Bayesian and Geostatistical Approaches to Combining Categorical Data Derived from Visual and Digital Processing of Remotely Sensed Images

ZHANG Jingxiong LI Deren

ABSTRACT This paper seeks a synthesis of Bayesian and geostatistical approaches to combining categorical data in the context of remote sensing classification. By experiment with aerial photographs and Landsat TM data, accuracy of spectral, spatial, and combined classification results was evaluated. It was confirmed that the incorporation of spatial information in spectral classification increases accuracy significantly. Secondly, through test with a 5-class and a 3-class classification schemes, it was revealed that setting a proper semantic framework for classification is fundamental to any endeavors of categorical mapping and the most important factor affecting accuracy. Lastly, this paper promotes non-parametric methods for both definition of class membership profiling based on band-specific histograms of image intensities and derivation of spatial probability via indicator kriging, a non-parametric geostatistical technique.

KEY WORDS Bayesian; remote sensing image; visual and digital processing

CLC NUMBER TP751.2

Introduction

Thematic mapping of land cover and land use based on remote sensing is fundamental to many applications concerning the enumeration of natural resources and environmental modeling at a range of scales\(^1\)\(^2\)\). The research has carried out on utilizing data of multiple sources, such as field observations, historical data, and various digital maps stored in geographic information systems, in addition to remotely sensed data of multiple spatial, spectral, and temporal resolutions\(^3\)\(^4\)\(^5\)\(^6\)\(^7\)\(^8\)\(^9\)\(^10\)\(^11\)\).

The discussion in this paper is focused on utilizing data from multiple sources rather than capitalizing on different classification rules on the same data set\(^12\)\). For multi-source data combination, there are a variety of choices. At the most generalized level, class labels are pooled with the most frequent labels being retained in the final output maps. At the intermediate level, class probabilities are derived from individual sources, and then are combined in a certain way to generate combined vectors of probabilities. At the most detailed level, original data such as photographic intensities and spectral digital numbers may be combined to form stacks of vectors before they are subjected to further analysis. This paper works at the intermediate level through probabilistic categorical data, which provide a basis for handling uncertainty in categorical mapping, as elaborated in References \(^13\)\(^-\)\(^19\)\).

There are two types of data in remote sensing-based categorical mapping. One is classified samples that are usually irregularly and sparsely located but of high accuracy. Such data may be obtained from field surveys, historical maps, and interpretation of large-scale aerial photographs. The other type is comprised of digitally processed data, which are often in regular raster format but may be lower in accuracy than the...
former type. The first type of data is usually employed by data analysts to train image classifiers or guide clustering, which generates large quantity of data of the second type.

1 Combining spatial categorical data

Consider a field over the domain $D$, where one of $c$ (being the total number of candidate classes) mutually exclusive and exhaustive classes $\{U_1, \cdots, U_c\}$ is possible at each location. It is generally not known beforehand which one of the $c$ classes exists at a location. Thus, at any location $x \in D$, a random variable $U(x)$ is defined, taking any of the values from the set $\{U_1, \cdots, U_c\}$. Given a particular set of knowledge and observational data concerning domain $D$, it is possible to assign discrete classes to individual locations, leading to a classified or realized field of $U(x)$ denoted by $\{u(x), x \in D\}$.

For a probabilistic extension of the random categorical field $U(x)$, define the occurrence probability $p_k(x) = \text{prob}(U(x) \in U_k)$, for individual classes $k = 1, \cdots, c$. It is useful to construct a probability vector field $\{p_1(x), \cdots, p_c(x) | x \in D\}$, whose elements are evaluated as class occurrence probabilities for location $x$. Each component of a probabilistic vector of dimension $c$ is analyzed in turn as usually or together as compositional data.

