ELSEVIER

Remote Sensing of Environment

www.elsevier.com/locate/sna

Predicting the Areal Extent of Land-Cover Types Using Classified Imagery and Geostatistics

*S. de Bruin**

Remote sensing is an efficient means of obtaining large-area land-cover data. Yet, remotely sensed data are not error-free. This paper presents a geostatistical method to model spatial uncertainty in estimates of the areal extent of land-cover types. The area estimates are based on exhaustive but uncertain (soft) remotely sensed data and a sample of reference (hard) data. The method requires a set of mutually exclusive and exhaustive land-cover classes. Land-cover regions should be larger than the pixels' ground resolution cells. Using sequential indicator simulation, a set of equally probable maps are generated from which uncertainties regarding land-cover patterns are inferred. Collocated indicator cokriging, the geostatistical estimation method employed, explicitly accounts for the spatial cross-correlation between hard and soft data using a simplified model of coregionalization. The method is illustrated using a case study from southern Spain. Demonstrated uncertainties concern the areal extent of a contiguous olive region and the proportion of olive vegetation within large pixel blocks. As the image-derived olive data were not very informative, conditioning on hard data had a considerable effect on the area estimates and their uncertainties. For example, the expected areal extent of the contiguous olive region increased from 65 ha to 217 ha when conditioning on the reference sample. ©*Elsevier Science Inc., 2000*

INTRODUCTION

Current concerns about environmental changes have lead to an increased demand for land-cover data at regional to global scales (e.g., DeFries and Townshend, 1994; Vogelmann et al., 1998). Satellite remote sensing is an efficient means to obtain these data in a timely and consistent manner. Yet, remotely sensed land-cover data are not error-free, as they rely largely on the spectral responses of land-cover types that may not all be spectrally distinguishable. Data accuracy may further degrade as a result of errors in the source data and imperfect image processing. If remotely sensed land-cover data are used to evaluate environmental changes one should, therefore, account for the uncertainties in these data.

Foody et al. (1992), Maselli et al. (1994), Van der Wel et al. (1998), De Bruin and Gorte (2000), and others explored how posterior probability vectors, a by-product of probabilistic image classification, can be used to represent *local uncertainty* about class labels of individual pixels. This paper goes one step further and presents a geostatistical approach to assess *spatial uncertainty* (Goovaerts, 1997, 1999; Deutsch and Journel, 1998), that is, the joint uncertainty about land cover at several pixels taken together. This is particularly useful in regional analyses that require spatially aggregated land-cover data. Examples of these are assessments of the areal extent of land-cover types over spatial units with fixed geometry (e.g., political units or square cells) or the size of contiguous regions having one vegetation cover. Sequential indicator simulation (SIS) enables the generation of multiple maps that honor the available data and allow spatial patterns and uncertainties in the mapped land cover to be inferred. Because in SIS spatial structures are described in terms of variograms, the approach is notably different from the one proposed by Canters (1997), who used image segmentation to derive spatial structures.

Recently, Kyriakidis (1999) used SIS to map thematic classification accuracy through integration of image-reported (soft) and higher accuracy (hard) class labels. Data integration was accomplished by using simple indicator kriging with varying local means (SKlm) (Goovaerts and Journel, 1995; Goovaerts, 1997) obtained from spatially degraded classified imagery. In this study, the soft indicator data are derived from an image classifier's posterior proba-

* Wageningen UR, Centre for Geo-Information, Wageningen The Netherlands

Address correspondence to S. de Bruin, Wageningen UR, Centre for Geo-Information, P.O. Box 47, 6700AA Wageningen, AH, Netherlands. E-mail: sytze.debruin@staff.girs.wau.nl

Received 17 November 1999; revised 10 March 2000.

REMOTE SENS. ENVIRON. 74:387–396 (2000)

655 Avenue of the Americas, New York, NY 10010

0034-4257/00/$–see front matter
PII S0034-4257(00)00132-2

bility vectors. Data integration is based on a collocated cokriging approach (Almeida and Journel, 1994; Goovaerts and Journel, 1995) that, unlike SKlm, explicitly accounts for the spatial cross-correlation between hard and soft data. As a consequence, collocated cokriging estimates are potentially less influenced by sharp local contrasts in the soft data, which are very common in classified imagery (speckling).

This paper explores the use of SIS with collocated indicator cokriging to evaluate uncertainty in area estimates derived from classified remotely sensed imagery. First, the consequences of spatial uncertainty on area predictions are explained. Next, two sections briefly outline the methods of collocated cokriging of indicator data and SIS. Finally, the approach is illustrated by predicting the areal extent of a contiguous olive region around a given point and within pixel blocks covering a study area in southern Spain.

AREA PREDICTION UNDER UNCERTAINTY

An obvious way to derive area estimates over a region from remotely sensed imagery is by counting the number of pixels that have been assigned to a given land cover. Bayes' decision rule, which is sometimes referred to as *maximum likelihood rule*, assigns each pixel to the class having the largest conditional probability of membership (Duda and Hart, 1973). It typically leads to an over-representation of the most frequent class and under-representation of less frequent categories (Goovaerts, 1997). Soares (1992) developed a classification algorithm that does not have this drawback. However, if the only aim is to estimate class areas over regions that contain a large number of pixels, there is no need for class allocation altogether, provided that the conditional probability vectors are available.

The regional proportion of category s_k over a region A equals the number of pixels where s_k occurs divided by the total number (N) of pixels in A [see Eq. (1)]:

$$f(A;s_k)=\frac{1}{N}\sum_{i=1}^{N} f(\mathbf{u}_i;s_k), \tag{1}$$

where $f(\mathbf{u}_i;s_k)$ is defined by Eq. (2):

$$f(\mathbf{u}_i;s_k)=\begin{cases}1 & \text{if } s(\mathbf{u}_i)=s_k\\ 0 & \text{otherwise}\end{cases} \tag{2}$$

with $\mathbf{u}_i$ denoting the ith pixel location and $s(\mathbf{u}_i)$ being the land-cover class at $\mathbf{u}_i$. As the true category $s(\mathbf{u}_i)$ is unknown, it is modeled by the random variable (RV) $S(\mathbf{u}_i)$. Consequently, it is modeled by the RV $F(\mathbf{u}_i;s_k)$. The (conditional) expectation ($E[.]$) and variance ($Var[.]$) of each are given by Eq. (3) and Eq. (4):

$$\begin{aligned}E[F(\mathbf{u}_i;s_k)]&=1\cdot\text{Prob}\{S(\mathbf{u}_i)=s_k\}+0\cdot\text{Prob}\{S(\mathbf{u}_i)\neq s_k\}\\&=p(\mathbf{u}_i;s_k|\mathbf{x}_i)\ \text{and}\end{aligned} \tag{3}$$

$$\begin{aligned}Var[F(\mathbf{u}_i;s_k)]&=\text{Prob}\{S(\mathbf{u}_i)=s_k\}\cdot\text{Prob}\{S(\mathbf{u}_i)\neq s_k\}\\&=p(\mathbf{u}_i;s_k|\mathbf{x}_i)\cdot(1-p(\mathbf{u}_i;s_k|\mathbf{x}_i))\end{aligned} \tag{4}$$

