Article

Geostatistics for Mapping Leaf Area Index over a Cropland Landscape: Efficiency Sampling Assessment

Beatriz Martinez 1,*, Eduardo Cassiraga 2, Fernando Camacho 3 and Javier Garcia-Haro 1

1 Departament de Física de la Terra i Termodinàmica, University of Valencia, Dr. Moliner, 50, 46100, Valencia, Spain; E-Mail: j.garcia.haro@uv.es
2 Grupo de Hidrogeología, Departamento de Ingeniería Hidráulica y Medio Ambiente, Universidad Politécnica de Valencia, Camino de Vera, s/n, 46022, Valencia, Spain; E-Mail: efc@dihma.upv.es
3 Earth Observation Laboratory (EOLAB), Parc Científic Universitat de València, C/Catedrático José Beltrán, 2, 46980 Paterna, Valencia, Spain; E-Mail: fernando.camacho@eolab.es

* Author to whom correspondence should be addressed; E-Mail: beatriz.martinez@uv.es; Tel.: +34-963-543-256; Fax: +34-963-543-385.

Received: 30 September 2010; in revised form: 15 November 2010 / Accepted: 16 November 2010 / Published: 23 November 2010

Abstract: This paper evaluates the performance of spatial methods to estimate leaf area index (LAI) fields from ground-based measurements at high-spatial resolution over a cropland landscape. Three geostatistical model variants of the kriging technique, the ordinary kriging (OK), the collocated cokriging (CKC) and kriging with an external drift (KED) are used. The study focused on the influence of the spatial sampling protocol, auxiliary information, and spatial resolution in the estimates. The main advantage of these models lies in the possibility of considering the spatial dependence of the data and, in the case of the KED and CKC, the auxiliary information for each location used for prediction purposes. A high-resolution NDVI image computed from SPOT TOA reflectance data is used as an auxiliary variable in LAI predictions. The CKC and KED predictions have proven the relevance of the auxiliary information to reproduce the spatial pattern at local scales, proving the KED model to be the best estimator when a non-stationary trend is observed. Advantages and limitations of the methods in LAI field predictions for two systematic and two stratified spatial samplings are discussed for high (20 m), medium (300 m) and coarse (1 km) spatial scales. The KED has exhibited the best observed local accuracy for all the spatial samplings. Meanwhile, the OK model provides comparable results when a well stratified sampling scheme is considered by land cover.
Keywords: geostatistics methods; leaf area index; cropland landscape; spatial sampling

1. Introduction

Accurate leaf area index (LAI) measurements are critical for understanding biological and physical processes associated with vegetation at a variety of spatial and temporal scales [1,2]. LAI is a key structural characteristic of vegetation because it defines the size of the interface for exchange of energy and mass between the canopy and the atmosphere [3]. Hence, it is an important determination for soil-vegetation-atmosphere transfer schemes (SVATS) for regional and global climate modeling, weather forecasting, global carbon cycle, and global change monitoring [4-6]. However, field measurements of LAI are cumbersome and time consuming at large spatial extents, and in that respect, the use of satellite remote sensing is the most effective means of estimating LAI fields at larger areas.

Mapping of continuous variables like LAI from high-resolution imagery (e.g., Landsat TM or ETM+) has largely been depended on aspatial empirical relationships based on calibrated relationships between spectral vegetation indices (SVIs) [7] or multiple spectral bands combination [8,9]. In fact, the estimates of vegetation properties maps from ground-based measurements (i.e., up-scaling process) is generally accepted in the ‘bottom-up’ process of the direct validation strategy of biophysical satellite derived products [10,11].

Traditional methods for empirical modeling of continuous variables, such as LAI, from SVIs rely on ordinary least squares (OLS) regression [12] using the normalized difference vegetation index (NDVI) and the simple ratio (SR) index [13,14]. Although the OLS is a technique with important limitations for such applications [7,15], it has been widely used due to its simple application. Alternatively, improvements over existing uses of regression analysis have been made, such as incorporating multiple bands into a re-weighted iterative OLS algorithm (IRLS) [11,16], including multiple spectral bands into a single predictive index such as canonical correlation analysis (CCA) [7,17,18] or selecting an appropriate regression model [15,19,20].

However, the geographic data often shows spatial correlation being even more relevant than the spatial techniques for mapping vegetation attributes [21], such as the kriging approaches. Numerous studies have demonstrated increased precision of maps through geostatistical techniques that utilize exhaustive secondary information [20,22-26].

