solve the differentiation equations for $\theta = (\mu_x, \sigma_x)$, the maximum values are found:

$$\mu_x^{\text{new}} = \frac{\sqrt[3]{\phi + \Psi}}{6\lambda} + \frac{2\left(-3\delta\lambda + \alpha^2\right)}{3\lambda\sqrt[3]{\phi + \Psi}} + \frac{\alpha}{3\lambda} \text{ and } \sigma_x^{\text{new}} = \sqrt{(\mu_x^{\text{new}})^2 - \frac{2C}{B}\mu_x^{\text{new}} + \frac{E}{B}}, \quad (6)$$

where $\phi = -36\,\delta\alpha\lambda + 108\,\eta\lambda^2 + 8\,\alpha^3$, $\Psi = 12\sqrt{3}\sqrt{4\,\delta^3\lambda - \delta^2\alpha^2 - 18\,\delta\alpha\lambda\eta + 27\,\eta^2\lambda^2 + 4\,\eta\alpha^3}\lambda$, $\lambda = AB$, $\alpha = 2\,AC + DB, \delta = AE + B^2 + 2\,DC, \eta = CB + dE$.

Here, $A = 4\,\frac{a^2\Omega\beta_{yx}^2}{\sigma_y^2}$, $B = 4\,a^2$, $C = 2\,ab\Omega$, $D = 2\frac{ab\Omega\beta_{yx}^2}{\sigma_y^2} - 4\frac{a^2\Omega(y-\mu_y-\beta_{yz}(z-\mu_z)-\beta_{yw}(w-\mu_w))\beta_{yx}}{\sigma_y^2}$, and $E = 2\,a\Omega + b^2\Omega$.

- Check for convergence of θ^{new} values. If $\left|\theta^{\text{new}} - \theta^{\text{old}}\right| \le \varepsilon$ is not satisfied, then let $\theta^{\text{old}} \leftarrow \theta^{\text{new}}$, and the algorithm returns to E-step, where ε is the stop criterion which has been selected to be 10^{-5}.

4. Implementation

The GBN is applied to the Speulderbos forest in The Netherlands where the LAI is available as a time series from July 2007 until September 2009. The site is well described elsewhere [6]. The time study contains two winter seasons (October-March) and two summer seasons (May-August). To implement the approach of this work, we consider missing values, by removing some of MODIS LAI observations successively and not successively, as the satellite missing cases is expected. The missing values are estimated using EM-algorithm, and they compared with the original LAI_M. Missing satellite imageries mainly occur during the winter season, due to the atmospheric conditions. Moreover, satellite images may not be available in other seasons due to the incomplete track spatial coverage. Therefore, the EM-algorithm is applied to estimate missing MODIS LAI in three cases. The first and the second case are successive and not successive missing LAI_M during two winter seasons (first and second, respectively). The third case concerns not successive missing LAI_M during the study period (Jul., 2007-Sept., 2009). For the clarity, the figures of LAI estimates are including only the values of interest, i.e., LAI_{FD}, LAI_M and LAI_{BN}.

5. Results of applying EM-algorithm to estimate LAI_M missing within a GBN

5.1. GBN performance with LAI_M estimates during the first winter season

The accuracy of LAI_M and LAI_{BN} is tested using the root mean square error (RMSE) and the relative error (RE) with respect to the LAI field observation (LAI_{FD}) before and after performing EM-algorithm (Table 1). The averaged absolute error (AAE) of the estimating LAI_M with respect to the original LAI_M is calculated as well. Fig. 2(a) shows LAI values after performing EM-algorithm of estimating five successive missing LAI_M. The RMSE and the RE of LAI_{BN} are 1.53 and 13.2%, respectively. Whereas, in Fig. 2(c) five not successive missing LAI_M are estimated, where the RMSE and the RE of LAI_{BN} are 1.51 and 13.3%, respectively. Nevertheless, the deviation between LAI_{BN} and the LAI_{FD} becomes larger after performing the EM-algorithm to estimate eight successive missing LAI_M (Fig. 2(b)). The RMSE and the RE of LAI_{BN} after and before performing the EM-algorithm equals 1.68 against 1.57 and 17.6% against 14.7%, respectively. The estimated missing LAI_M represents the original LAI_M. This is observed especially with the case of not successive missing, with an AAE of 0.02.