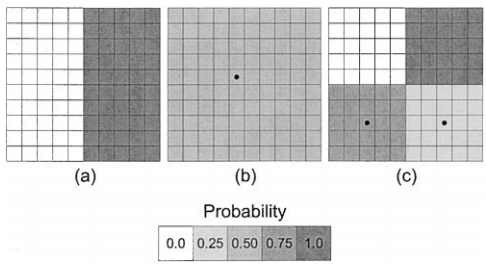


Figure 1. Pixel blocks showing conditional probabilities for a land-cover type s_k given the pixels' spectral feature vectors. In all three cases the expected proportion of s_k-covered pixels equals 0.5. The black dots in (b) and (c) indicate sample locations.

(Bernoulli distribution), where $p(\mathbf{u}_i;s_k|\mathbf{x}_i)$ is an estimate of the conditional probability for class s_k to occur at location $\mathbf{u}_i$ given the corresponding spectral feature vector $\mathbf{x}_i$. The expected regional proportion equals the sum of N expectations from Eq. (3) divided by N [see Eq. (5)]:

$$E[F(A;s_k)]=\frac{1}{N}\cdot\sum_{i=1}^{N} p(\mathbf{u}_i;s_k|\mathbf{x}_i) \tag{5}$$

Calculation of the variance of $F(A;s_k)$ is more involved though, as will be illustrated below.

Figures 1a–c represent pixel blocks of 100 pixels each. The pixels are shaded according to their values for $p(\mathbf{u}_i;s_k|\mathbf{x}_i)$. In all three cases the expected regional proportion $E[F(A;s_k)]$ equals 0.5. If the pixels were to be independent from each other, $Var[F(A;s_k)]$ would equal the sum of the variances [Eq. (4)] of the 100 individual pixels divided by 100. The results are shown in the first row of Table 1. Spatial independence, however, rarely occurs in image scenes and would seriously restrict the usefulness of remotely sensed imagery in a land resource survey.

Suppose that each gray shade in Fig. 1 represents an independent object with a homogeneous land cover (e.g., an agricultural field). Figure 1b thus represents one object ($N=1$) corresponding to an extreme case of spatial dependence of the pixels. The expectation $E[F(A;s_k)]$ still equals 0.5, but now $Var[F(A;s_k)]$ amounts to 0.25, being 100 times larger than for independent pixels (cf., Goodchild et al., 1992; Canters, 1997). This is not surprising as $F(A;s_k)$ can only take the value zero or one. On the other hand, $Var[F(A;s_k)]$ would reduce to zero if the true land cover would be sampled at the pixel locations indicated by black dots in Figs. 1b and 1c. The variance reduction in the case of independent pixels would amount to only 1% and 4%, respectively (see Table 1), since knowledge of the land cover at the sample locations would not affect the uncertainty at other locations.

The geostatistical methods presented herein after use prior models of spatial correlation to describe spatial continuity of land-cover types. They assume the existence of an exhaustive sample of soft data derived from the probability

Table 1. Variance of the Areal Proportion of Class s_k over Region A, $Var[F(A;s_k)]$, for Different Situations Indicated in Fig. 1

	Fig. 1a	*Fig. 1b*	*Fig. 1c*
(1) Independent pixels	0	2.5×10^{-3}	9.375×10^{-4}
(2) Multipixel objects	0	0.25	2.344×10^{-2}
(3) As (1), but with sampled ground truth	0	2.475×10^{-3}	9.0×10^{-4}
(4) As (2), but with sampled ground truth	0	0	0

vectors from an image classification and a relatively small sample of hard reference data. Area predictions are conditioned on both data types and on the spatial correlation models that tie the data together. The methods not only deal with area proportions within spatially confined units but also enable uncertainty in the geometry of contiguous regions of a given land cover to be modeled.

INDICATOR COKRIGING

Indicator Approach

The above example illustrates that uncertainty in area estimates from remotely sensed imagery can be considerably reduced if the estimates are conditioned on sampled ground truth (hard data). The example does not show that such conditioning involves updating the image-derived conditional probabilities. Indicator kriging provides a framework to generate posterior conditional probabilities by integrating hard and soft indicator data (Journel, 1986; Zhu and Journel, 1993; Goovaerts, 1997).

Indicator kriging of a categorical variable (e.g., land-cover class) requires that all data be coded as local prior probability values. Precise measurements of category s_k at hard data locations $\mathbf{u}_a$ are coded into a set of K binary (hard) indicator data defined as [see Eq. (6)]:

$$i(\mathbf{u}_a;s_k)=\begin{cases}1 & \text{if } s(\mathbf{u}_a)=s_k\\ 0 & \text{otherwise}\end{cases} \quad k=1,\ \ldots,\ K \tag{6}$$

These measurements are often supplemented by a large amount of indirect data, such as class probabilities conditioned on remotely sensed spectral responses. These are expressed as soft indicator data with values between 0 and 1, thereby indicating uncertainty about the actual category at the soft data location $\mathbf{u}_i$. For example [see Eq. (7)]:

$$y(\mathbf{u}_i;s_k)=p(\mathbf{u}_i;s_k|\mathbf{x}) \tag{7}$$

cf. Eq. (3).

Next, local prior probabilities are updated into posterior distributions using nearby hard and soft data. Collocated indicator cokriging is an updating procedure that incorporates exhaustively sampled soft data by using only the soft indicator datum that is collocated with the location being estimated. It has important advantages over full cokriging in that it avoids instability problems caused by highly redundant soft information and significantly simplifies modeling of spatial correlation (Almeida and Journel, 1994; Goovaerts and Journel, 1995).

