Bayesian maximum entropy data fusion of field-observed leaf area index (LAI) and Landsat Enhanced Thematic Mapper Plus-derived LAI

Aihua Li, Yanchen Bo & Ling Chen

School of geography, Beijing Normal University, Beijing, 100875, China

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

Version of record first published: 10 Sep 2012.

To cite this article: Aihua Li, Yanchen Bo & Ling Chen (2013): Bayesian maximum entropy data fusion of field-observed leaf area index (LAI) and Landsat Enhanced Thematic Mapper Plus-derived LAI, International Journal of Remote Sensing, 34:1, 227-246

To link to this article: http://dx.doi.org/10.1080/01431161.2012.712234

PLEASE SCROLL DOWN FOR ARTICLE
demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Bayesian maximum entropy data fusion of field-observed leaf area index (LAI) and Landsat Enhanced Thematic Mapper Plus-derived LAI

Aihua Lia, Yanchen Bo, and Ling Chen

School of geography, Beijing Normal University, Beijing 100875, China; 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 13 September 2011; accepted 28 March 2012)

Accurate high-resolution leaf area index (LAI) reference maps are necessary for the validation of coarser-resolution satellite-derived LAI products. In this article, we propose an efficient method based on the Bayesian Maximum Entropy (BME) paradigm to combine field observations and Landsat Enhanced Thematic Mapper Plus (ETM+) derived LAI surfaces in order to produce more accurate LAI reference maps. This method takes into account the uncertainties associated with field observations and with the regression relationship between ETM+-derived LAI and field measurements to perform a non-linear prediction of LAI, the variable of interest. In order to demonstrate the difference by soft data and hard data, we estimate the LAI reference maps by three BME interpolation methods, BME1, BME2, and BME3. BME1 and BME2 perform maximum estimation and mean estimation, respectively, by taking the ETM+-derived LAI as interval soft data and the field LAI measurements as hard data. BME3 is utilized when ETM+-derived LAI surfaces are processed as uniform probability soft data and field measurements are processed as Gaussian probability soft data. Three study sites are selected from the BigFoot project (NASA’s Earth Observing System validation programme) (http://www.fsl.orst.edu/larse/bigfoot/index.html). In regard to the mean and standard deviation of LAI surfaces, standard deviation predicted by BME methods has lower values than that derived by ETM+. The mean value of the BME-predicted LAI, which takes into account the uncertainties of field measurements, is lower than that of ETM+-derived LAI at each study site. A comparison with field measurements shows that BME1, BME2, and BME3 have root mean square errors (RMSE) of 0.455, 0.485, and 0.517 and average biases of $-0.017$, $-0.010$, and $-0.304$, respectively. The RMSEs and biases of the predicted LAI surfaces are less when compared to the ETM+-derived LAI, which has the average RMSE and bias of 0.642 and $-0.080$. When the field measurements are processed as soft data, the predicted LAI by BME3 has more bias than those of the predictions by BME1 and BME2, but has less RMSE than that of the ETM+-derived LAI by 0.125. In summary, BME is capable of incorporating the spatial autocorrelation and the uncertainties in the field LAI measurements into the LAI surface estimation to produce a more accurate LAI surface with less RMSE in validation. The maximum estimation has relatively better accuracy than the mean estimation. The results indicate that the BME is a promising method for fusing point-scale and area-scale data.

*Corresponding author. Email: boyc@bnu.edu.cn
1. Introduction

Leaf area index (LAI), defined as one-half of the total green leaf area per unit ground horizontal surface area (Chen and Black 1992), is one of the important biophysical variables necessary for land surface dynamic process models (Reich, Turner, and Bolstad 1999; Running et al. 1999; Xiao et al. 2009). Nowadays, several moderate-resolution LAI products are produced from different satellite sensors such as Polarization and Directionality of the Earth’s Reflectances (POLDER) (Roujean and Lacaze 2002; Lacaze 2005), the Moderate Resolution Imaging Spectroradiometer (MODIS) (Myneni et al. 2002), the Multiangle Imaging Spectroradiometer (MISR) (Knyazikhin et al. 1998; Hu et al. 2003), the Advanced Very High Resolution Radiometer (AVHRR) (Chen et al. 2002), and VEGETATION (Feng et al. 2006; Baret et al. 2007; Weiss et al. 2007). These moderate-resolution LAI products should be validated before application (Justice and Townshend 1994; Cihlar, Chen, and Li 1997; Liang 2004). Since validation is the only way to prove the real accuracy of the satellite products, it is worthwhile to produce the most accurate reference continuous maps, even with many inevitable uncertainties and great difficulties.