Geostatistics capitalized on the spatial correlation between observations to predict attribute values at unsampled locations using information related to one or several attributes. The kriging methods are based on Matheron’s regionalized variable theory [27], which can be utilized to provide the best linear unbiased estimates. Kriging methods have been adapted, extended, and generalized extensively since 1970. For example, kriging has been generalized to evaluate variability of soil physical and chemical properties at the field level [28,29], estimate hydraulic conductivity, moisture content and conductivity of a surface [30-32], predict spatial rainfall [23], predict the area extent of land cover types [33], map forest canopy height [21] or estimate LAI fields [20,34]. Geostatistics is presented as a good alternative to remote sensing for estimating vegetation variables [34,35].
The goal of this paper is to evaluate the performance of geostatistical techniques in the prediction of LAI maps over a cropland landscape using ground-based LAI data. For this purpose, the ordinary kriging (OK), collocated cokriging (CKC) and kriging with an external drift (KED) models are considered to estimate high-spatial resolution LAI maps from LICOR LAI-2000 ground based measurements over an agricultural area (Barrax test site in Albacete region, Spain). Our main concern is to evaluate the efficiency of the three techniques in the estimation of LAI maps from a non-stationary variable (as is the case of agricultural landscapes) as well as their dependence on auxiliary information. LAI ground-based measurements derived from an independent method (digital hemispherical photographs, DHP) and a LAI image (reference image) derived from applying an empirical relationship to a high-resolution image have been used to assess model estimates.

Measuring vegetation characteristics such as LAI is not a straightforward task, especially over a heterogeneous landscape. ‘Where to measure’ is an important question that has to be answered before a study can begin [34]. Special emphasis is given in this work to evaluate the influence of the spatial sampling on the geostatistical estimates over the Barrax cropland landscape. A simulation exercise based on the design of four different spatial samplings (i.e., two stratified and two systematics) from a reference LAI map is addressed. Advantages and limitations of the geostatistical methods in the predictions of the LAI reference for the four spatial samplings are discussed at high (20 m), medium (300 m) and coarse (1 km) spatial resolutions. An important concern of this study is the influence of the auxiliary information according to the spatial resolution.

2. Background

2.1. Ordinary Kriging (OK)

Kriging is a generic name for stochastic interpolation techniques in which a continuous regionalized variable of interest $Z_1(u)$ (i.e., spatial values of a random function $Z(u)$), usually referred as primary variable, is estimated in the study area $A$ at any unsampled location $u$ using the $z_1$ data measured at different locations, $z_1(u_\alpha), \alpha=1,\ldots, n$ [36]. The general kriging estimate is:

$$Z_1^*(u) - m_1(u) = \sum_{\alpha=1}^{n} \lambda_\alpha(u)[Z_1(u_\alpha) - m_1(u_\alpha)]$$ (1)

where $m_1(u)$ is the model trend considered and $\lambda_\alpha(u)$ is the weight assigned to datum $z_1(u_\alpha)$ given a neighborhood $W(u)$ centered on $u$. The $n$ weights are obtained by solving a system of linear equations that must satisfy two conditions. First, weights must be such that the estimate given in Equation (1) is unbiased, $E\{Z_1^*(u) - Z_1(u)\} = 0$. Second, the weights must be such that the error variance $\sigma^2_e(u) = \text{Var}\{Z_1^*(u) - Z_1(u)\}$, is the smallest possible. The coefficients of the kriging estimate depends on, (i) the spatial configuration of the data, (ii) the unobserved location relative to the data locations and (iii) the spatial correlation or the degree to which one location can be predicted from a second location as a function of spatial separation. Observations that are close in space to the estimate receive more weight than those that are further away. In order to determine the weights at unsampled locations, the variogram model is required [37]. The variogram is the main tool used in Geostatistics in order to account for the degree of spatial correlation (i.e., spatial continuity) of a regionalized variable, as a function of the separation distance and direction. In practice, the effectiveness of kriging depends upon
the appropriate selection of the model variogram parameters and how representative the observation points are of the phenomenon [38,39].

All kriging estimators are variants of the Equation (1), where the differences reside in the model considered for the trend $m_1(u)$ and the possibility of including correlated auxiliary information on the estimates. Meanwhile the simplest case is the simple kriging (SK), which considers the trend as known and constant throughout the study area $A$, the ordinary kriging (OK) accounts for local fluctuations of the mean by limiting the domain of stationarity of the mean to the local neighborhood $W(u)$. The OK estimator is thus written as a linear combination of the $n(u)$ regionalized variables $Z_1(u)$:

$$ Z_1^*(u) = \sum_{a=1}^{n(u)} \lambda_a^{OK}(u)Z_1(u_a) $$

where the unknown local mean $m_1(u)$ is filtered from the linear estimator by forcing the kriging weights to sum 1.

Due to its simplicity, OK is widely used as a general interpolation method, but results depend on the measurement of $Z_1$ itself and the validity of the assumptions about stationarity of the mean and the variance. When measurements are sparse or poorly correlated in space, the estimation of the primary variable is generally improved by accounting for one (i.e., secondary variable) or more auxiliary information originating from other continuous attributes. This auxiliary information is assumed to be correlated with the primary data and can be available at all primary samples locations and at all nodes of the estimation grid. In this situation, the estimation is usually done using two extensions of the kriging estimator, kriging with and external drift (KED) and collocated cokriging (CKC).