Collocated Indicator Cokriging

The ordinary collocated indicator cokriging (ocICK) estimate of the posterior probability vector of a categorical variable is shown in Eq. (8):

$$[p(\mathbf{u};s_k|(n))]_{ocICK}=\sum_{a=1}^{n(\mathbf{u})}\lambda_a^{OCK}(\mathbf{u};s_k)\cdot i(\mathbf{u}_a;s_k) + \lambda_{n(\mathbf{u})+1}^{OCK}(\mathbf{u};s_k)\cdot y(\mathbf{u};s_k) \quad k=1,\ \ldots,\ K \tag{8}$$

where (n) denotes the nearby hard and the collocated soft data. Using models of spatial dependence, the weights $\lambda_a^{OCK}(\mathbf{u};s_k)$ and $\lambda_{n(\mathbf{u})+1}^{OCK}(\mathbf{u};s_k)$ are determined by solution of an ordinary cokriging (OCK) system under the unbiasedness condition [see Eq. (9)]:

$$\sum_{a=1}^{n(\mathbf{u})}\lambda_a^{OCK}(\mathbf{u};s_k)+\lambda_{n(\mathbf{u})+1}^{OCK}(\mathbf{u};s_k)=1 \tag{9}$$

(e.g., Isaaks and Srivastava, 1989; Goovaerts, 1997). Any posterior probability outside the interval [0, 1] is reset to the closest bound, zero or one. Subsequently, the estimates $p(\mathbf{u};s_k|(n))$, $k=1,\ \ldots,\ K$ are standardized by their sum to meet the condition [see Eq. (10)]

$$\sum_{k=1}^{K}p(\mathbf{u};s_k|(n))=1 \tag{10}$$

(Goovaerts, 1997; Deutsch and Journel, 1998). Note that condition in Eq. (9) guarantees unbiasedness only if the hard and soft indicator variables have the same mean within each search neighborhood.

Unlike full cokriging, solution of the OCK system by ocICK does not require a spatial-dependence model for the soft indicator data, but only for the hard indicator data and the cross-correlation between hard and soft data. Spatial dependence modeling is usually done by fitting functions through sample (cross-)semivariance values. The cross-variogram $\gamma_{IY}(\mathbf{h};s_k)$ for category s_k between a hard indicator, I, and soft indicator, Y, is computed from paired observations in a number of direction and distance classes a vector $\mathbf{h}$ apart [see Eq. (11)]:

$$\gamma_{IY}(\mathbf{h};s_k)=\frac{1}{2N(\mathbf{h})}\sum_{a=1}^{N(\mathbf{h})}[i(\mathbf{u}_a;s_k)-i(\mathbf{u}_a+\mathbf{h};s_k)] \cdot[y(\mathbf{u}_a;s_k)-y(\mathbf{u}_a+\mathbf{h};s_k)] \tag{11}$$

where $N(\mathbf{h})$ is the number of data pairs in the class of distance and direction. Though not used in this study, indicator cross-variograms can also be computed between indicators of different categories $s_{k'}\neq s_k$. The indicator variogram is computed by substituting $i(.)$ for $y(.)$ in Eq. (11).

The value $2\gamma_I(\mathbf{h};s_k)$ indicates how often two locations a vector $\mathbf{h}$ apart belong to different categories $s_{k'}\neq s_k$ (Goovaerts, 1997, 1999). The linear model of (co)regionalization is used to ensure positive definiteness of the covariance matrix in the kriging system (e.g., Isaaks and Srivastava, 1989; Goovaerts, 1997; but see Yao and Journel, 1998).

Since ocICK requires the covariance of the soft indicator data only at $\mathbf{h}=0$, the only constraint the linear model of coregionalization must satisfy is that [see Eq. (12)]

$$|sill[\gamma_{IY}(\mathbf{h};s_k)]|\leq\sqrt{sill[\gamma_Y(\mathbf{h};s_k)]\cdot sill[\gamma_I(\mathbf{h};s_k)]} \qquad (12)$$

(Goovaerts, personal communication), where $sill[.]$ denotes the semivariance for distances larger than the range (i.e., the distance where the variogram levels off). Modeling can be further simplified using a Markov-type assumption, which states that dependence of the soft indicator on the hard indicator is limited to the collocated hard indicator datum (Zhu and Journel, 1993; Almeida and Journel, 1994; Goovaerts, 1997). The cross-variogram between hard and soft indicator data, $\gamma_{IY}(\mathbf{h};s_k)$, is then inferred directly from $\gamma_I(\mathbf{h};s_k)$, using a coefficient of proportionality obtained from calibrating the soft data to the hard data. The validity of this approximation must be checked (see e.g., Goovaerts and Journel, 1995). Note that ocICK with a Markov coregionalization model is equivalent to ordinary kriging of the residuals when the drift given by the cross-correlation coefficient between hard and soft indicator data has been subtracted (Coléou, 1999).

SIS

The posterior probability estimates $p(\mathbf{u};s_k|(n))$ computed by indicator kriging model the *local* uncertainty about the category that occurs at each interpolated location. As opposed to the kriging variance, which is independent of data values (e.g., Goovaerts, 1997, 1999), measures derived from these probability vectors reflect the uncertainty that is due to both data geometry and data values. Regional analyses, however, often require spatially aggregated data. This implies that local uncertainties must be combined to reflect joint uncertainty at several locations taken together. Such spatial uncertainty can be modeled by stochastic simulation (i.e., generating multiple equiprobable realizations of the joint distribution of attribute values in space) (Zhu and Journel, 1993; Journel, 1996; Goovaerts, 1997, 1999).

Simulation of multiple realizations of a categorical variable can be performed using SIS. Such simulation proceeds as follows (Gómez-Hernández and Srivastava, 1990; Goovaerts, 1997; Deutsch and Journel, 1998; Kyriakidis, 1999):

1. Define a random path through all nodes (pixels) to be simulated, visiting each node only once;
2. At each node $\mathbf{u}$ along this random path:
   (a) Determine the posterior probability $p(\mathbf{u};s_k|(n))$ for each category s_k, $k=1,\ldots,K$, conditional to the neighboring hard and soft indicator data, for example using ocICK [Eq. (8)].
   (b) Generate a value $s^{(l)}(\mathbf{u})=s_k^{(l)}$ via Monte Carlo sampling of the above distribution. The simulated value is added to the conditioning data set to be used as a hard datum in all subsequent determinations;
3. Move to another node along the random path and repeat step 2.

The realization is completed when all nodes have been given a simulated value.

The set of realizations generated by SIS provides an uncertainty model of the spatial distribution of (categorical) attribute values. Spatial features, such as contiguous nodes (pixels) assigned to the same category, are considered certain if seen in all realizations. Conversely, features are deemed uncertain if seen only on a few simulated maps. Returning to the problem of area prediction referred to previously, this model of spatial uncertainty can be used to assess uncertainty in area estimates derived from remotely sensed imagery. This will be demonstrated below.