1.1 Bayesian methods

Multivariate normal statistical theory describes the probability that an observation vector $Z(x)$ of dimension $b$ (i.e., number of spectral bands) at location or pixel $x$ will occur, given that it belongs to a class $U_k (k = 1, \cdots, c)$ and as there is the following function:

$$
p_k(z(x)) = (2\pi)^{-b/2} \det | \text{cov}_Z |^{-1/2} \exp\left(-\frac{\text{dis}^2}{2}\right)$$  

(1)

where $p_k(z(x))$ is the probability of occurrence of the measurement vector $z(x)$, $\text{cov}_Z$ is the variance and covariance matrix of variable $Z$ as estimated from training data, and $\text{dis}^2$ denotes the squared Mahalanobis distance between an observation $z(x)$ and the means of class-specific $Z$ variable $m_{Z,k}$:

$$
\text{dis}^2 = (z(x) - m_{Z,k})^T \text{cov}^{-1}_Z (z(x) - m_{Z,k})
$$  

(2)

Assuming that prior probability, $p(U_i)$ or simply $p_i$, is available, which states the probability that an observation will be drawn from class $U_i$. The posteriori probability of class $U_i$ conditional to data $z(x)$ is:

$$
p(U_i | z(x)) = \frac{p(U_i) p_i(z(x))}{p(z(x))} = \frac{p_i(z(x))}{p(z(x))}$$  

(3)

where $p(z(x))$ sums $p(U_i, z(x))$ across all classes. It is usual practice to assign observation $z(x)$ to class $U_i$ that has the highest probability of occurrence given the $b$-dimensional vector $z(x)$ observed. Thus, a posterior probability of class membership in a Bayesian-type classifier is based not only on the resemblance of a pixel to the class signature, but also on prior weighting of output classes based on their anticipated sizes.

While Gaussian distributions, such as that specified in Eq. (1), are reasonable for data analysis, it might be more flexible for using non-parametric approaches to building density functions from the data. A histogram graphing relative frequencies on the ordinate versus values or interval bounds on the abscissa may be exploited for deriving spectral signatures specific to candidate classes, so that probabilistic membership values $p_i(z(x))$ are estimated from histograms built upon training data.

Given certain multi-source data, it is possible to denote data-specific membership functions as vectors $p\rho(U_i | z'(x)) (k=1, \cdots, c)$, where $\rho$ denotes a particular data sources ($\rho = 1, \cdots, T$, where $T$ is the number of data sources available). Assuming class-conditional independence among data sources, a product role has been proved:

$$
p(U_i | \text{combined data}) = \prod_{\rho=1}^{T} p\rho(U_i | z'(x)),
$$  

(4)

by which the combined classifier is to assign a membership to a class for which $p(U_i | \text{combined data})$ turns out the greatest. Note that class probabilities calculated by Eq. (4) should be normalized by their sum over all candidate classes. This product possesses a form of consensus agreement among the classifiers in the
sense that the combined classifier is unlikely to assign a predominant membership to particular class if just one classifier says so. Eq. (3) is just a special case for Eq. (4) with two sets of probabilistic categorical data.

A criticism against the product rule set forth in Eq. (4) is about the lack of mechanism by itself to handle varying levels of uncertainty in different data sources, although some exponents were advised to weight sources of data according to their comparative reliability\(^{26-27}\). The geostatistical approaches to be described later offer some adaptive and data-drive solution to the question of weighing sources of data in the light of spatial covariance and cross-covariance.

### 1.2 Geostatistical alternatives

Again, suppose \( c \) mutually exclusive classes can be found over a domain and \( n \) classified samples are available. Each observation is classified as a member of one of the possible classes \( \{ U_1, \cdots, U_c \} \). An important formalism for categorical data lies in the transformation of the classes \( U(x) \) at the sample location \( x \) into binary indicators:

\[
i(U_k;x) = \begin{cases} 1 & \text{if } u(x) \in U_k \\ 0 & \text{if } u(x) \notin U_k \end{cases}, k = 1, \cdots, c. \tag{5}\]

For a categorical variable, the expectation of its indicator transform leads to the direct evaluation of the conditional probability that a certain class \( U_k \) prevails at location \( x \) as:

\[
E(i(U_k;x)) = \frac{1}{n} \sum_{s=1}^{n} i(U_k;x_s) = \sum_{s=1}^{n} \lambda_i(U_k(x_s)) = p_k(x), k = 1, \cdots, c. \tag{6}\]