2.2. Kriging with an External Drift (KED)

Kriging with an external drift is a variant of kriging that accounts for the local spatial trend of the primary variable [40,41] along with the use of exhaustive secondary information, $Z_2$. The distinctive feature of KED is that the algorithm employs a non-stationary random function model, where stationarity is limited within each search neighborhood, yielding more local detail than with ordinary kriging [40]. KED incorporates the secondary variable by assuming that it represents a trend surface to which the mean of the primary variable is linearly related, i.e.,

$$ m_1^*(u) = a + bz_2(u) $$

where $m_1^*(u)$ is the mean of the principal variable at location $u$; $z_2(u)$ is the value of the secondary variable at the same location; and $a$ and $b$ are unknown coefficients. The expression of the estimate of kriging with an external drift is the same as the ordinary kriging estimate (although the kriging coefficients are different since they result from the solution of a different system of equations):

$$ Z_1^*(u) = \sum_{a=1}^{n(u)} \lambda_a^{KED}(u)Z_1(u_a) $$

where the kriging weights are estimated implicitly through the kriging system within each neighborhood and are the solution of a system of $n_1(u) + 2$ linear equations that takes into account the trend components [40,41].
The implementation of this method also requires knowing exhaustively the secondary variable, thus its quality depends on the quality of the secondary variable field. Note that this method does not account for any spatial correlation between the variables.

2.3. Collocated Cokriging (CKC)

The collocated ordinary cokriging [41] (CKC) assumes the trend components of the primary variable \( Z_1 \) and a single secondary attribute \( Z_2 \) to be unknown and stationarity in a local neighborhood area \( W(\mathbf{u}) \). The primary variable is estimated by the collocated cokriging as a linear combination of the \( n_1 \) nearest primary data and the secondary datum collocated at the point being estimated [38]. The expression of this estimate is:

\[
Z_1^*(\mathbf{u}) = \sum_{a=1}^{n_1(\mathbf{u})} \lambda_1^\text{CKC}(\mathbf{u})Z_1(\mathbf{u}_a) + \lambda_2^\text{CKC}(\mathbf{u})[Z_2(\mathbf{u}) - m_2 + m_1]
\]  

(5)

where \( \lambda_1 \) and \( \lambda_2 \) are the weights assigned to the primary and secondary variable respectively, with the constraint \( \sum_{a=1}^{n_1(\mathbf{u})} \lambda_1 + \lambda_2 = 1 \). The weights \( \lambda_1 \) generally decrease as the corresponding primary data location get farther away from the location \( \mathbf{u} \) being estimated, whereas the weight of the secondary variable, \( \lambda_2 \), increases. If primary and secondary variables are uncorrelated, the collocated cokriging reverts to the kriging method \( \lambda_2 = 0 \).

This method incorporates the secondary information as a covariate, i.e., it considers the information on spatial continuity of the principal variable and the linear correlation between the primary and the secondary variables. The collocated cokriging requires the knowledge of the secondary variable exhaustively. Table 1 summarizes the main characteristics of the considered geostatistical techniques.

<table>
<thead>
<tr>
<th>Model</th>
<th>Trend Non-Stationarity</th>
<th>Secondary variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>OK</td>
<td>Unknown local mean ( m_1(\mathbf{u}) )</td>
<td>No</td>
</tr>
<tr>
<td>CKC</td>
<td>Unknown local mean ( m_1(\mathbf{u}) )</td>
<td>Yes</td>
</tr>
<tr>
<td>KED</td>
<td>( m_1^*(\mathbf{u}) = a + bZ_2(\mathbf{u}) )</td>
<td>Yes</td>
</tr>
</tbody>
</table>

3. Methods

3.1. Study Area

The study area, Barrax, is a 5 km x 5 km agricultural area centered at 39°3′N, 2°5′W in the Albacete region (Spain) (Figure 1). Activities in Barrax were extended as part of the Echival Field Experiment in a Desertification-threatened Area (EFEDA) project (1993–1995) [42]. Digital Airborne Imaging Spectrometer Experiments (DAISEX) from ESA, were carried out in this area during 1998, 1999 and 2000 [43] and more recently in the context of SPerctra bARrax Campaign (SPARC) 2003 and 2004, (www.uv.es/leo/sparc/people.htm, [44]), Sentinel-2 and Fluorescence Experiment (SEN2FLEX) 2005 (www.uv.es/leo/sen2flex/) and Sentinel-3 Experiment (SEN3Exp) 2009 (www.esa.int/esaCP/SEMOW432BZF_index_0.html). This area was also one of the VALERI (Validation of LAnd European Remote Sensing Instruments) experiment sites in 2003 [16]. Currently, the Barrax test site is
considered in validation exercises of main projects such as GEOLAND in the context of Global Monitoring of Environment and Security (GMES) (www.gmes-geoland.info/).

The Barrax test site concentrates in a small area different uniform and large agricultural crops up to 1 km which show LAI values ranging from 0 to 6 [43]. The dominant cultivation in the area is approximately 60% of dry land and 40% of irrigated land crops (corn, alfalfa, potato, sugar beet, onion, wheat, legumes, and others).