CASE STUDY

Study Area, Data, and Methods

The case study concerns part of the drainage basin of the river Guadalhorce in the province of Malaga, southern Spain. The area is approximately 110 km^2 in extent and is centered around the village of Alora. The major part of the study area is covered by digital color orthophotography derived from aerial photographs taken in 1996. The latter were supplied by the Instituto de Cartografía de Andalucía. De Bruin and Gorte (2000) did a land-cover classification of the study area using 1995 Landsat Thematic Mapper (TM) imagery. The classification scheme distinguished 10 mutually exclusive and exhaustive land-cover classes. The per-pixel class membership probabilities conditional to the remotely sensed spectral responses were stored to enable further analyses of local classification uncertainty. Here, in the first instance, we will consider the three main crop types in the area: citrus fruits, arable crops, and olive. Later, attention is focused on the olive crop.

Hard land-cover indicator data [Eq. (6)] were collected by visual interpretation of the digital color orthophotography. First, an equilateral triangular grid with a spacing of 420 m was laid over the area having orthophoto coverage. At each grid node the land-cover category was determined within a square cell of 900 m^2. The cells precisely matched ground resolution cells of the geo-referenced 1995 Landsat TM image. Only cells in which a unique land-cover category could be clearly identified were retained (514 cells). Subsequently, the grid was densified for improved variogram estimation at short distances. The locations of 200 additional sample points were optimized using spatial simulated annealing (Van Groenigen and Stein, 1998). The ob-

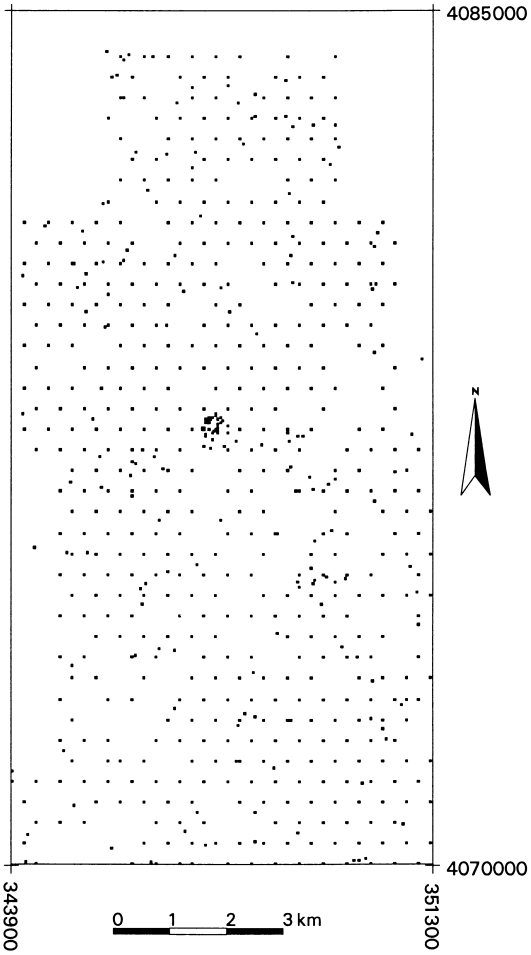


Figure 2. Locations of the 688 land-cover samples. Coordinates (m) correspond to UTM zone 30.

jective was to have at least 100 point pairs in distance class 90 m to 180 m and 400 point pairs in distance class 180 m to 270 m, in each of two direction classes (0±45° and 90±45°). Five points were lost because they were positioned within a cell that was also sampled by another point. In another 21 cells the land cover could not be properly determined. The total reference set thus amounted to 688 cells with high-accuracy (hard) land-cover data (Fig. 2).

The image-derived land-cover class probabilities (De Bruin and Gorte, 2000) were calibrated against the hard indicator data by means of logistic regression. This was done to approximate equality of the stationary means of hard and soft indicator data, and thus validity of unbiasedness condition in Eq. (9). The thus-transformed class probabilities served as soft indicator data in all subsequent analyses. Indicator variogram modeling for the three main crop types was done using GSTAT 2.0 (Pebesma, 1998; Pebesma and Wesseling, 1998). The Markov coregionalization model was used to infer the cross-variograms between hard and soft indicator data. The resulting models were visually checked against sample cross-variogram values. If the Markov approximation was inappropriate, constraint [Eq. (11)] was used to fit a linear model of coregionalization.

The error matrix of the 1995 classification (De Bruin and Gorte, 2000), which was prepared from 87 homogeneous multipixel reference sites, illustrates the difficulty of correctly classifying olive from remotely sensed imagery. The class had 65% omission errors and included 39% false commissions as a result of spectral confusion with other land-cover classes. Therefore, the olive class was selected to demonstrate the effect of using hard data in geostatistical estimation and simulation.

The expanded GSLIB cokriging program *newcokb3d* (Ma and Journel, 1999) was used to implement ocICK for estimating local probabilities of the occurrence of olive. Sequential indicator simulation with ocICK was performed using the GSLIB program *sisim*. The latter program was modified to enable the use of linear models of coregionalization. Estimates of the spatial uncertainty about the presence or absence of olive vegetation were obtained from 500 SIS realizations both with and without conditioning on the hard indicator data.

RESULTS

Figure 3 shows the indicator (cross-)variograms for the three main crop types in the study area. The continuous curves in the upper three plots (Figs. 3a–c) were obtained by fitting positive linear combinations of spherical functions through the sample semivariances. The indicator variogram for citrus (Fig. 3a) is anisotropic (i.e., the pattern of spatial connectivity changes with direction; the axis of greatest spatial continuity being in a north–south direction). The solid lines in Figs. 3d to 3f are Markov models of the indicator cross-variograms $\gamma_{IY}(\mathbf{h};s_k), s_k=$*citrus*, *arable*, *olive*. The models show good correspondence with the experimental data for citrus and arable crops, but the approximation does not fit the olive data. The cross-variogram $\gamma_{IY}(\mathbf{h};$*olive*$)$ cannot be considered as being proportional to $\gamma_I(\mathbf{h};$*olive*$)$. A better fit, obtained with a linear model of coregionalization, is shown in Fig. 4. This model puts the most weight (73%) on the relatively unimportant long-range component (1350 m) of $\gamma_I(\mathbf{h};$*olive*$)$. Besides having low overall predictive ability (see above), the image-derived soft indicator data particularly fail to detect short-range variations in the presence or absence of olive vegetation.