Indicator kriging performs the estimation of the indicator transform, that is, the probabilities of finding individual classes \( U_k \) (\( k = 1, \cdots, c \)) at a point \( x \) using:

\[
p_k(x) = i(U_k;x) = \sum_{i=1}^{c} \lambda_i(i(U_k;x_s) - p_k) \tag{7}\]

where \( p_k \) is the marginal probability of class \( U_k \) inferred from sample data, \( i(U_k;x_s) \) represents the indicator transform of a sample point \( x_s(s = 1, 2, \cdots, n) \) to class \( U_k \), \( \lambda_i \) is the weight associated with the sample point \( x_s \), which are, in turn, determined by:

\[
\sum_{i=1}^{c} \lambda_i \text{cov}_{i}(x_s, x_j) = \text{cov}_{i}(x, x_j), \text{for } i = 1, \cdots, n \tag{8}\]

where \( \text{cov}_{i}(x_s, x_j) \) elements are covariance of indicator variable \( i \) between sampled locations \( x_s \) and \( x_j \), while \( \text{cov}_{i}(x, x) \) stands for covariance between location \( x \) and unsampled location \( x_s \), as implemented in Reference [28].

Further value of geostatistics lies in another technique known as co-kriging, which is applied for combining information from ground-based sampling and remote sensing in Reference [29]. Instead of incurring a rather complicated full co-kriging solution built upon two sets of all \( c \) indicator variables, consider an individual category, such as \( U_k(k = 1, \cdots, c) \), at a time. Class probabilities (of \( U_k \)) for primary and secondary sources are simplified as \( p_i^1(x) \) and \( p_i^2(x) \), respectively. Using \( n_1 \) data \( \{ p_i^1(x_s^1), s_1 = 1, \cdots, n_1 \} \) of variable \( p_i^1(x) \) and \( n_2 \) data \( \{ p_i^2(x_s^2), s_2 = 1, \cdots, n_2 \} \) of variable \( p_i^2(x) \), simple co-kriging can provide an estimation for \( p_k(x) \) at an unsampled location \( x \) as:

\[
p_k(x) = p_k^1 + \sum_{s_1=1}^{n_1} \lambda_{s_1}(p_i^1(x_s^1) - p_k^1) + \sum_{s_2=n_1+1}^{n_1+n_2} \lambda_{s_2}(p_i^2(x_s^2) - p_k^2) \tag{9}\]

where \( \lambda_{s_1} \) and \( \lambda_{s_2} \) are the weights assigned to \( s_1 \)-th datum \( p_i^1(x_s^1) \) and \( s_2 \)-th \( p \) datum \( p_i^2(x_s^2) \) for prediction of class probability at location \( x \), and \( p_k^1 \) and \( p_k^2 \), are the marginal class probabilities evaluated from primary and secondary sources. The kriging weights are obtained by solving the following formula system:

\[
\sum_{i=1}^{n_1} \lambda_{s_1} \text{cov}_{x_s^1}(x_s^1, x_s^1) + \sum_{i=n_1+1}^{n_1+n_2} \lambda_{s_2} \text{cov}_{x_s^2}(x_s^2, x_s^2) = \text{cov}_{x_s^1}(x, x_s^1), \text{for } s_1 = 1, \cdots, n_1; \tag{10}\]

\[
\sum_{i=1}^{n_1} \lambda_{s_1} \text{cov}_{x_s^1}(x_s^1, x_s^1) + \sum_{i=n_1+1}^{n_1+n_2} \lambda_{s_2} \text{cov}_{x_s^2}(x_s^2, x_s^2) = \text{cov}_{x_s^1}(x, x_s^2), \text{for } s_2 = n_1+1, \cdots, n_1+n_2 \]

where \( \text{cov}_{x_s^1} \) and \( \text{cov}_{x_s^2} \) are auto-covariance for primary and secondary variables, and \( \text{cov}_{x_s^1 x_s^2} \) being their across-covariance.
Analysis is simplified by retaining only the secondary data co-located with the individual grid nodes to be estimated, because this avoids the problem of modeling the spatial covariance of the secondary data. A further approximation is achieved by building on a Markov-like hypothesis whereby co-located data are used to screen data of the same type located further away, so that the cross-covariance between the indicator data and the probabilistic classification has a linear relationship with the covariance of the indicator data.