**Figure 1.** Experiment site location over a HRVIR2/SPOT4 scene (black box on the SPOT scene). The background image consists of a RGB (NIR, RED, GREEN) color composite SPOT image acquired on July 3th, 2003.

### 3.2. Field Measurements

The ground data set was collected in the SPARC 2003 campaign during the period 10–16 July 2003. The SPARC campaign lay within the frame of different international initiatives. This data set has been integrated in the ESA Campaign DataBase (CDB). The ESA-CDB is a generalized version of the ESA ENVISAT Calibration/Validation database, currently available at http://nadir.nilu.no/calval. LAI data was measured with the optical instrument LICOR LAI-2000 plant canopy analyzer [45], which relies on the indirect LAI estimates based on the inversion of gap fraction data [46]. Gap fraction is the fraction of view in some direction from below the canopy that is not blocked by foliage [47]. The LICOR LAI-2000 allows the LAI to be estimated by comparing differential light measurements above and below canopy at different zenith angles and assuming a modified Beer-Lambert formulation [48].

On the other hand, LAI estimates from digital hemispherical photographs [49,50] were available. The digital hemispherical camera is also an indirect optical method, with a continuous view zenith angle and an extreme wide-angle camera lens (e.g., generally 180°) that allows the gap fraction to be evaluated in all viewing directions. These characteristics make the hemispherical camera a potential instrument for the estimation of light regime parameters (e.g., fAPAR) and other measures more directly related to canopy architecture, such as fractional vegetation cover (FVC), LAI or foliage
inclinuation angle [51,52]. Since previous studies have demonstrated the potential of the DHP technique as compared with the LAI-2000 [53], these measurements will be used as validation data set to assess the uncertainty of the LAI fields predicted by the OK, CKC and KED models.

3.3. Auxiliary Information

Spectral information from two relevant HRVIR2/SPOT4 channels: red (R: 610–680 nm), and near infrared (NIR: 780–890 nm) data was used as auxiliary information. The image corresponds to 3 July, 2003 and was georectified to a sampling size of 20 m × 20 m in the UTM 30N-WGS84 projection. The SPOT radiance values were transformed into reflectance at the top of the atmosphere (TOA).

The NDVI [54] was computed from the SPOT TOA reflectance data to use it as a secondary variable in LAI predictions from CKC and KED models since it is highly correlated with vegetation parameters such as green leaf biomass and green leaf area [55,56]. Although this index has shown to be sensitive to background scattering and soil darkening particularly in sparsely vegetated areas [57], this influence becomes less significant in cropland landscapes. Figure 2 shows the NDVI derived from the SPOT TOA reflectance data over the study area.

Figure 2. NDVI derived from the SPOT TOA reflectance data over the study area.

3.4. Sampling Strategy

Designing a sampling strategy includes a series of considerations such as the number of samples and their optimal spatial distribution. This selection often involves tradeoffs between target accuracy and
logistics constraints. The most widely used methods in different disciplines are the random, systematic and stratified spatial samplings. The random method is very simple and straightforward to implement, but it provides a sampling variance which is usually large [58,59]. Meanwhile, the systematic sampling involves establishing sample points at regular intervals, providing the same variance with lower cost than with random samplings. The major disadvantage is that each cover type in the study area does not have an equal chance of being included in the sampling [60]. Finally, the stratified random sampling is potentially the most efficient approach. In essence, it is based on breaking the total area into subunits, either as a set of regular blocks or into ‘natural’ areas based on factors such as soil type or land cover type.

A stratified random sampling method was designed in order to assure a reliable representation of the existent LAI variability, paying special attention to the possible loss in efficiency with inappropriate stratification [8,60]. The stratification combines information from a land cover classification (with accuracy of 95%) and NDVI information [53]. The major land uses classes showing different NDVI levels were sampled (see a more detailed description in [11]). 96 statistically representative Elementary Sampled Units (ESUs) were measured using the LICOR LAI-2000, Figure 3 (left) and 39 ESUs with the hemispherical camera, Figure 3 (right). A total of six and nine different crop types were sampled with the LICOR LAI-2000 and the camera, respectively. The LICOR LAI-2000 measurements were distributed over a larger area than those made with the hemispherical camera. The dimensions of the ESUs were of 20 m × 20 m for both instruments (coinciding with the SPOT image pixel size) over which measurements were averaged to produce a unique value per ESU. A minimum of 12 measurements were taken to characterize the variability within the ESUs.

**Figure 3.** Elementary sampling units (ESUs) acquired with the LICOR LAI-2000 (left) and the hemispherical camera (right). The LAI estimates from the hemispherical camera are assumed to be a suitable validation set of the OK, CKC and KED predictions.