Field LAI measurements and high-resolution LAI surface maps are two kinds of the so-called true data against which the moderate-resolution satellite-derived LAI products are evaluated (Berterretche et al. 2005). Field measurements are the most essential validation data. A system of land validation core sites that represent a range of biome types has been established (Morissette, Privette, and Justice 2002). Some network sites, such as BigFoot, VALERI, and FLUXNET (Cohen, Maiersperger, and Pflugmacher 2006; Baret et al. 2008; Garrigues et al. 2008), are available for LAI validation. Heterogeneity makes pixel-scale validation not simply equivalent to field measurement average. The field measurement data are usually interpolated into surface data, but the accuracy of geostatistical methods to obtain LAI surface maps from field data is limited to the number and the spatial distribution of measurement points (Christakos 2000; Liang 2004). As another choice of true reference map, the high-resolution LAI surface, which can be routinely retrieved from high-resolution Earth Observation (EO) data, covers extensive regions, but it has a lower accuracy when compared to field measurements. Ideally, the combination of these two data types makes accurate parameter retrievals possible.

Regression analysis has been widely used to link field data and Landsat ETM+ imagery of the dominant canopy to provide a continuous higher-resolution reference surface for LAI validation (Cohen et al. 2003). Regression analyses such as ordinary least squares (OLS), inverse OLS, and reduced major axis (RMA) (Curran and Hay 1986; Van Huffel 1997; Cohen et al. 2003) are non-spatial methods. Since most regression analyses rely on spectral vegetation indices (SVIs), such as the normalized difference vegetation index (NDVI), which are limited by their asymptotic nature in relation to LAI (Chen and Cihlar 1996), the regression relationships between LAI and SVIs are uncertain. Kriging with an external drift (KED) and sequential Gaussian conditional simulation (SGCS) are two spatial methods for mapping LAI surfaces (Berterretche et al. 2005). The latter can preserve the anisotropy observed in semivariograms of measured LAI and maintain the global pattern. Neither the regression analyses nor the traditional geostatistical methods have the ability to take into account the errors of field LAI measurements and the weakness of regression models. Wilson et al. (2011) utilize a hierarchical Bayesian model (HBM) to link field data and remote-sensing data. HBM offers a framework to explicitly incorporate data collected at different scales without losing information by aggregation (Agarwal et al. 2005). However, the information from neighbours is not incorporated into the integration procedure. Therefore, a new improved strategy that can integrate uncertain data in the fusion procedure of multi-source data is necessary to map more accurate LAI reference surface for moderate-resolution LAI validation.
As one of the spatio-temporal knowledge synthesis and mapping methods, Bayesian maximum entropy (BME) (Christakos 2000; Christakos and Serre 2000; Christakos et al. 2001) has been successfully applied in fusing field observations and categorical data (Bogaert and D’Or 2002; D’Or 2003; Douaik et al. 2004) or multi-sensor data (Bogaert et al. 2009). It estimates variables by non-linear prediction and non-Gaussian distribution hypothesis. The most interesting strength is its capability to deal with soft data with uncertainties from different sources because BME is based on the Bayesian rule. Model errors and measurement errors usually exist in data. Taking into account the uncertainties in multi-source data fusion fits more to reality and can achieve more accurate fusion results (Christakos 2000).

This article focuses on reconstructing LAI reference maps through blending field LAI measurements and ETM+-derived LAI surfaces based on BME. In order to explore the differences that arise in the ways the uncertainties of data are processed, we estimate the LAI reference maps by three methods: BME1, which utilizes BME to estimate the maximum value at each estimated point when the ETM+-derived LAI surfaces and the field LAI measurements are processed as interval soft data and hard data, respectively; BME2, which makes a mean estimation at each point when the ETM+-derived LAI surfaces and the field LAI measurements are processed as uniform probability soft data and hard data, respectively; BME3, which makes a mean estimation when the ETM+-derived LAI surfaces interval soft data is processed as uniform probability soft data and the field measurements are processed as Gaussian probability soft data. The comparison between BME1 and BME2 is to examine the difference between the maximum and mean estimations. The comparison between BME2 and BME3 is to examine the difference between the hard field measurements and the soft field measurements. Three sites, which are dominated by forest, grass, and crop, are selected from the BigFoot project (Cohen et al. 2006) for this study. The goal of this study is to explore the potential of BME in combining different levels of reliability and different scales of LAI data for much more accurate LAI reference maps.

2. Methodology

2.1. Soft data and hard data

Soft data are also called fuzzy data, and are associated with different sources of uncertainties. On the contrary, hard data are used as accurate data (Christakos 2000). In remote-sensing product validation, measurement error is the first source of uncertainty (Gower, Kucharik, and Norman 1999; Fernandes et al. 2003). Although repeated measurements can cancel out some random errors, it cannot remove measurement bias. With regard to LAI ground measurements, an indirect optical method with some uncertainties is usually preferred over the direct accurate method, since the latter is very laborious (Chen and Cihlar 1995; Chen et al. 1997; Jonckheere et al. 2004). The second source of uncertainty is brought by parameter retrieval methods which include empirical, physical, and hybrid models (Chen et al. 1997; Gower, Kucharik, and Norman 1999; Weiss et al. 2004; Chen et al. 2006). The mismatch in the temporal or spatial scale is the third source of uncertainty, especially when multi-scale data are required to be merged for specific applications (Chen et al. 1997; Liang 2004; Lee 2005; Lee, Yeatts, and Serre 2009; Yu et al. 2009). Other sources of uncertainty include the transformation of secondary data into soft data (Cassiani and Christakos 1998; Lee 2005) and other experimental knowledge (Christakos 2000; Kovitz and Christakos 2004).