Table 1. The 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 of Successive and not Successive Missing LAI_M Estimated (SME) during the first winter season.

Cases	RMSE				RE%				AAE		
	without missing	5 SME	8 SME	5 not SME	without missing	5 SME	8 SME	5 not SME	5 SME	8 SME	5 not SME
LAI_M	3.26	3.26	3.23	3.26	44.1%	44.0%	43.6%	44.1%	0.05	0.1	0.02
LAI_{BN}	1.57	1.53	1.68	1.51	14.7%	13.2%	17.6%	13.3%			

5.2. GBN performance with LAI_M estimates during the second winter season

Here, we found that the LAI_{BN} with performing EM-algorithm is still close to the LAI_{FD} (Table 2). Fig. 3(a) shows LAI_{BN} and LAI_M after estimating five successive missing LAI_M, where the RMSE and the RE of LAI_{BN} are1.59 and 14.7%, respectively. While the RMSE and the RE of the LAI_{BN} after five not successive missing LAI_M estimated are 1.51 and 14.4%, respectively (Fig. 3(c)). Moreover, the estimated missing LAI_M is close to the original LAI_M with AAE values less than 0.08. The differences between LAI_{BN} and LAI_{FD} has occurred after applying the EM-algorithm to estimate eight successive missing LAI_M (Fig. 3(b)). The RMSE and the RE of LAI_{BN} after and before performing the EM-algorithm equals 1.69 against 1.57 and 17.0% against 14.4%, respectively.

Table 2. The 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 of Successive and not Successive Missing LAI_M Estimated (SME) during the second winter season.

Cases	RMSE				RE%				AAE		
	without missing	5 SME	8 SME	5 not SME	without missing	5 SME	8 SME	5 not SME	5 SME	8 SME	5 not SME
LAI_M	3.26	3.25	3.24	3.22	44.1%	44.0%	43.8%	43.6%	0.04	0.08	0.04
LAI_{BN}	1.57	1.59	1.69	1.51	14.7%	14.7%	17.0%	14.4%			