2 Tests

Aerial photographs in color at the 1 : 24 000 scale (Fig. 1) and Landsat TM image data (Fig. 2) were collected over an Edinburgh suburb near Blackford Hill, which were also used for previous research. The aerial photographic data were digitized on a desktop scanner and subsequently rectified by use of 1-order polynomials at a resolution of 5 m, a few ground control points and accurate positional data derived from large-scale plans. The resulting root mean squared error (RMSE) was about 3.2 m, sufficiently for its resolution.

Fig. 1 An aerial photographic subset showing an Edinburgh suburban test site

Fig. 2. An extract of Landsat TM image at an Edinburgh suburb (bands 3, 4 and 5 are shown in grey level but not to scale)

In order to facilitate computer-assisted processing of aerial photographic data, textual features at window sizes of 3 by 3, and 5 by 5 were calculated from the photographic intensities. The resultant data were then analyzed by use of the ERDAS software system to generate supervised classification. Such raw output was checked on screen to detect and correct any apparent errors and inconsistencies to permit reasonable accuracy in land cover mapping based on the digitized aerial photographic data. Upon accomplishment of aerial image classification, a 30 m resolution reference data set containing land cover proportions in each pixel was generated by aggregation over the 5 m resolution data, to allow for combination with the satellite data, as described below.

The Landsat TM sub-image was co-registered to the aerial image using 1-order polynomials, with a resultant RMSE of about 27 m or 0.9 pixel size. As bands 3, 4 and 5 contain much of the information for discriminant analysis of the land cover types, only these bands were retained in the satellite image subset and overlaid to the land cover proportion reference data. Fortran 77 subroutines were developed, which performed the generation of random sets of training and testing data, supervised classification with and without incorporation of contextual information contained in the training data set, and accuracy assessment for classification outputs.

2.1 Results with 5 classes

The study focused, firstly, on a five-class classification scheme, comprising grassland, built-up land, woodland, shrubland, and water bodies. A set of 231 pure samples was collected from the reference data. This data set was divided randomly into training and testing sets, comprising 114 and 117 pixels, respectively. Each of the three Landsat TM bands (bands 3, 4 and 5) was considered as a separate data source, from which radiance histograms could be constructed to derive probabilistic class membership values. One histogram was formed for each class in each Landsat TM band, resulting in a total of 15 histograms. A program was written for combined classification fed with band-specific probabilistic class membership values, resulting in combined probabilistic membership values. When desired, the probabilistic classification could be hardened with the application of a maximization operation.
to assign each pixel to the class with which had the highest membership value. The magnitudes of the highest class membership values could also be used to give an impression of the spatial distribution of the certainty of class allocations\[18.2s\].

Accuracy assessment is an important component of image classification and should be undertaken in a statistically sound manner\[a2'aa\]. Using the testing data, accuracy was evaluated in terms of the percent correctly classified (PCC) pixels and the Kappa coefficient of agreement for the hardened classification. The results (Table 1) revealed the classification based on spectral data alone was not quite accurate. To improve the accuracy of these evidential classifications, spatial context was incorporated into the analyses. This was achieved in two ways. First, a simple $3 \times 3$ majority filter was applied to the classification output\[34\]. This was found to markedly improve classification accuracy (Table 1). Second, probabilistic membership values of the training pixels, with their locational coordinates, were used to calculate experimental semivariograms, to which proper variogram models were fitted to enable derivation of spatial contextual information.

Table 1 Summary of classification accuracy assessment (5-class example)

| Classification          | Percentage of correctly classified pixels/% | Kappa coefficient of agreement |
|-------------------------|---------------------------------------------|--------------------------------|
| Spectral                | 30.8                                        | 0.165                          |
| Majority filtering      | 54.7                                        | 0.433                          |
| Spatial prior           | 60.7                                        | 0.456                          |

Kriging was undertaken with the cell size of the output grids equal to the 30 m spatial resolution of the Landsat TM sensor, resulting in spatial contexts in the form of probabilistic membership values at individual pixels for the test area. The classifications derived earlier were then combined with the data in spatial contexts to produce contextually enhanced classifications. The contextually enhanced classifications provided more spatially coherent representations than their non-contextual counterparts. It was also apparent that incorporating contextual information data into the classifications could result in a significant improvement in classification accuracy (Table 1). It was apparent that both methods for incorporating contextual information could be used to significantly improve the accuracy of spectral classifications.