Soft data can be expressed in terms of interval values and probability statements in mathematical computation, other empirical charts, and expert knowledge (Christakos
230 A. Li et al.

Exploratory data analysis is usually operated by frequency distribution (histogram) statistical analysis. The probabilistic data are approximately normally distributed or Student’s $t$-distributed (Lee, Balling, and Gober 2008; Bogaert et al. 2009). Some data need to be treated as interval data with upper and lower bounds when they are not fit for probability distribution but with physical meanings (Douaik, Van Meirvenne, and Toth 2005).

### 2.2. BME algorithm

Field measurements without considering their measurement errors are used as the input to the traditional geostatistical method, which interpolates point data of a target parameter into its spatial context (Journel 1989; Christakos and Li 1998; Christakos 2000). However, plenty of information can be acquired for a certain study area. As a recent hot spot in research fields, multi-source information integration makes great advances in regional interpolation. In this article, we intend to integrate most of the useful data or information to reconstruct the LAI surface for an advanced LAI reference map. For certain validation sites, field point measurements and remote-sensing-derived LAI surface maps with different spatial scales are two kinds of data type. Due to the limitations in the spatial representative of the point measurements in the heterogeneous landscape, it is necessary to integrate multi-scale data with different accuracy levels.

BME is one of the methods in the spatio-temporal knowledge synthesis and mapping family. It is an extension of spatial geostatistics (Christakos 1990a, 1990b; Christakos and Li 1998; Christakos 2000). BME has been successfully applied in soil science (Bogaert and D’Or 2002; D’Or 2003; Douaik et al. 2004; Douaik, Van Meirvenne, and Toth 2005), environmental risk assessment (Lee 2005; Yu et al. 2009; Bogaert et al. 2009), environmental health (Puangthongthub et al. 2007, Lee, Yeatts, and Serre 2009; Pang, Christakos, and Wang 2010), and so on (Money 2008; Lee, Balling, and Gober 2008). Major differences between BME and the conventional geostatistical methods have been concluded from five respects (Table 1), which include data source, input data type, how the covariance model is used, whether the characteristic of prediction is linear, and whether there is a distribution hypothesis for input data. BME is a probabilistic approach that could perform interpolation with uncertain data (i.e. soft data). Since any auxiliary data with errors can be considered as soft data, BME theoretically provides a new way to synthesize multi-source related data. Additionally, BME that does not require its input is normally distributed and can predict variables beyond the reach of linear Gaussian spatial statistics.

Figure 1 shows the major steps of applying the BME algorithm (Christakos 2000). The first step is to select data to get experiential covariance functions and then model the fit covariance functions. The second step is to introduce a Lagrange multiplier to compute

<table>
<thead>
<tr>
<th>Methods</th>
<th>BME</th>
<th>Geostatistical methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data source</td>
<td>Multi-source data sets of different accuracy level</td>
<td>Field measurements</td>
</tr>
<tr>
<td>Input data type</td>
<td>Uncertain and accurate data (multi-scale data)</td>
<td>Accurate data (single scale)</td>
</tr>
<tr>
<td>Covariance model</td>
<td>For joint probability</td>
<td>For weight coefficient</td>
</tr>
<tr>
<td>Distribution hypothesis</td>
<td>No hypothesis</td>
<td>Normal distribution</td>
</tr>
<tr>
<td>Linear/non-linear</td>
<td>Non-linear</td>
<td>Linear</td>
</tr>
</tbody>
</table>
the maximum entropy with the constraint condition of minimizing the variance and then to obtain the joint probability density function (PDF). The final step is to add specific information (hard and soft data) to the posterior PDF and to calculate the mean or maximum value.