3.5. LAI Predictions from OK, CKC and KED Models

The OK, CKC and KED models were performed with the GSLIB software [39,40]. The first step was to model the spatial continuity of the data by adjusting a theoretical model to experimental
variograms. In case of OK and KED, only the variogram of the primary variable is needed. The interpolation of the primary variable using the CKC is more complex. In this case, the cross variogram between the primary and secondary variable is required. The difficulty of the cokriging model is that it is not possible to build the three variogram models (e.g., primary, secondary and primary-secondary variables) independently of each other. In our case, the modeling effort can be further alleviated by the Markov-type approximation \[41,61\], which entails the symmetry of the cross covariance model \(C_{12}(h) \approx \rho_{12}(0)C_{11}(h)\), where \(\rho_{12}\) refers to the correlation coefficient between primary and secondary variable and \(C_{11}(h)\) corresponds to the covariance of the primary variable. Such symmetry is a consequence of the fact that the cross variogram is assumed to be proportional to the variogram of the primary variable.

3.6. LAI Predictions Assessment

In order to assess the predictions derived from the different models, first, the LAI estimates from the hemispherical camera have been used as validation data. The LAI estimates from the hemispherical camera have been compared with the estimates of the three geostatistics models at each ESU location. Second, a different exercise has been conducted in order to compare the proposed models with current state of art techniques. In this case, a high-spatial resolution LAI image (reference image) derived from a non-spatial method has been compared with the different predictions. The methodology is based on the prediction of a biophysical high-spatial resolution image from an empirical transfer function (TF) which relates the biophysical in situ measurements and concomitant radiance values from a high-resolution image \[11,16\]. The method, currently proposed in the VALERI project, is based on a multivariate ordinary least square method which uses an IRLS algorithm. A TF that relates the LAI ground truth measurements from the LICOR LAI-2000 and the G, R and NIR HRVIR2/SPOT4 bands has shown its validity in the characterization of the Barrax area with a LAI accuracy lower than 0.7 \[11\].

4. Results

4.1. LAI Field Measurements

Figure 4 shows the LAI histograms distribution for both data sets, LICOR LAI-2000, \((\text{LAI}_{\text{LICOR}})\) and DHP \((\text{LAI}_{\text{DHP}})\). An overall good characterization of LAI values is observed in both cases, comprising the range of expected values for agricultural areas \[51\]. Similar mean LAI values were observed for both data sets (around 2.9) with a standard deviation of 1.47 and 1.38, respectively. \(\text{LAI}_{\text{DHP}}\) shows an oversample of very low (0.2–1) LAI values.
**Figure 4.** LAI values estimated from LICOR LAI-2000 (top) and digital hemispherical photographs (bottom) over the Barrax test site.

### 4.2. Modeling Spatial Continuity

Figure 5 shows the Gaussian model fitted to the directional variograms derived for the maximum and minimum anisotropy directions observed from the LAI LICOR field measurements. A clear anisotropy is observed around the 30° NE and 60° NW directions with *range* values (i.e., a measure of the spatial continuity) for each direction reaching almost a distance of 1,600 m and 1,200 m, respectively.
Figure 5. Gaussian models fitted to the directional variograms derived for the maximum (left) and minimum (right) anisotropy directions observed from the LAI_LICOR field measurements.

Figure 6 displays the correlation between the primary (i.e., LAI_LICOR) and secondary (i.e., NDVI) variables considered. A linear relationship is found where the NDVI explain the 74% of the variance shown in the LAI data set ($r = 0.86$, $p$-value < 0.05). Although, the NDVI and LAI do not behave linearly at high values, the NDVI has been selected as secondary information since a major presence of crops with LAI lower than four were found.

Figure 6. A linear relationship between the primary (LAI) and secondary (NDVI) variables is found with $r = 0.86$ at a significance level of 95%.

4.3. LAI Predictions

Figure 7 presents the LAI maps predicted from the OK, CKC and KED models using the LAI_LICOR data. A visual comparison with the NDVI image (Figure 2) can be useful to evaluate overall pattern agreements. The OK estimates (Figure 7, left) reveal that the primary variable sampling is not suitable
for predicting large scale variations, although it reproduces the spatial pattern of the study area at small scales. As it was expected, the CKC (Figure 7, center) and KED (Figure 7, right) predictions prove the relevance of the auxiliary information to reproduce the spatial pattern at local scales, providing the KED model a better prediction on the transition between crops, where a non-stationary trend is observed. The KED model does the best job of predicting the LAI spatial variability and the pattern of the area described by the secondary variable. In fact, the KED model also provides significant improvements in the characterization of the bare areas with respect to the CKC model when the secondary variable is considered.