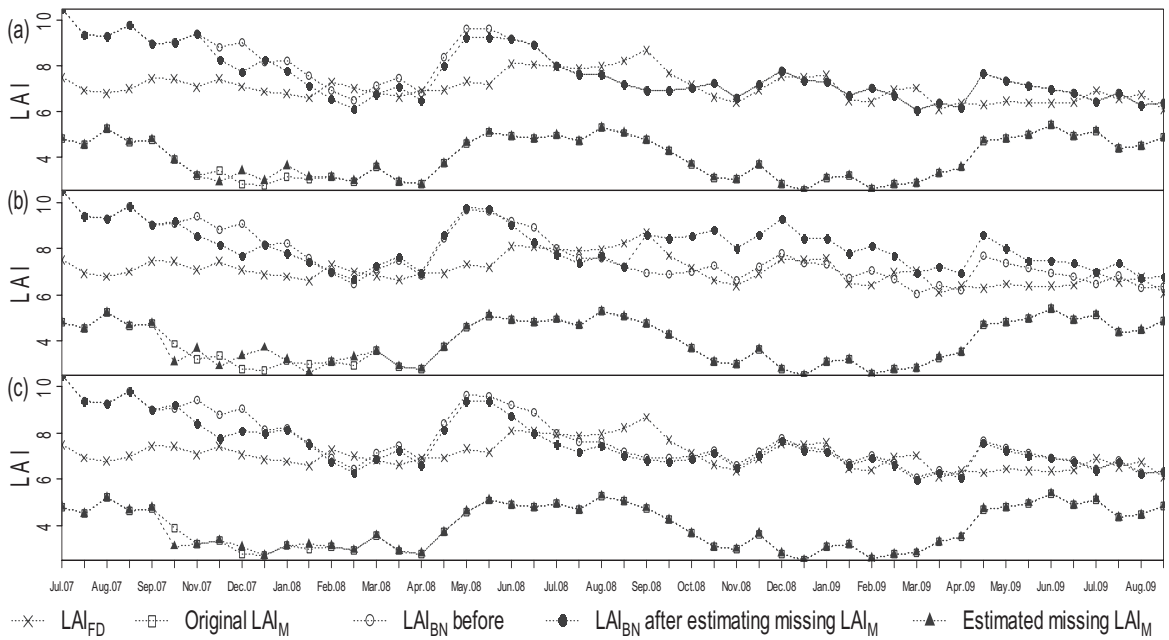

Fig. 2. LAI_{BN} and LAI_M values of the Speulderbos forest obtained before and after performing the EM-algorithm during the first winter season; (a) 5 successive missing LAI_M estimated, (b) 8 successive missing LAI_M estimated, and (c) 5 not successive missing LAI_M estimated.

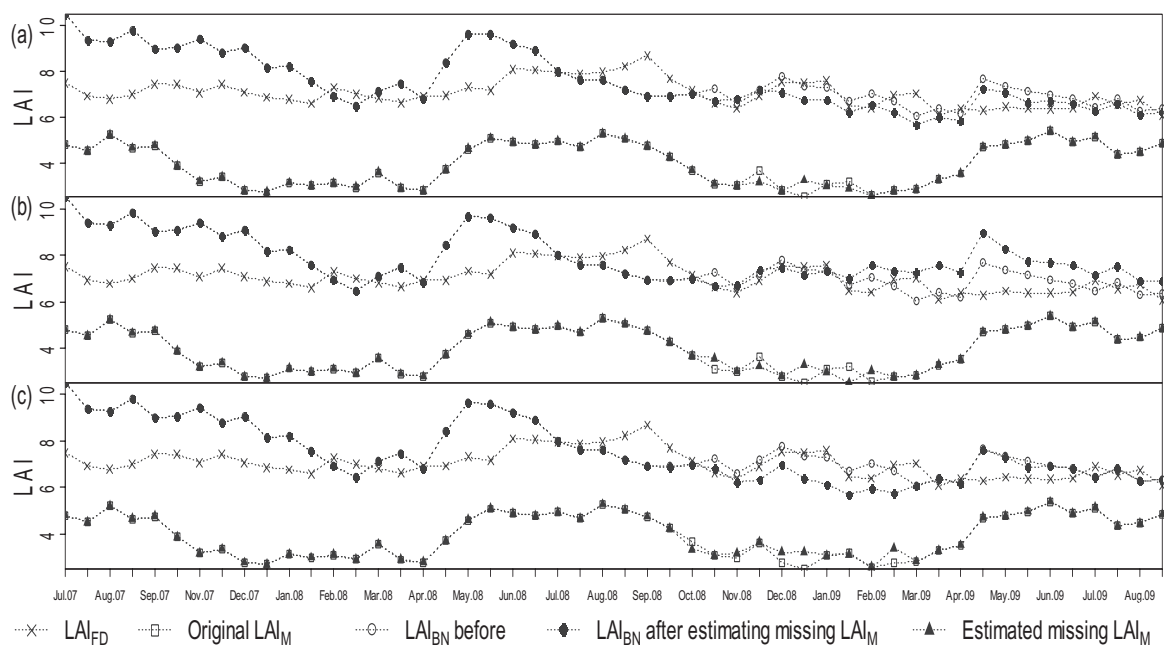

Fig. 3. LAI_{BN} and LAI_M values of the Speulderbos forest obtained before and after performing the EM-algorithm during the second winter season; (a) 5 successive missing LAI_M estimated, (b) 8 successive missing LAI_M estimated, and (c) 5 not successive missing LAI_M estimated.

5.3. GBN performance with LAI_M estimates of not successive missing during the whole time period

Finally, the EM-algorithm is carried out to estimate the 16 LAI_M of not successive missing. The differences between LAI_{BN} and LAI_{FD} reduces after applying the EM-algorithm (Fig.4). The RMSE and the RE of LAI_{BN} is 1.49 against 1.57 and 14.0% against 14.7%, respectively. Moreover, the RMSE and the RE of the LAI_M equal 3.27 against 3.26 and 44.4% against 44.1%, respectively, with an AAE of 0.16.