The variable $x_{\text{map}}$ consists of a vector of points $x_{\text{soft}}$, $x_{\text{hard}}$, and $x_k$, which denote values at soft and hard data points and unknown value at estimation point, respectively. BME is based on the Bayesian rule and has the goal to obtain the posterior PDF. However, BME does not need to get the prior PDF and model the likelihood function, while the joint PDF is calculated. The objective of the first stage of BME is to compute the joint PDF $f_G(x_{\text{map}})$ given the general knowledge $G$, $g_\alpha(x_{\text{map}})$ is a set of functions $x_{\text{map}}$ such as mean and covariance moments. $f_G(x_{\text{map}})$ is achieved by maximizing the entropy with constraint to variance. The Shannon information entropy $\phi$ is given by Equation (1). The Lagrange multiplier $\lambda_\infty$ is introduced to maximize the entropy (Equation (2)). $L[f_G(x_{\text{map}})]$ is the object function for maximizing entropy after introducing the Lagrange multiplier. $E[g_\alpha(x_{\text{map}})]$ is the expected value for $g_\alpha(x_{\text{map}})$.

\[
\phi [x_{\text{map}}, f_G(x_{\text{map}})] = -f_G(x_{\text{map}}) \log [f_G(x_{\text{map}})],
\]

\[
L[f_G(x_{\text{map}})] = - \int f_G(x_{\text{map}}) \log [f_G(x_{\text{map}})] \, dx_{\text{map}} \\
- \sum_\alpha \lambda_\alpha \left[ \int g_\alpha(x_{\text{map}}) f_G(x_{\text{map}}) \, dx_{\text{map}} - E[g_\alpha(x_{\text{map}})] \right].
\]

Figure 1. Concept framework of BME theory.
Specific observations consist of hard and soft data collected for the points to be estimated. If soft data are composed of probability distribution data, then the posterior PDF $f^*$ at estimation point can be written as Equation (3). Finally, from Equation (3), several estimators such as the mean (Equation (4)) and the mode (Equation (5)) can be extracted. $\bar{x}_{k|\kappa}$ is the estimation of $x_k$.

$$f^* (x_k | x_{\text{soft}}, x_{\text{hard}}) = \frac{f_G (x_{\text{soft}}, x_{\text{hard}}, x_k)}{f (x_{\text{soft}}, x_{\text{hard}})},$$

(3)

$$\bar{x}_{k|\kappa} = \int x_k f_k (x_k) \, dx_k,$$

(4)

$$\bar{x}_{k|\kappa} = \max (f_k (x_k)).$$

(5)

3. Data

3.1. Study sites

The BigFoot project is funded by the Earth Science Enterprise to collect and organize data to be used in NASA’s Earth Observing System (EOS) validation program. Validation data sets submitted to the Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center (DAAC) are available at http://daac.ornl.gov/BIGFOOT_VAL/bigfoot.shtml. Three study sites, the agricultural cropland site (AGRO), the tall grass prairie site (KONZ), and the temperate mixed forest site (HARV), are selected from the nine sites of the BigFoot validation project for this study. The AGRO site is located just south of Champaign, Illinois. This temperate cropland site has alternate crops of corn and soy bean. The KONZ site is located at Konza Prairie near Manhattan, Kansas. The HARV forest site is located at Harvard Forest near Petersham, Massachusetts. At each site, there are about one hundred 25 m × 25 m plots where land cover, LAI, absorbed radiation, and net primary production are measured at five to nine subplots per plot. Subplot measurements are averaged to provide a single value for each measured variable at each plot. The field LAI measurements, Landsat ETM+-derived LAI, and land-cover surfaces in 2000 are involved in this study. Each LAI and land-cover surface has a grain of 25 m and covers a 7 km × 7 km extent. The specifications of HARV, AGRO, and KONZ sites are described in Table 2.

3.2. Uncertainties in field measurements and ETM+-derived LAI surface maps

There are many ways to measure LAI in the field. The destructive harvest method for the crop and allometric equations for the forest are direct methods to obtain more accurate field measurements (Chen et al. 1997; Jonckheere et al. 2004). But considerable efforts are required to obtain spatially representative estimates. As a result, field estimates of LAI are more often obtained through indirect optical measurement instruments (Chen et al. 1997; Gower, Kucharik, and Norman 1999). These methods, such as LI-COR LAI-2000 probe and hemispherical photography, relate the amount of foliage to the probability of a beam of light that will pass through the canopy, the so-called gap probability. LAIs from direct methods are reported as true LAIs and LAIs from optical measurements are reported as effective LAIs (Chen et al. 1997; Gower, Kucharik, and Norman 1999; Weiss et al. 2004; Chen et al. 2006; Garrigues et al. 2008) associated with several sources of uncertainties, such as foliage clumping, saturation of the optical signal in dense canopies, simplification of leaf optical properties, disregard of forest understory, and redundant measurements of
Table 2. Specifications of the HARV, AGRO, and KONZ sites.

<table>
<thead>
<tr>
<th>Sites</th>
<th>Location</th>
<th>Vegetation type</th>
<th>Data sets used</th>
<th>Dates when data obtained</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>HARV</td>
<td>Harvard Forest LTER, Massachusetts, USA</td>
<td>Temperate mixed forest</td>
<td>Field measurements, ETM+ derived LAI, ETM+ land cover</td>
<td>18 June 2000, 4 August 2000</td>
<td>LAI-2000</td>
</tr>
<tr>
<td>AGRO</td>
<td>Bondville, Illinois, USA</td>
<td>Cropland: corn and soybean</td>
<td>Field measurements, ETM+ derived LAI, ETM+ land cover</td>
<td>14 July 2000, 11 August 2000</td>
<td>Direct harvest</td>
</tr>
</tbody>
</table>
non-green plant elements, which can result in some positive and negative biases (Garrigues et al. 2008).