As can be observed in Fig. 5, conditioning on hard olive indicator data has a considerable effect on the estimated local probabilities of occurrence. In the neighborhood of hard indicator data, the short range variability of the kriging estimates (Fig. 5b) is lower than that of the image-derived probabilities (Fig. 5a). At the same time the local uncertainty is lower. Beyond the range of influence of the hard indicator data (see Fig. 2), Figs. 5a and 5b are identical.

The effect of the hard indicator data on estimating spatial uncertainty is even more pronounced. Figure 6

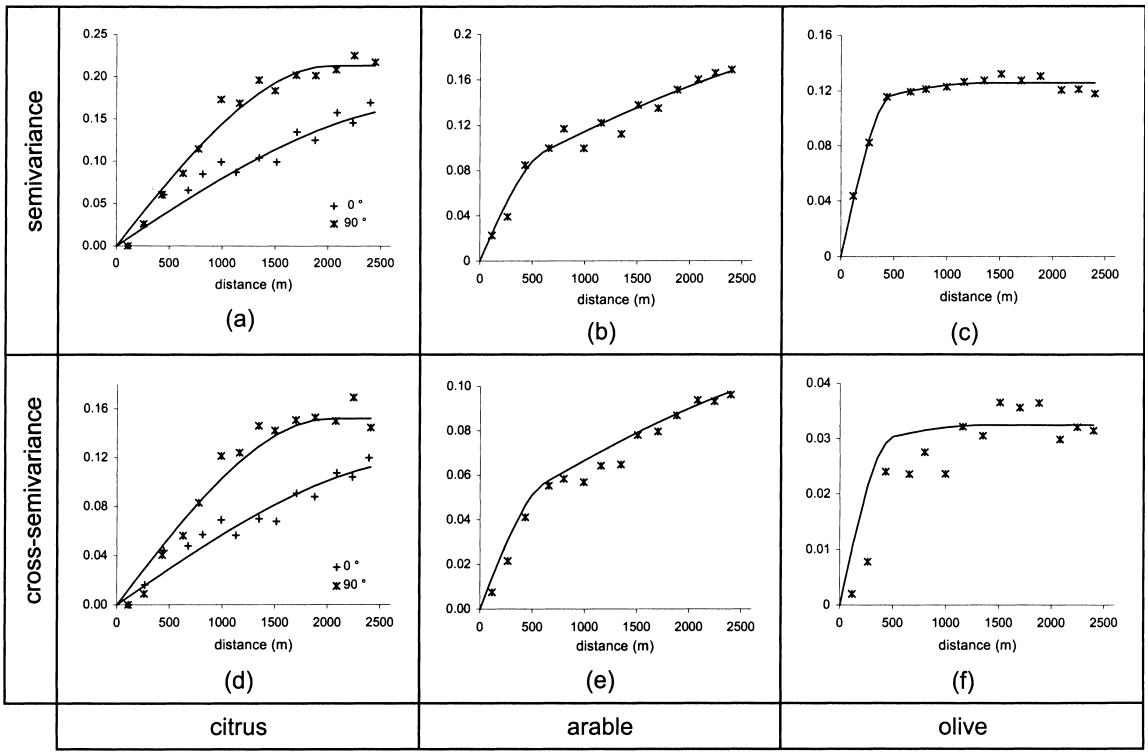


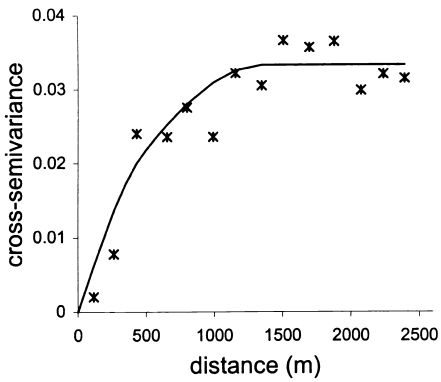
Figure 3. Experimental indicator (cross-)variograms (symbols), fitted variogram models (a–c), and Markov models of the indicator cross-variograms (d–f) for the three main crop types in the study area.

illustrates some results of 500 SIS realizations (247×500 pixels each), conditioned on both the hard and soft indicator data. The attribute of interest concerns the area of a contiguous olive-covered region around one of the sample locations (point #213). This location was known to be covered by olive vegetation. The area estimate is subject to spatial uncertainty because it depends on the land cover at multiple locations taken together. Therefore, it cannot be directly calculated from a probability field (e.g., Fig. 5b), but an approximate answer can be obtained from the statistics of a set of equiprobable realizations. The results of the multiple SIS computations are summarized in a histogram (Fig. 6a) and a cumulative distribution graph (Fig. 6b) of the simulated area. Olive-labeled pixels were considered connected if they were within the immediate 8-pixel neighborhood (eight nearest neighbors) of each other. The mean area amounted to 217 ha and the (sample) variance was 7,638 ha^2. However, the latter figure is of little practical value since the area distribution exhibits bimodality with distinct peaks around 150 ha and 330 ha. This bimodality is caused by two regions being connected or not in the individual simulations.

The SIS computations were repeated without conditioning on the 688 hard indicator data. The results are summarized in Figs. 7a and 7b. The mean area and variance now amounted to 65 ha and 3,513 ha^2, respectively. The area distribution has a high peak at 0 ha and a second, lower peak around 70 ha. The first peak is due to uncertainty about the land cover at location #213 itself. In 31% of the simulations it was classified as not having olive vegetation. The difference with the distribution of Fig. 6 is a consequence of the absence of hard indicator data that via the model of coregionalization relate the uncertain image-derived data to locations having known land cover.

For comparison, the original 1995 land-cover classification (De Bruin and Gorte, 2000) would have reported a contiguous olive region of 80 ha around point #213. Although the error matrix of that classification indicates the difficulty of correctly classifying olive vegetation, it is not clear how this uncertainty affects the area estimate.

Figure 4. Experimental indicator cross-variogram for olive (symbols) and linear model of coregionalization fitted under constraint [Eq. (11)].