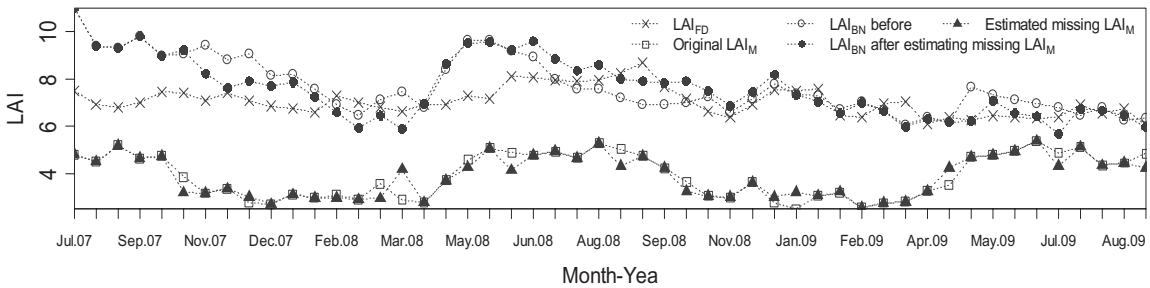

Fig. 4. LAI_{BN} and LAI_M values of the Speulderbos forest obtained before and after performing the EM-algorithm for not successive missing LAI_M estimated during the period from July 2007 until September 2009.

6. Discussion and conclusion

In this study the EM-algorithm is formulated within GBN and the missing LAI_M is estimated. Our results show that the missing LAI_M is estimated successfully such that it represents the origin LAI_M trend. 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. A common criticism of the EM-algorithm is that the convergence can be quite slow [9]. In order to save computing time, it is essential to start with good initial parameters. For this reason we resorted expression (3) such that we can identify the initial values as a closest value to the estimate LAI_M values, however, in some cases it required 804 iterations. From the results of performing EM-algorithm to estimate the missing LAI_M, we observed that the small difference between the LAI_M estimates 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 [6]. We conclude that the missing LAI_M values are estimated successfully using the EM-algorithm. The more than five successive missing LAI_M has an influence on GBN output such that LAI_{BN} does not match the LAI_{FD}. Further, we conclude that LAI_{BN} is improved after performing the EM-algorithm with not successive missing LAI_M during the whole time period study.

References

[1] Wamelink GWW, Wieggers HJJ, Reinds GJ, Kros J, Mol-Dijkstra JP, van Oijen M, et al. Modelling impacts of changes in carbon dioxide concentration, climate and nitrogen deposition on carbon sequestration by European forests and forest soils. *For Ecol Manag*. 2009;**258**:1794-805.
[2] Bonan GB. Importance of leaf area index and forest type when estimating photosynthesis in boreal forests. *Remote Sens Environ*. 1993;**43**:303-14.
[3] Landsberg JJ, Waring RH. A generalised model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning. *For Ecol Manag*. 1997;**95**:209-28.
[4] Myneni RB, Hoffman S, Knyazikhin Y, Privette JL, Glassy J, Tian Y, et al. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sens Environ*. 2002;**83**:214-31.
[5] Kalacska M, Sanchez-Azofeifa A, Caelli T, Rivard B, Boerlage B. Estimating leaf area index from satellite imagery using Bayesian networks. *IEEE T Geosci Remote*. 2005;**43**:1866-73.
[6] Mustafa YT, Van Laake PE, Stein A. Bayesian Network Modeling for Improving Forest Growth Estimates. *IEEE T Geosci Remote*. 2011;**49**:639-49.
[7] Jensen FV, Nielsen TD. *Bayesian networks and decision graphs*. 2nd ed. New York: Springer; 2007.
[8] Beale EML, Little RJA. Missing Values in Multivariate Analysis. *J R Stat Soc B Met*. 1975;**37**:129-45.
[9] Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from Incomplete Data via the EM Algorithm. *J R Stat Soc B Met*. 1977;**39**:1-38.