The high-resolution LAI surface is another choice of true reference map. The ideal and key methodology is a spatial scaling technique to evaluate the accuracy of satellite-derived LAI by fusing field measurements and high-resolution remotely sensed imagery into a new high-resolution and more accurate LAI surface. The high-resolution imagery-derived LAI surface is a bridge to link the point measurements and the coarse pixel values in the validation process. The ground measurements are used to calibrate the products from high-resolution imagery, which are then aggregated to the sensor coarse resolutions as reference maps (Liang 2004). The algorithm of ETM+-derived LAI in BigFoot validation sites is based upon two key techniques. First, the multiple date imagery is utilized and then integrated into a single index by canonical correlation analysis. An orthogonal regression method, called reduced major axis (RMA), is used to build the relationship between the integrated NDVI and LAI from field measurements. So the accuracy of LAI reference map is primarily affected by the errors in the ground measurements and the modelled relationship between the ETM+-derived SVI and the ground measurements.

LAI-2000 and harvest methods are used to collect the field-observed LAI data from the HARV and AGRO sites, respectively (Cohen, Maiersperger, and Pflugmacher 2006). Both methods are used in the KONZ site (Cohen, Maiersperger, and Pflugmacher 2006). Here, the bias between effective LAI by the LAI-2000 probe and true LAI by the harvest method has not been taken into account, since our study focuses on improving the reference LAI maps by combining field measurements and ETM+-derived LAI surfaces.

3.3. Creating interval soft data and modelling covariance functions

Field-observed LAI is measured multiple times at each plot. Usually, field measurements are processed as hard data without any uncertainties using average values. We can make field-observed LAI fit for Gaussian distribution, which could take into account measurement errors. Other soft data are obtained from ETM+-derived LAI surfaces by the regression relationship between ETM+-derived LAI surface and field measurement data. Seventy per cent (100) of the 149 field measurements in the HARV site and 80% (80) of the 99 field measurements in both the AGRO and KONZ sites for the dominant vegetation types are selected randomly for BME modelling, and the ones remaining are used for the validation of the BME results. Regression model statistics for each site are shown in Table 3. Traditional OLS regression (Figure 2) is performed for the interval soft data. First, we calculate the variance of residuals $\sigma^2$, and then the linear estimation values $\text{LAI}_{\text{estimation}}$. Secondly, the lower ($\text{LAI}_a$) and upper ($\text{LAI}_b$) values are calculated from Equations (6) and (7). Note that some vegetation types that lack field measurements in corresponding sites are not included in the regression models. LAIs of these vegetation types in ETM+-derived LAI surfaces are given as the constant.

<table>
<thead>
<tr>
<th>Site</th>
<th>Slope</th>
<th>Intercept</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HARV</td>
<td>0.61</td>
<td>1.92</td>
<td>0.41</td>
</tr>
<tr>
<td>AGRO</td>
<td>0.98</td>
<td>0.0484</td>
<td>0.86</td>
</tr>
<tr>
<td>KONZ</td>
<td>0.59</td>
<td>0.835</td>
<td>0.29</td>
</tr>
</tbody>
</table>
Figure 2. Regression model (trend line in red) for field-observed LAI and corresponding ETM+-derived LAI: HARV (a), AGRO (b), KONZ (c).
To demonstrate the differences between hard data and soft data used in the BME interpolation, three BME interpolations by BME1, BME2, and BME3 are performed at each site. In Figure 3, field measurements processed as Gaussian probability soft data are shown in blue. The distance between the upper and lower red squares is the interval data. The green dashed line denotes the mean value of the soft data. The Gaussian probability soft data can be only obtained in some vegetation types where field measurements are available, as shown in Figure 3. For each vegetation type, one pixel or plot is selected as an example to show the soft data (Figure 3).

Experiential covariance models are modelled with positive linear combinations of effects: nugget, spherical, and exponential models (Figure 4). The model parameters are shown in Table 4. The first covariance component of ENF is the nugget, which is due to the existence of observation random errors. The second covariance component has a covariance sill ($c_1$) of 0.7 and a spatial range ($a_1$) of 750 m, while the third component has a covariance sill ($c_2$) of 0.8 and a spatial range ($a_2$) of 250 m. The DBF, MF, and GL have similar spatial covariance models to ENF. Spatial heterogeneity of these four vegetation types changes more slowly than that of OS, CO, and SY, which have the similar spatial covariance models.