The relatively low accuracy in classification with 5-classes seems to be due to the incompatibility between conceptual and data classes\[35,34\]. As commonly known, the classification scheme chosen for a project, representing information classes aimed for, specifies classes into which the images are to be classified. In other words, information classes reflect how reality is conceived in a given project and thus user’s requirements therein\[26\]. In a supervised mode, conceptual classes are linked with data classes by training, during which a set of ground data is used for class descriptions to drive classification\[37\]. By means of an error matrix based on cross-tabulating between classified data and training data, it was found that PCC for training data was only as good as about 62.3%, signaling severe incompatibility between information and data classes. To improve the classification significantly, a sensible way is to reduce incompatibility between data and information classes by coarsening the categorical semantic scale, i.e. decreasing the number of classes by proper merging, as shown in the next section.

2.2 Results with 3 classes

A 3-class scheme including mostly built/barren (including water bodies), predominantly vegetated, and mixed types of land cover was adopted due to their typicality of spectral signatures in both the aerial photographs and the Landsat TM image subset. Training pixels (195) and testing pixels (197) were located with the reference data described previously. Per-band histograms detailing the frequencies of reflectively digital numbers for each of the land cover types under consideration were derived from the training data. Which constituted the spectral signatures of the three classes. The satellite sub-image was classified, treating each band as separate source of
spectral measurement, using the product rule of classification. This gave rises to a percentage of 87.3\% correctly classified pixels, as calculated from cross-tabulation with the testing data. Usually, post-processing is used to improve initial spectral classification. In this test, a slight improvement in PCC (89.3\%) was reported in post-processing classification by use of majority filtering.

The Bayesian classification above is assumed to be of equal prior probabilities. It is possible to input priors as quantified by the proportions of training pixels belonging to individual classes to a Bayesian classifier. This resulted in a lower PCC of 85.8\%, prompting for possible improvements by incorporating spatial rather than non-spatial priors in Bayesian image classification. For this, spatial auto-and cross-covariance for land cover distributions were calculated from the training pixels and class probabilities derived from spectral classification, and subsequently fitted with Gaussian models of varying ranges and sills.

The derived variogram models were used in combination with the training data to drive indicator kriging, resulting in vectors of class-occurrence probabilities at individual grid nodes, i.e. the Landsat TM image pixels. The kriged map was evaluated, indicating a PCC of 89.5\%, greater than those obtained previously from classifications modified with non-spatial priors and post-processed with majority filtering.

To build further on spatial information contained in the training sample and spectral classification, co-kriging may be explored for combining indicator data consisting of geo-referenced training sample and probabilistic vectors evaluated at individual pixels. A PCC of 94.4\% was recorded for a full co-kriging approach to data combination, which is a remarkable improvement over spectral and spatial classification alone. Implementation of full co-kriging calls for co-regionalization modeling, which is often tedious and computationally costly. As a compromise, a simplified version of co-kriging using primary data, i.e. training data, and only part of the secondary data, i.e. the spectral classification, which collate with the grid nodes being estimated in the run of co-kriging. The resulting PCC registered 93.9\%, only slightly inferior to that obtained with full co-kriging but at substantial reduced computational costs.

The contextual information derived from indicator kriging based on training data is usually represented as regular arrays of prior probabilities of candidate categories existing over space. Such regularized information at grid nodes may be combined with spectral classifications, which, in this experiment, was already available on regular grids, indicating possibly stream-lined computation. The kriged contextual information in the form of prior probabilities was utilized by use of, firstly, the product rule to enhance spectral classification, resulting in PCC of 92.9\%, confirming an impressive improvement brought by combination of spatial and spectral information over either data set alone\cite{38}. Secondly, kriged spatial priors recorded at regular grid nodes can be combined with spectral classifications by use of collocated co-kriging. As the co-regionalization using only collocated primary and secondary data requires only variance and covariance quantities pooled from the primary and secondary data grids, it is possible to simplify calculation of weights and store them in computer for straightforward linear combination of probabilistic data from the primary and secondary sources. The weights are listed in Table 2, which results from simple co-kriging. Subsequently, a PCC of 94.9\% was reported, which is the best registered for combined classification, let alone classifications using either spectral or spatial information alone.