**Figure 7.** LAI fields from OK (left), CKC (center) and KED (right) models.

In order to assess the error associated to each model, different statistics have been computed using the LAI\(_{DHP}\) data (Table 2). Summary statistics for these predictions indicates that although the CKC model improves the correlation with the LAI\(_{DHP}\) ground-based measurements it provides a high root mean square error (RMSE = 1.2). In contrast, the lowest RMSE differences (0.8) and highest linear correlation coefficient (0.9) given by the KED predictions versus the LAI\(_{DHP}\) measurements indicate the effectiveness of the KED model in the prediction of the LAI parameter. On the other hand, the predicted LAI fields have been compared with the LAI reference map (see Section 3.6). Table 3 shows the summary statistics obtained from the OK, CKC and KED predictions versus the LAI reference map. The comparison with the reference field reveals the low correlation and high mean error introduced by the OK predictions (\(r = 0.3\), RMSE = 2.4) whereas the KED shows the best agreement (\(r = 0.9\), RMSE = 1.2). The high values observed in the intercept parameter may be perceived as a problem associated to insufficient sampling of bare areas and dry lands (e.g., barley and fallow). While they represent about 60% of the region only a few ESUs have been considered for characterizing these areas (less than 10% of sampling points). A significant advantage of the KED model is its better characterization of bare areas.

**Table 2.** Statistics derived between the OK, CKC and KED LAI fields versus LAI estimated from DHP at a significance level of 95%.

<table>
<thead>
<tr>
<th>Model</th>
<th>Slope</th>
<th>Intercept</th>
<th>(r)</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>OK</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>1.3</td>
</tr>
<tr>
<td>CKC</td>
<td>0.9</td>
<td>0.4</td>
<td>0.8</td>
<td>1.2</td>
</tr>
<tr>
<td>KED</td>
<td>0.9</td>
<td>0.5</td>
<td>0.9</td>
<td>0.8</td>
</tr>
</tbody>
</table>
Figure 8 shows the RMSE between the LAI predicted from the OK (left), CKC (center) and KED (right) and the LAI reference map. Both OK and CKC are limited to reproduce the low LAI values in dry lands because they are highly sensitive to the undersampling of bare areas. One possible solution would be the addition of ESUs with LAI values of zero, to non vegetated areas as it is demonstrated in section 4.4. Although both CKC and KED models incorporate the secondary information, the KED model provides significantly lower RMSE due to the use of a non-stationary random function model within each search neighborhood as opposed to the unknown but constant trend model in a local neighborhood area. Special emphasis is dedicated in the next section to analyze the potential of the kriging models with different spatial samplings, particularly their effectiveness in the bare areas characterization.

**4.4. Efficient Configuration**

This section aims to quantify some aspects that influence the accuracy of the proposed models such as the sampling strategy and the spatial resolution. Two systematic and two stratified sampling strategies have been designed in order to better analyze the compromise between variance and cost (Table 4). The LAI measurements for each spatial scheme have been directly obtained from the LAI reference map. The analysis consists of assessing which model and spatial sampling strategy better reproduces this LAI field considered as ground truth.

The two systematic spatial samplings consist of a regular grid of ESUs separated 500 m (SP1) and 200 m (SP2), respectively (Figure 9, top). The separated distances are chosen lower than the range provided by the NDVI variogram computed for this area since this parameter agrees with the largest element size in the study area [62]. In this case, a good sampling of the main representative fields is expected. In the two stratified spatial patterns, SP3 and SP4 (Figure 9, bottom), samples are distributed...
over the study area in order to realistically capture the variability of the dominant vegetation types and conditions in the scene. Both spatial schemes differ in the number of in-field samples. The SP3 spatial pattern samples only those fields per crop that show different NDVI values. Each field is characterized by three ESUs in order to describe the spatial variability. As a result, 61 ESUs on vegetation canopies are sampled over 18 fields (e.g., 5 corn, 4 alfalfa, 4 sugar beet, 3 onion, 1 potato, and 1 garlic). The SP4 spatial scheme samples all the representative fields (i.e., 36) using a single ESU per field. Moreover, additional ESUs are included in both stratified patterns in order to solve the bare areas characterization. A total of 51 ESUs showing a NDVI value in the image lower than 0.1 has been added to the SP3 and SP4 spatial samplings. A LAI value of zero is assigned to every ESU.

Table 4. Spatial sampling schemes derived from the LAI reference map.

<table>
<thead>
<tr>
<th>Spatial sampling</th>
<th>Type</th>
<th>Regular spacing</th>
<th>Number of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP1</td>
<td>Systematic</td>
<td>200 m</td>
<td>676</td>
</tr>
<tr>
<td>SP2</td>
<td>Systematic</td>
<td>500 m</td>
<td>121</td>
</tr>
<tr>
<td>SP3</td>
<td>Stratified</td>
<td>-</td>
<td>61 (veg.)+ 51 (soil)</td>
</tr>
<tr>
<td>SP4</td>
<td>Stratified</td>
<td>-</td>
<td>36 (veg.)+ 51 (soil)</td>
</tr>
</tbody>
</table>

Figure 9. Systematic (top) and stratified (bottom) sampling schemes designed from the LAI reference map.
Figure 10 shows the anisotropy variograms representing the maximum and minimum anisotropy directions of the four spatial sampling schemes where similar continuity behavior is observed for all of them. A Gaussian model is fitted to the directional variograms where a clear anisotropy is observed around the 30°NE and 60°NW directions with range values for each direction reaching almost a distance of 1600 m and 1200 m.