4. Results and discussion

Figure 5 shows the ETM+-derived LAI surfaces of the three measurement sites and their corresponding estimation maps of the three interpolation methods mentioned above. ETM+-derived LAI surfaces are shown in Figures 5(a), (e), and (i). The interpolation results of BME1 are shown in Figures 5(b), (f), and (j). Correspondingly, the results of BME2 are shown in Figures 5(c), (g), and (k). The results of BME3 are shown in Figures 5(d), (h), and (l). The predicted LAI maps and the original ETM+-derived LAI maps at the AGRO site do not show obvious difference (Figures 5(e)–(h)). However, at the HARV site (Figures 5(a)–(d)), the result of BME1 is more approximate to the ETM+-derived LAI, whereas BME2 retrieves a relatively smaller number of high LAI values. The histograms are used to report the variability of the mean and the standard deviation with the ETM+-derived LAI and the estimated LAIs. BME1 and BME2 have a similar mean and standard deviation in each site, as shown in Figure 6. Mean values of LAI predicted by BME1 and BME2 are higher than that derived by ETM+, except in the KONZ site, and the standard deviations predicted by BME1 and BME2 in each site are lower than that derived by ETM+, as shown in Figure 6. The mean of LAI predicted by BME1 and BME2 increases by about 0.23 more than that derived by ETM+ at HARV site and is close to that derived by ETM+ at AGRO site. Compared to ETM+-derived LAI and the LAIs predicted by BME1 and BME2, BME3 obtains the lowest mean at each site.

Cross-validation was not involved in our evaluation and the field observations excluding those for regression are used for the validation of predicted LAI. Figure 7 shows the scatter plots of the ETM+-derived LAI and the predicted LAI versus the field-observed LAI. It shows that the most predicted values are approximate to the observed values. Fewer outliers exist in the predicted LAIs than in the ETM+-derived LAIs. Summary statistics (Equations (8)–(11)) are used to quantitatively evaluate the accuracy of the results. These statistics include the $R$ square ($R^2$), root mean square error (RMSE), mean bias (Bias), correlation (CR), and the variance ratio (VO). VO is calculated as the standard deviation
Figure 3. Interval ETM+-derived LAI data (green solid line is about the mean values, red is about the upper and lower limitations) and Gaussian probability field measurements data (blue solid line): HARV site (a), AGRO site (b), KONZ site (c).
Figure 4. Nested covariance models of different land-cover types: Dots are experiential models and lines are fit models.
Table 4. Parameters of the covariance models.

<table>
<thead>
<tr>
<th>Biome</th>
<th>Nested covariance models</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENF</td>
<td>( C(s) = c_{\text{nugget}} + c_1 \exp\left(-\frac{3s}{\alpha_1}\right) + c_2 \exp\left(\frac{3s}{\alpha_2}\right) )</td>
<td>( c_{\text{nugget}} = 0.1; c_1 = 0.7; \alpha_1 = 750; c_2 = 0.8; \alpha_2 = 250 )</td>
</tr>
<tr>
<td>DBF</td>
<td>( c_{\text{nugget}} = 0.1; c_1 = 0.4; \alpha_1 = 1100; c_2 = 0.6; \alpha_2 = 250 )</td>
<td></td>
</tr>
<tr>
<td>MF</td>
<td>( c_{\text{nugget}} = 0.03; c_1 = 0.4; \alpha_1 = 750; c_2 = 0.8; \alpha_2 = 250 )</td>
<td></td>
</tr>
<tr>
<td>GL</td>
<td>( c_{\text{nugget}} = 0.1; c_1 = 0.5; \alpha_1 = 1500; c_2 = 0.45; \alpha_2 = 400 )</td>
<td></td>
</tr>
<tr>
<td>OS</td>
<td>( C(s) = c_{\text{nugget}} + c_1 \exp\left(-\frac{3s}{\alpha_1}\right) + c_2 \text{sph}\left(1 - \left(\frac{3s}{2\alpha_1}\right)^2 - \left(\frac{s}{2\alpha_2}\right)^2\right))</td>
<td>( c_{\text{nugget}} = 0.01; c_1 = 0.7; \alpha_1 = 1800; c_2 = 0.2; \alpha_2 = 500 )</td>
</tr>
<tr>
<td>CO</td>
<td>( c_{\text{nugget}} = 0.1; c_1 = 0.3; \alpha_1 = 750; c_2 = 1.2; \alpha_2 = 900 )</td>
<td></td>
</tr>
<tr>
<td>SY</td>
<td>( c_{\text{nugget}} = 0.09; c_1 = 0.25; \alpha_1 = 950; c_2 = 1.0; \alpha_2 = 750 )</td>
<td></td>
</tr>
</tbody>
</table>

Notes: ENF, evergreen needleleaf forest; DBF, deciduous broadleaf forest; MF, mixed forest; GL, grassland; OS, open shrubland, CO, corn; SY, soya bean; exp, exponential covariance function; sph, spherical covariance function.