| Land cover types | Primary | Secondary |
|------------------|---------|-----------|
| Built/barren     | 0.668   | 0.562     |
| Mixed            | 0.678   | 0.602     |
| Vegetated        | 0.990   | 0.021     |

A summary of classification accuracy achieved by various classifiers is provided in Table 3,
where kappa coefficients of agreements are also computed. Post-processing of spectral classifications and co-kriging of training data and spectral classifications tend to smooth out the resultant categorical maps more than product rule-based classifications by used of gridded spatial priors and spectral classifications. In fact, the Bayesian product rule has a consensus mechanism not to coarsen data resolution by imposing a neighborhood for smoothing.

After comparative studies about different combined classifications have been completed, it is important to reflect on some deeprooted issues. The inaccuracy revealed by accuracy assessment does not necessarily represent a direct result of deficiency with the data or the classifiers. Rather, a substantial amount of inaccuracy originates from incompatibility between class definitions on different sources, such as photo-interpretation and digital spectral classification. This is perhaps one of the main reasons for less-than-expected performance of any digital image classifiers.

### Table 3 Accuracy assessment for spectral, spatial, and combined classification (3-class example)

| Classifiers | Percentage of correctly classified pixels/% | Kappa coefficient of agreement |
|-------------|------------------------------------------|------------------------------|
| Bayesian classification of spectral data | 87.3 | 0.740 |
| Equal priors | 85.8 | 0.701 |
| Non-spatial priors | 89.5 | 0.701 |
| Indicator kriging with training samples | 89.3 | 0.778 |
| Post-processing (majority filtering) of spectral classification | 93.8 | 0.878 |
| Combined classification | 94.9 | 0.901 |
| Training samples and spectral classification | 93.9 | 0.878 |
| Co-located co-kriging | 94.4 | 0.888 |
| Full co-kriging | 92.9 | 0.858 |
| Kriged prior and spectral classification | 94.9 | 0.901 |
| Bayesian product rule | 93.9 | 0.878 |
| Co-located co-kriging | 94.4 | 0.888 |

Further research may benefit from methods that reconcile the differences between categorical data of seemingly compatible definitions but disparate underlying semantics. Rough sets-based approximation is likely to shed light on this issue, so is the distills from ground truth data by commensuration between interpretation and measurement-based categorization.

### 3 Conclusions

This paper has endeavored to put Bayesian and geostatistical approaches for categorical data integration to comparative and comprehensive examination. It was found that the Bayesian product rule worked very well with kriged spatial priors and spectral classification, while co-kriging or its simplified version with collocated data configurations did slightly better though at extra expenses. Indeed, geostatistical strategies should be promoted for their strong theoretical basis and great potentials to applications.

Further research is needed on the topic of semantic homogenization across different categorical data sources, as semantic discrepancy tends to undermine performance of combined classification. Also necessary is investigation into the issues of scale and scaling, as different categorical maps are created with different resolution and thus scale, which, when combined, are bound to be affected by the incompatibilities of scaling spatially and semantically.

### REFERENCES

1. DeFries R, Hansen M, Townshend J, et al. (2000) A global 1 km data set of percent tree cover data derived from remote sensing. *Global Change Biology*, 6: 247-254
2. Stehman S V, Sohl T L, Loveland T R (2003) Statistical sampling to characterize recent united states landcover change. *Remote Sensing of Environment*, 86 (4): 517-529
3. Barr S, Barnsley M (2000) Reducing structural clutter in land cover classification of high spatial resolution remotely-sensed images for urban land use mapping. *Computers and Geosciences*, 26(4): 433-449
4. Hutchinson C F (1982) Techniques for combining
landsat and ancillary data for digital classification improvement. *Photogrammetric Engineering and Remote Sensing*, 48; 123-130

5 Skidmore A K (1989) An expert system classifies eucalypt forest types using thematic mapper data and a digital terrain model. *Photogrammetric Engineering and Remote Sensing*, 55; 1 449-1 464