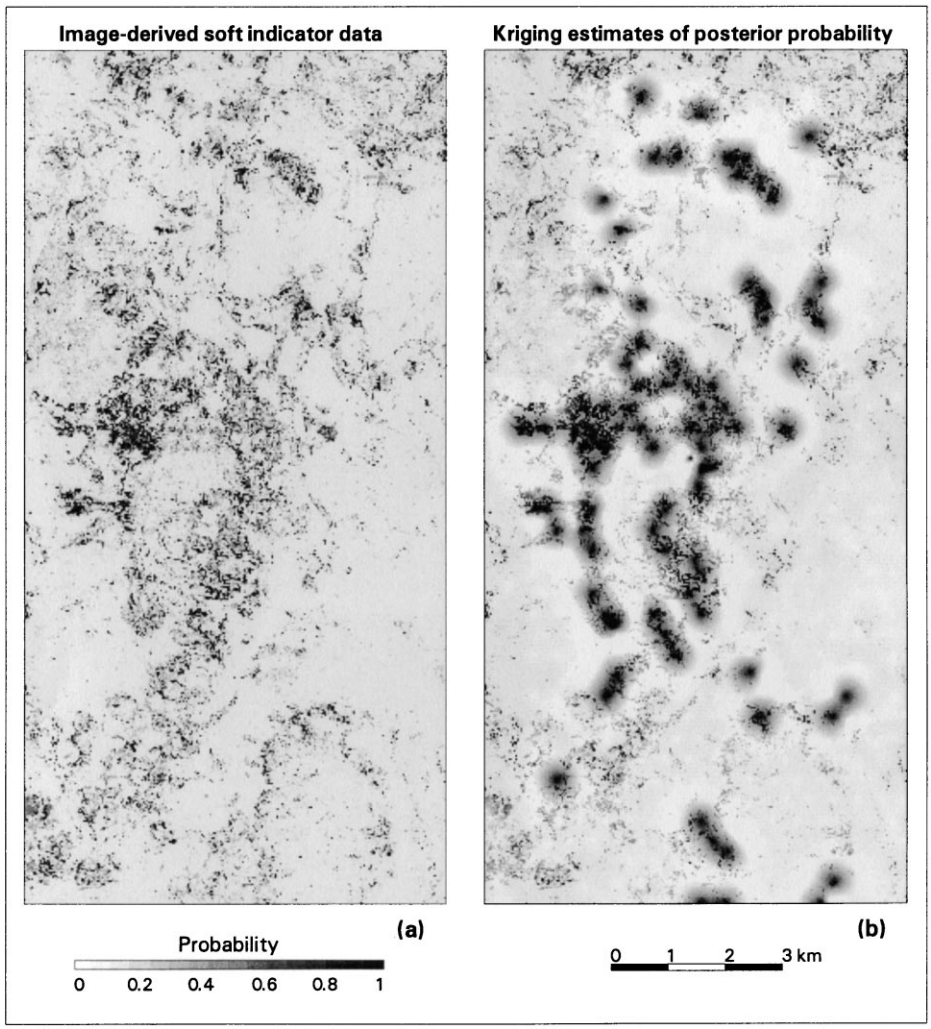


Figure 5. Image-derived soft indicator data for olive (a) and ocICK estimates of the local probabilities of occurrence conditional to the nearby hard and the collocated soft indicator data (b).

Assessment of uncertainty in area estimates requires spatial error modeling and cannot be based on global measures derived from an error matrix.

As indicated earlier, area predictions over spatial units with fixed geometry also involve spatial uncertainty. Figure 8 shows estimates of the proportions of olive vegetation and their variances in square pixel blocks of 100 pixels (9 ha) each. The former were calculated as block averages of the ocICK posterior probability estimates $p(\mathbf{u};olive|(n))$ shown in Fig. 5b. Alternatively, they could have been obtained from some form of block indicator kriging (Isaaks and Srivastava, 1989; Goovaerts, 1997; Deutsch and Journel, 1998). The variances were calculated from the 500 maximally conditioned SIS realizations. Note that unlike block kriging variances, these conditional variances reflect the uncertainty that is due to both data geometry and data values.

CONCLUSIONS

This paper presents a geostatistical method to model uncertainties in image-derived estimates of the areal extent of land-cover types. These uncertainties have a spatial character and may concern, for example, the size of land-cover regions (e.g., habitats) or the proportion of land-cover types within spatial units (e.g., pixel blocks). The area estimates are based on exhaustive but uncertain (soft) remotely sensed data and a sample of exact (hard) data. The latter data are particularly important if the image-derived data are not very informative. Collocated indicator cokriging allows the updating of soft probability data using a simplified model of coregionalization between hard and soft data. A Markov-type assumption may further alleviate the modeling efforts. The case study, however, demonstrated that the Markov approximation does not always fit the experimental cross-variogram. Sequential indicator simulation enables the generation of a set of alternative equiprobable maps from which uncertainties regarding land-cover patterns can be inferred. The method can be implemented using public domain software (Deutsch and Journel, 1998; Pebesma, 1998; Ma and Journel, 1999).

Assessment of uncertainty about land cover is rarely a goal in itself. More often, the variable of interest is an ecological response variable that ultimately may be used in developing land-use policies. Estimates of the uncertainty in such a variable can be obtained by using multiple