Figure 5. ETM+-derived LAI surfaces and prediction surfaces of BME1, BME2, and BME3 are shown from left to right, respectively: HARV site (top), AGRO site (middle), KONZ site (bottom).

\[ \text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left[L_{\text{predicted}}(x_i) - L_{\text{observed}}(x_i)\right]^2}{n}}, \quad (8) \]
Figure 6. Histograms of ETM+–derived LAI and predicted LAI at HARV (a), AGRO (b), and KONZ (c) sites, respectively. With regard to each site, the histograms of ETM+–derived LAI surface and three predicted LAI surfaces of BME1, BME2, and BME3 are shown from left to right.

Note: The first value is mean and the second one is standard deviation.
Figure 7. ETM+-derived LAI and predicted LAI versus observed LAI at HARV (a), AGRO (b), and KONZ (c) sites, respectively. Predicted LAI1, LAI2, and LAI3 are the results of BME1, BME2, and BME3, respectively.
Bias = \sum_{i=1}^{n} \left[ L_{\text{predicted}}(x_i) - L_{\text{observed}}(x_i) \right] / n, \quad (9)

CR = \frac{\sum_{i=1}^{n} \left[ L_{\text{predicted}}(x_i) - \bar{L}_{\text{predicted}}(x) \right] \left[ L_{\text{observed}}(x_i) - \bar{L}_{\text{observed}}(x) \right]}{\sqrt{\sum_{i=1}^{n} \left[ L_{\text{predicted}}(x_i) - \bar{L}_{\text{predicted}}(x) \right]^2} \sqrt{\sum_{i=1}^{n} \left[ L_{\text{observed}}(x_i) - \bar{L}_{\text{observed}}(x) \right]^2}}, \quad (10)

\text{VO} = \frac{\text{Std}_{\text{predicted}}}{\text{Std}_{\text{observed}}}. \quad (11)

In Table 5, we can see that $R^2$ and CR values of BME methods are higher than those of ETM+-derived LAI, and the RMSEs of BMEs are lower than that of ETM+-derived LAI at each site. The average RMSEs of BME1, BME2, and BME3 in all sites are 0.455, 0.485, and 0.517, respectively, which are lower than the RMSE (0.642) of ETM+-derived LAI. Lower RMSEs means fewer extreme data. The biases of LAI predicted by BME1 and BME2 are reduced from $-0.080$ to $-0.017$ and $-0.010$, respectively, compared to ETM+-derived LAI. At all study sites, the average bias of BME1 is higher than that of BME2 by 0.007 and the average bias of RMSE of BME1 is lower than that of BME2 by 0.03. So the maximum estimation (BME1) can get more approximate results to the field measurements than the mean estimation (BME2). BME3 has the largest bias of $-0.304$ in all LAI surfaces, including ETM+-derived LAI and predicted LAIs. Both BME1 and BME2 consider field measurements as hard data without uncertainty; their RMSEs and biases are almost less than those of BME3, which takes into account the uncertainty of field measurements. BME3 preferably transfers the field measurements into soft data. So the predicted LAI by BME3 gains the negative bias and decreases the RMSE. The statistical values are worse than those of BME1 and BME2 because there are no hard (accurate) data used in the prediction by BME3.

Table 5. Summary statistics of the LAI predictions compared to the field measurements.