6 Srinivasan A, Richards J A (1990) Knowledge-based techniques for multi-source classification. *International Journal of Remote Sensing*, 11(3); 505-525

7 Ehlers M (1991) Multi-sensor image fusion techniques in remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 46(1); 57-74

8 Peddle D, Franklin S E (1991) Image texture processing and data integration for surface pattern discrimination. *Photogrammetric Engineering and Remote Sensing*, 57(4); 413-420

9 Wang Y, Civeo D L (1994) Evidential reasoning-based classification of multi-source spatial data for improved land cover mapping. *Canadian Journal of Remote Sensing*, 20(4); 381-395

10 Vogelmann J E, Sohl T, Howard S M (1998) Regional characterization of land cover using multiple sources of data. *Photogrammetric Engineering & Remote Sensing*, 64(1); 45-57

11 Tso B, Mather P M (2001) Classification methods for remotely sensed data. London; Taylor & Francis.

12 LeBlanc M, Tiboharini R (1996). Combining estimates in regression and classification. *Journal of American Statistical Association*, 91; 1 611-1 650

13 Verstraete M M, Pinty B, Myneni R B (1995) Potential and limitations of information extraction on the terrestrial biosphere from satellite remote sensing. *Remote Sensing of Environment*, 58; 201-214

14 Zhang J, Goodchild M F (2002) Uncertainty in geographical information. London and New York; Taylor & Francis.

15 Kyriakidis P C, Liu X, Goodchild M F (2004) Geostatistical mapping of thematic classification uncertainty. In: Lunetta R S, Lyon J G (eds). Remote Sensing and GIS Accuracy Assessment. Boca Raton; CRC Press, 145-162

16 Switzer P, Kowalk W S, Lyon R J P (1982) A prior probability method for smoothing discriminant analysis classification maps. *Mathematical Geology*, 14(5); 433-444

17 Goovaerts P (2002) Geostatistical incorporation of spatial coordinates into supervised classification of hyperspectral data. *Journal of Geographical Systems*, 4; 99-111

18 Steele B M (2000) Combining multiple classifiers: an application using spatial and remotely sensed information for land cover type mapping. *Remote Sensing of Environment*, 74; 545-556

19 Kyriakidis P C (2001) Geostatistical models of uncertainty for spatial data. In: Hunsaker C T, Goodchild M F, Friedl M A, et al. (eds). Spatial Uncertainty in Ecology; Implications for Remote Sensing and GIS applications. New York; Springer-Verlag, 175-213

20 Aitchison J (1984) The statistical analysis of geochemical compositions. *Mathematical Geology*, 16; 531-564

21 Pawlowsky G, Olea R A, Davis J C (1995) Estimation of regionalized compositions; a comparison of three methods. *Mathematical Geology*, 27(1); 105-127

22 Strahler A H (1980) The use of prior probabilities in maximum likelihood classification of remotely sensed data. *Remote Sensing of Environment*, 10; 135-163

23 Card D H (1982) Using known map category marginal frequencies to improve estimates of thematic map accuracy. *Photogrammetric Engineering and Remote Sensing*, 48; 431-439

24 Lugosi G, Nobel A B (1996) Consistency of data-driven histogram methods for density estimation and classification. *Annals of Statistics*, 24; 687-706

25 Masselli F, Conese C, De Filippis T, et al. (1995) Integration of ancillary data into a maximum-likelihood classifier with nonparametric priors. *ISPRS Journal of Photogrammetry and Remote Sensing*, 50(2); 2-11

26 Lee T, Richards J A, Swain P H (1987) Probabilistic and evidential approaches for multisource data analysis. *IEEE Transactions on Geoscience and Remote Sensing*, GE-25(3); 283-293

27 Benediktsson J A, Sveinsson J R, Swain P H (1997) Hybrid consensus theoretic classification. *IEEE Transactions on Geoscience and Remote Sensing*, 35(4); 833-843

28 Bregt A K, Bouma J, Jellinek M (1987) Comparison of thematic maps derived from a soil map and from kriging of point data. *Geoderma*, 39; 281-291

29 De Bruin S (2000) Predicting the areal extent of landcover types using classified imagery and geostatistics. *Remote Sensing of Environment*, 74; 387-396

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