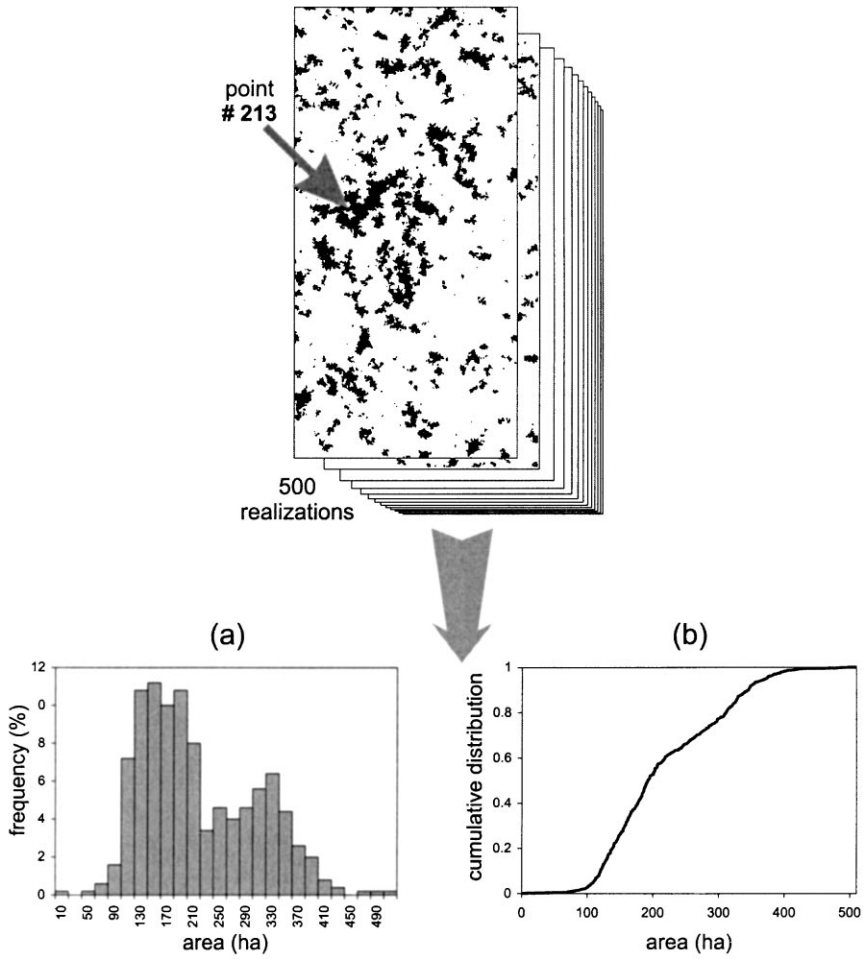


Figure 6. Histogram (a) and cumulative distribution (b) of the area of a contiguous region with olive vegetation (around sample #213). The distribution was calculated from 500 SIS realizations; all conditioned on nearby hard and collocated soft indicator data.

SIS-generated land-cover maps as input to ecological response models. The uncertainty estimates thus obtained can then be used in risk-based policy (Goovaerts, 1999; Kyriakidis, 1999).

The indicator approach presented in this paper requires the land-cover regions to be considerably larger than the pixels' ground resolution cells (*H*-resolution; Strahler et al., 1986). In the opposite case, it may be relevant to model vegetation quantities as continuous variables so that an approach similar to that proposed by Dungan (1998) could be adopted. Alternatively, it may be more appropriate to model mixtures of discrete land-cover classes, in which case geostatistical estimation could be performed using some form of compositional kriging (De Gruijter et al., 1997).

The method requires an exhaustive set of mutually exclusive land-cover classes (e.g., olive vs. nonolive). Uncertainty is due to incomplete data about the true land-cover type but does not concern the class definitions; these should be clear-cut. If the latter is not the case, the concept of expected membership in a fuzzy set can be used to

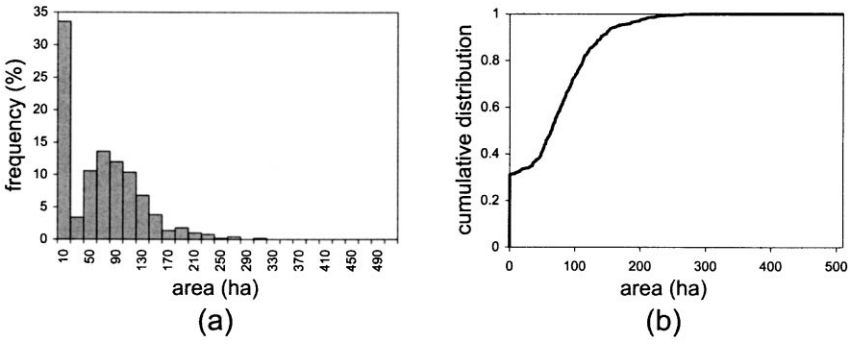


Figure 7. Histogram (a) and cumulative distribution (b) of the area of a contiguous region with olive vegetation (around sample #213). The distribution was calculated from 500 semiconditional SIS realizations (i.e., without considering the hard indicator data).

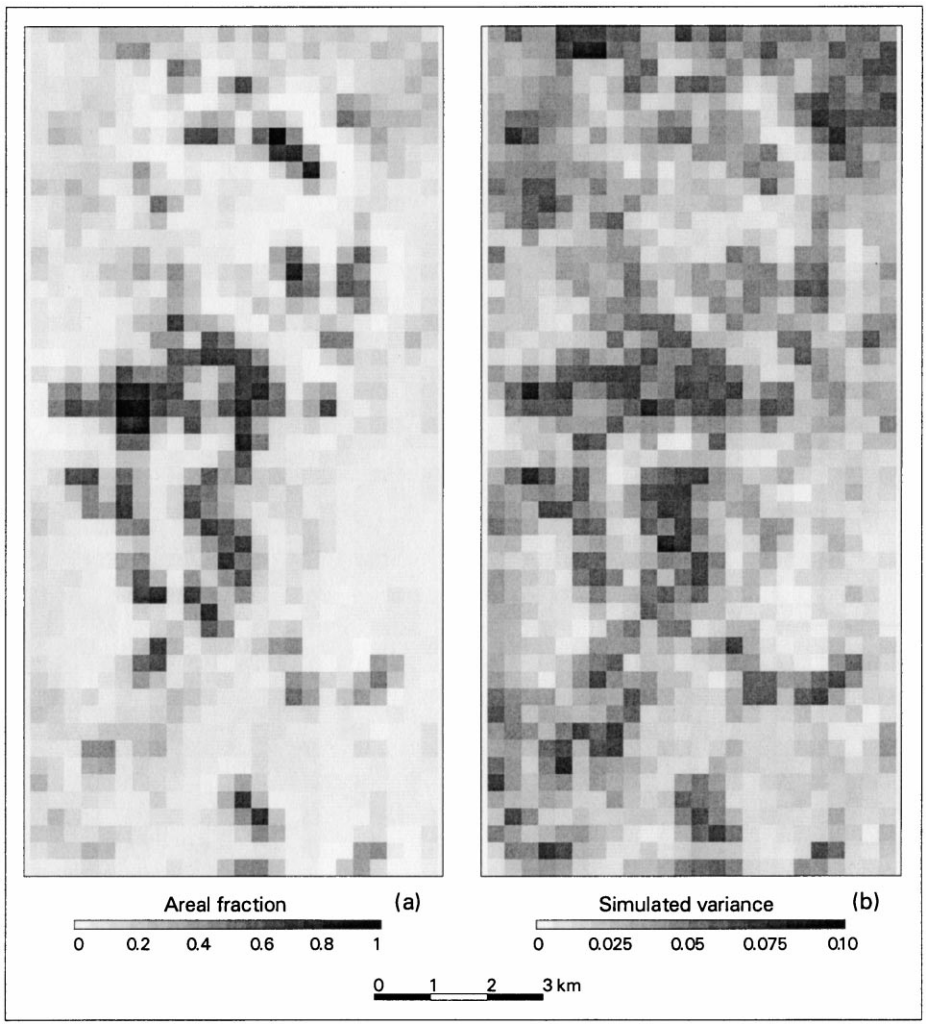


Figure 8. Estimates of the proportions of olive vegetation (a) and their variances (b) in pixel blocks of 10×10 pixels (9 ha).

combine uncertainty about values of random variables with uncertainty about the class intentions. Examples of such a combination of fuzziness and probabilistic uncertainty in areas other than land-cover mapping have been reported by Lark and Bolam (1997) and De Bruin (2000).