**Figure 10.** Directional variogram representing the maximum (left) and minimum (right) anisotropy directions of the four spatial sampling schemes.

Figure 11 shows the LAI field predicted from the OK (top), CKC (center) and KED (bottom) models for the four spatial sampling strategies. The visual inspection of these predictions reveals a strong influence of the primary variable when comparing the OK and CKC predictions for the SP1 sampling (Figure 9 top and center). In this case, the exhaustive SP1 spatial sampling leads to a screen effect of the secondary variable in the CKC predictions as the influence of the secondary variable is not significant in the predictions. This effect is less marked for the SP2 spatial design. In this case, the secondary variable contributes to the predictions with information about the pattern at local scales. In terms of overall pattern, KED best exhibited the observed local accuracy for all the spatial samplings. The main differences observed between the predictions with the KED model refer to the intervariability of the fields constrained to the spatial distribution of the primary variable.

In order to get a deep insight into the optimal configuration, the predicted LAI fields have been compared with the LAI reference image. Table 5 shows the statistical parameters derived from this comparison. The worst results are observed for the systematic SP2 spatial sampling with the OK model showing the lowest correlation (<0.6) and highest RMSE (>1.3). The CKC model captures the spatial pattern well at local scales for all the spatial schemes except for the SP1. The KED model provides the best correspondence with the reference LAI retrievals for all the spatial samplings with a linear slope and intercept near one and zero, respectively. The high Pearson coefficient (>0.9) and decrease of the RMSE differences (<0.2) with respect to the CKC estimates, reveals the KED model as the most appropriate for LAI predictions at high-spatial resolution for the considered spatial schemes (see Table 5).
Figure 11. LAI maps derived from the OK (left), CKC (center) and KED (right) models for the four sampling schemes.

Table 5. Statistical parameters between the OK, CKC and KED predictions and the reference LAI field.

<table>
<thead>
<tr>
<th>Model</th>
<th>Slope</th>
<th>Intercept</th>
<th>$r$</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>OK</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SP1</td>
<td>0.8</td>
<td>0.2</td>
<td>0.9</td>
<td>0.7</td>
</tr>
<tr>
<td>SP2</td>
<td>0.3</td>
<td>0.8</td>
<td>0.6</td>
<td>1.3</td>
</tr>
<tr>
<td>SP3</td>
<td>0.3</td>
<td>0.8</td>
<td>0.6</td>
<td>1.3</td>
</tr>
<tr>
<td>SP4</td>
<td>0.3</td>
<td>0.8</td>
<td>0.7</td>
<td>1.2</td>
</tr>
<tr>
<td>CKC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SP1</td>
<td>1.3</td>
<td>0.01</td>
<td>0.9</td>
<td>0.8</td>
</tr>
<tr>
<td>SP2</td>
<td>0.9</td>
<td>0.18</td>
<td>0.9</td>
<td>0.5</td>
</tr>
<tr>
<td>SP3</td>
<td>0.9</td>
<td>0.18</td>
<td>0.9</td>
<td>0.5</td>
</tr>
<tr>
<td>SP4</td>
<td>0.9</td>
<td>0.1</td>
<td>0.9</td>
<td>0.4</td>
</tr>
<tr>
<td>KED</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SP1</td>
<td>0.9</td>
<td>0.04</td>
<td>0.9</td>
<td>0.13</td>
</tr>
<tr>
<td>SP2</td>
<td>0.9</td>
<td>0.07</td>
<td>0.9</td>
<td>0.2</td>
</tr>
<tr>
<td>SP3</td>
<td>0.9</td>
<td>0.1</td>
<td>0.9</td>
<td>0.18</td>
</tr>
<tr>
<td>SP4</td>
<td>0.9</td>
<td>0.1</td>
<td>0.9</td>
<td>0.2</td>
</tr>
</tbody>
</table>
However, due to the high number of samples included in the systematic designs, the stratified ones seem the most reliable in order to capture the variability of the study area. Figure 12 shows the percentage of pixels with RMSE lower than 1.0 for the three models and both the SP3 (black) and SP4 (gray) stratified samplings. Although both spatial designs show similar RMSE for the whole area and for the three models, some discrepancies are observed at pixel level. The most significant differences refer to the OK model estimates, which provide a high percentage of pixels (around 40%) with RMSE greater than 1.0 for both spatial samplings. This reveals that the OK is not suited at local scales. Substantial improvements are observed for the CKC with a percentage of pixels above 80% with RMSE lower than 1.0. In this case, differences are still observed between sampling schemes, the SP4 being more effective than the SP3. Finally, the KED seems to be less affected by the sampling design and provides a high percentage (96%) of pixels with RMSE lower than 1.0. In this case, the choice of the stratified spatial design would be subjected to the reduction of field efforts.