<table>
<thead>
<tr>
<th>Sites</th>
<th>Number of plots</th>
<th>Methods</th>
<th>$R^2$</th>
<th>RMSE</th>
<th>Bias</th>
<th>CR</th>
<th>VO</th>
</tr>
</thead>
<tbody>
<tr>
<td>HARV</td>
<td>48</td>
<td>ETM+-derived LAI</td>
<td>0.57</td>
<td>0.688</td>
<td>$-0.054$</td>
<td>0.754</td>
<td>1.290</td>
</tr>
<tr>
<td></td>
<td>BME1</td>
<td>0.59</td>
<td>0.518</td>
<td></td>
<td>$-0.030$</td>
<td>0.770</td>
<td>0.996</td>
</tr>
<tr>
<td></td>
<td>BME2</td>
<td>0.59</td>
<td>0.344</td>
<td></td>
<td>0.014</td>
<td>0.766</td>
<td>0.656</td>
</tr>
<tr>
<td></td>
<td>BME3</td>
<td>0.57</td>
<td>0.351</td>
<td></td>
<td>$-0.311$</td>
<td>0.754</td>
<td>0.657</td>
</tr>
<tr>
<td>AGRO</td>
<td>19</td>
<td>ETM+-derived LAI</td>
<td>0.82</td>
<td>0.631</td>
<td>0.049</td>
<td>0.905</td>
<td>1.070</td>
</tr>
<tr>
<td></td>
<td>BME1</td>
<td>0.89</td>
<td>0.460</td>
<td></td>
<td>0.099</td>
<td>0.942</td>
<td>0.988</td>
</tr>
<tr>
<td></td>
<td>BME2</td>
<td>0.84</td>
<td>0.582</td>
<td></td>
<td>0.062</td>
<td>0.917</td>
<td>1.050</td>
</tr>
<tr>
<td></td>
<td>BME3</td>
<td>0.82</td>
<td>0.623</td>
<td></td>
<td>$-0.215$</td>
<td>0.906</td>
<td>1.060</td>
</tr>
<tr>
<td>KONZ</td>
<td>19</td>
<td>ETM+-derived LAI</td>
<td>0.45</td>
<td>0.436</td>
<td>$-0.275$</td>
<td>0.669</td>
<td>1.100</td>
</tr>
<tr>
<td></td>
<td>BME1</td>
<td>0.89</td>
<td>0.114</td>
<td></td>
<td>0.043</td>
<td>0.945</td>
<td>0.657</td>
</tr>
<tr>
<td></td>
<td>BME2</td>
<td>0.78</td>
<td>0.157</td>
<td></td>
<td>$-0.062$</td>
<td>0.883</td>
<td>0.628</td>
</tr>
<tr>
<td></td>
<td>BME3</td>
<td>0.45</td>
<td>0.216</td>
<td></td>
<td>$-0.376$</td>
<td>0.671</td>
<td>0.548</td>
</tr>
<tr>
<td>All sites</td>
<td>86</td>
<td>ETM+-derived LAI</td>
<td>0.88</td>
<td>0.642</td>
<td>$-0.080$</td>
<td>0.939</td>
<td>1.070</td>
</tr>
<tr>
<td></td>
<td>BME1</td>
<td>0.92</td>
<td>0.455</td>
<td></td>
<td>$-0.017$</td>
<td>0.961</td>
<td>0.970</td>
</tr>
<tr>
<td></td>
<td>BME2</td>
<td>0.92</td>
<td>0.485</td>
<td></td>
<td>$-0.010$</td>
<td>0.958</td>
<td>0.969</td>
</tr>
<tr>
<td></td>
<td>BME3</td>
<td>0.91</td>
<td>0.517</td>
<td></td>
<td>$-0.304$</td>
<td>0.952</td>
<td>0.964</td>
</tr>
</tbody>
</table>
The VO of the LAI predicted by each BME used in this study is less than that of the ETM+-derived LAI (Table 5), which agrees with the fact that all the LAI surfaces predicted by BME methods have lower standard deviations. The VO of the LAI predicted by BME1 is close to 1.0 at both HARV and AGRO sites, which implies that the variance of observations is preserved in the predictions for these two sites. However, VO is much smaller in KONZ site, although $R^2$ values of BME1 and BME2 are up to 0.89 and 0.78, respectively. The lower VO means that the predicted LAI at KONZ site has a smaller dynamic range than the field LAI measurements.

5. Conclusions

Improved strategies for regression analysis and geostatistical methods have been proposed to create accurate high-resolution LAI reference maps. However, neither of these methods takes into account the uncertainties of input data and models. In addition, regression analysis is limited by the empirical relationship of NDVI and field LAI measurements, whereas the traditional geostatistical methods with rigorous Gaussian hypothesis are limited to linear estimation.

In this article, BME is utilized at three BigFoot sites to produce more accurate high-resolution LAI maps by integrating in situ LAI and Landsat ETM+-derived LAI. Multiple field measurements in one plot are processed as Gaussian soft data. The relationship between field measurements and Landsat ETM+-derived LAI is not perfect and this imperfection brings uncertainties. So we transformed ETM+-derived LAI into soft data based upon OLS regression. Comparisons of ETM+-derived LAI surfaces at three study sites show that the RMSEs can be reduced to at least 0.13 by using BME. Furthermore, the BME methods exhibit a value of VO close to 1.0 in both HARV and AGRO sites by getting the maximum estimation from the posterior PDF. It means that the variance in the observations is preserved in the prediction in each of the two sites.

In summary, BME is able to eliminate the extreme data to reduce RMSE and result in smaller variance. The maximum estimation has relatively higher accuracy than the mean estimation. BME provides a novel approach to account for the uncertainties associated with field observations and with those from the relationship between observed values and regression model outputs to produce more accurate LAI reference maps. BME has the potential to fuse the field measurements and remotely sensed data into the land surface, but its effectiveness is worth further examination.

Acknowledgments

This work was jointly supported by the Key Project of Earth Observation and Navigation Field of the State Programme for High-Tech Research and Development (863 Programme) (2009AA122103), National Key Basic Research Programme of China (the Chinese 973 Programme) under grant (2007CB714402), and the foundation of the State Key Laboratory of Remote Sensing Science. We greatly thank the University of North Carolina Bayesian Maximum Entropy (UNC-BME) laboratory at the UNC at Chapel Hill for some programmes. The authors thank the two anonymous reviewers for their valuable comments.

References