Finally, uncertainty about a spatial phenomenon always depends on the decisions made to model that phenomenon. Important decisions in the present study are the stationarity decision, which is very common in geostatistics, and the models of spatial dependence. Model decisions may be inappropriate and sometimes difficult to validate. Although cross-validation or validation against a separate evaluation set may be helpful, they also suffer from severe restrictions. Hence the importance of clearly documenting all aspects of the model. (Goovaerts, 1997). The frequent assumption of independent pixels obviously impedes proper assessment of spatial uncertainty in image-derived area estimates. Using well-documented geostatistical methods, the modeling alternative presented in this paper exploits spatial dependence rather than ignoring it.

The author is grateful to Jaap de Gruijter, who gave helpful advice in an early stage of this research. He, Arnold Bregt, Martien Molenaar, and Alfred Stein are acknowledged for commenting on an earlier version of the manuscript.

REFERENCES

Almeida, A. S., and Journel, A. G. (1994), Joint simulation of multiple variables with a Markov-type coregionalization model. *Mathematical Geology* 26:565–588.

Canters, F. (1997), Evaluating the uncertainty of area estimates derived from fuzzy land-cover classification. *Photogramm. Eng. Remote Sens.* 63:403–414.

Coléou, T. (1999), Links between external drift, Bayesian kriging, collocated cokriging. CGG-Petrosystems Hints, http://www.cgg.com/software/pts/hints/KR1.html, accessed 9-07-99.

De Bruin, S. (2000), Querying probabilistic land cover data using fuzzy set theory. *Int. J. Geographical Information Science* (in press).

De Bruin, S., and Gorte, B. G. H. (2000), Probabilistic image

classification using geological map delineations applied to land cover change detection. *Int. J. Remote Sens.* (in press).

De Gruijter, J. J., Walvoort, D. J. J., and Van Gaans, P. F. M. (1997), Continuous soil maps—A fuzzy set approach to bridge the gap between aggregation levels of process and distribution models. *Geoderma* 77:169–195.

DeFries, R. S., and Townshend, J. R. G. (1994), NDVI-derived land cover classifications at a global scale. *Int. J. Remote Sens.* 15:3567–3586.

Deutsch, C. V., and Journel, A. G. (1998), *GSLIB Geostatistical Software Library and User's Guide,* 2d ed., Oxford University Press, New York.

Duda, R. O., and Hart, P. E. (1973) *Pattern Classification and Scene Analysis*, John Wiley & Sons, New York.

Dungan, J. L. (1998), Spatial prediction of vegetation quantities using ground and image data. *Int. J. Remote Sens.* 19:267–285.

Foody, G. M., Campbell, N. A., Trodd, N. M., and Wood, T. F. (1992), Derivation and application of probabilistic measures of class membership from the maximum-likelihood classification. *Photogramm. Eng. Remote Sens.* 58:1335–1341.

Gómez-Hernández, J. J., and Srivastava, R. M. (1990), ISIM3D: An ANSI-C three-dimensional multiple indicator conditional simulation program. *Computers and Geosciences* 16:395–440.

Goodchild, M. F., Sun, G., and Yang, S. (1992), Development and test of an error model for categorical data. *Int. J. Geographical Information Systems* 6:87–104.

Goovaerts, P. (1997), *Geostatistics for Natural Resources Evaluation*, Oxford University Press, New York.

Goovaerts, P. (1999), Geostatistics in soil science: State-of-the-art and perspectives. *Geoderma* 89:1–45.

Goovaerts, P., and Journel, A. G. (1995), Integrating soil map information in modelling the spatial variation of continuous soil properties. *European Journal of Soil Science* 46:397–414.

Isaaks, E. H., and Srivastava, R. M. (1989), *Applied Geostatistics*, Oxford University Press, Oxford.

Journel, A. G. (1986), Constrained interpolation and qualitative information—The soft kriging approach. *Mathematical Geology* 18:269–286.

Journel, A. G. (1996), Modelling uncertainty and spatial dependence: Stochastic imaging. *Int. J. Geographical Information Systems* 10:517–522.

Kyriakidis, P. C. (1999), Stochastic imaging for assessing the impact of imprecise spatial information on ecological models. Conference on Spatial Statistics for Production Ecology, April 19–21, 1999, Wageningen.

Lark, R. M., and Bolam, H. C. (1997), Uncertainty in prediction and interpretation of spatially variable data on soils. *Geoderma* 77:263–282.

Ma, X., and Journel, A. G. (1999), An expanded GSLIB cokriging program allowing for two Markov models. *Computers and Geosciences* 25:627–639.

Maselli, F., Conese, C., and Petkov, L. (1994), Use of probability entropy for the estimation and graphical representation of the accuracy of maximum likelihood classifications. *ISPRS J. Photogramm. Remote Sens.* 49:13–20.

Pebesma, E. J. (1998), Gstat user's manual. Utrecht University, Department of Physical Geography, http://www.geog.uu.nl/gstat/.

Pebesma, E. J., and Wesseling, C. G. (1998), GSTAT: A program for geostatistical modelling, prediction and simulation. *Computers and Geosciences* 24:17–31.

Soares, A. (1992), Geostatistical estimation of multiphase structures. *Mathematical Geology* 24:149–160.

Strahler, A. H., Woodcock, C. E., and Smith, J. A. (1986), On the nature of models in remote sensing. *Remote Sens. Environ.* 20:121–139.

Van der Wel, F. J. M., Van der Gaag, L. C., and Gorte, B. G. H. (1998), Visual exploration of uncertainty in remote sensing classification. *Computers and Geosciences* 24:335–343.

Van Groenigen, J. W., and Stein, A. (1998), Constrained optimization of spatial sampling using continuous simulated annealing. *J. Environmental Quality* 27:1078–1086.

Vogelmann, J. E., Sohl, T. L., Campbell, P. V., and Shaw, D. M. (1998), Regional land cover characterization using Landsat Thematic Mapper data and ancillary data sources. *Environmental Monitoring and Assessment* 51:415–428.

Yao, T., and Journel, A. G. (1998), Automatic modeling of (cross) covariance tables using fast Fourier transform. *Mathematical Geology* 30:589–615.

Zhu, H., and Journel, A. G. (1993), Formatting and integrating soft data: Stochastic imaging via the Markov-Bayes algorithm. In *Geostatistics Tróia '92,* Vol. 1 (A. Soares, Ed.), Kluwer Academic Publishers, Dordrecht, pp. 1–12.