Figure 12. Percentage of pixels with RMSE between OK, CKC y KED with respect to the high-resolution LAI reference map lower than 1.0 for the SP3 (top) and SP4 (bottom) sampling patterns.

4.5. Spatial Resolution Influence

In this section advantages and limitations of the methods in LAI field predictions at different scales are discussed. For this purpose, the LAI fields derived from the three geostatistical methods and consideration of the four spatial samplings have been up-scaled to moderate (300 m) and coarse (1 km) spatial resolutions. These estimates have been compared with the up-scaled LAI reference map at these resolutions (see Figure 13). As it was expected the KED model does the best job for all the spatial samplings at high, medium and coarse spatial resolution followed by the CKC model. In both cases, the SP3 and SP4 patterns show a good agreement with the LAI reference estimates ($r > 0.8$ and RMSE < 0.8) as compared with the most intensive scheme SP1. One important outcome is the high correspondence between the OK and the reference estimates for the SP4 scheme at medium and coarse
spatial resolution. This result suggests that if sampling is sufficiently dense and well stratified by land cover (using for example site auxiliary information), a simple method such as OK would be sufficient to accurately derive ground-truth maps at coarse resolution. In this case, the study area could be characterized without needing a high resolution image (e.g., NDVI) and only using less than 40 ESUs on vegetation canopies. Nevertheless, there are two problems using this approach: firstly, it may be difficult to avoid sample clustering due to site access constraints; secondly, there may only be one cover class but with a large range in LAI, as in open forests.

**Figure 13.** Correlation coefficient and RMSE between the geostatistical and the LAI reference map for the four spatial sampling at high (20 m), medium (300 m) and coarse (1 km) spatial resolutions.

5. Conclusions

This paper evaluates the potentials and limitations of three geostatistical approaches, the ordinary kriging (OK), the collocated cokriging (CKC) and kriging with an external drift (KED) on the estimation of leaf area index fields. The study is focused on some crucial aspects that influence reported accuracy of the maps, such as the spatial sampling protocol, auxiliary information, and spatial resolution in the estimates.

The analysis is performed over a small agricultural area (5 × 5 km²) centered at Barrax test site in the Albacete region (Spain) where LICOR LAI-2000 field measurements were available. A high-resolution NDVI image computed from SPOT TOA reflectance data is considered as an auxiliary variable in LAI predictions. The predicted LAI fields uncertainty have been assessed using two different data sets (i) a LAI ground-truth data set derived from a digital hemispherical camera and (ii) a high-spatial resolution LAI field derived from a non-spatial method. Although this area is particularly challenging due to the discontinuity between crops, the CKC and KED predictions have proven the relevance of the auxiliary information to reproduce the spatial pattern at local scales. The KED model provides more realistic estimates on the transition between crops, where a non-stationary trend is
observed. This method offers a desirable tradeoff since it incorporates, in an effective way, the local spatial trend of the primary variable (LAI) as well as the secondary variable (NDVI) to characterize poorly sampled regions such as bare and dry areas.

The results from the efficient configuration assessment (model and spatial sampling) have demonstrated the poor influence of the auxiliary information on the estimates when the primary variable is exhaustively sampled (e.g., SP1), producing a screen effect in the auxiliary information. In this case, the OK model provides an acceptable correlation with the LAI reference field and similar to those provided by the CKC model. The KED model best exhibits the observed local accuracy for all the spatial samplings showing the stratified ones are most reliably translated to the field. Thus, the choice of the sampling scheme would be constrained to the reduction of field efforts. In this case, it seems more effective to capture the spatial variability of the area with at least one sample per field (i.e., SP4) than the inter-variability of some fields (i.e., SP3) which could be retrieved from the secondary variable.

Due to the high dependence of the heterogeneity on the estimates, an important question concerns the influence of spatial resolution on the estimates. The results from the optimal spatial configuration exercise at medium and coarse spatial resolutions have demonstrated the KED model as a suitable alternative for estimating LAI fields using the NDVI as exhaustive auxiliary information. KED estimates from very dense primary variables (e.g., scheme SP1) have not shown significant variations from a stratified sampling (e.g., scheme SP3 and SP4) with a maximum of 61 samples on vegetated canopies. Moreover, the CKC model has proven to be a second alternative when stratified sampling schemes are considered (e.g., SP3 and SP4). The main disadvantage of these models is their limitation due to the lack of availability of secondary information.

Results also suggest that the OK model, despite its deficiencies, can be successfully applied to reproduce the LAI pattern at a 1 km resolution scale if sampling is sufficiently dense and well stratified by land cover. Future work is needed to test this approach at other sites and with other sampling schemes.

Acknowledgments

This work has been partly support by DESURVEY (EC-003950), DULCINEA (CGL2005-04202), ÁRTEMIS (CGL2008-00381/CLI) and SPARC 2003 projects. Thanks are also due to the J. Moreno team for supporting the field experiment.

References and Notes


© 2010 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).