J Geograph Syst (2002) 4:99–111

Journal of Geographical Systems

Geostatistical incorporation of spatial coordinates into supervised classification of hyperspectral data

P. Goovaerts

Department of Civil and Environmental Engineering, The University of Michigan, Ann Arbor, MI 48109-2125, USA (e-mail: goovaert@engin.umich.edu)

Received: 12 April 2001 / Accepted: 7 September 2001

Abstract. This paper presents a methodology to incorporate both hyperspectral properties and spatial coordinates of pixels in maximum likelihood classification. Indicator kriging of ground data is used to estimate, for each pixel, the prior probabilities of occurrence of classes which are then combined with spectral-based probabilities within a Bayesian framework. In the case study (mapping of in-stream habitats), accounting for spatial coordinates increases the overall producer's accuracy from 85.8% to 93.8%, while the Kappa statistic rises from 0.74 to 0.88. Best results are obtained using only indicator kriging-based probabilities, with a stunning overall accuracy of 97.2%. Significant improvements are observed for environmentally important units, such as pools (Kappa: 0.17 to 0.74) and eddy drop zones (Kappa: 0.65 to 0.87). The lack of benefit of using hyperspectral information in the present study can be explained by the dense network of ground observations and the high spatial continuity of field classification which might be spurious.

Key words: Hyperspectral, streams, supervised classification, kriging

1 Introduction

During the last decade the development of computational resources has fostered the use of numerical methods to process the large bodies of data that are collected in the geosciences. A key feature of geoscience information is that each observation relates to a particular location in space. Knowledge of an attribute value, say a heavy metal concentration or reflectance value, is of little interest unless the location of the measurement is known and accounted for in the analysis.

The author is grateful to Andrew Marcus for the Lamar River data set as well as for his constructive comments on the manuscript. Comments by two anonymous reviewers also improved the presentation of this paper.

An application where data locations are likely worth incorporating is the stratification or classification of spectral imagery, such as the mapping of land cover classes or urban areas. A supervised classification typically proceeds in three steps:

1) the spectral signature of each class is first derived from a set of training data where both ground truth and spectral information are available,
2) a classification criterion (e.g. discriminant function, k-nearest-neighbor method) is developed and used to estimate the probability of occurrence of each class for any pixel to be classified,
3) the pixel is then assigned to the class with the maximum probability of occurrence (maximum likelihood criterion).

The main limitation of such spectral classifiers is that the proximity of a pixel to a class is computed only in the spectral space, without any consideration for the spatial coordinates of that pixel. As a result, classifications often display noisy or unrealistic features, such as isolated pixels assigned to a particular class. Although such features might be corrected *"a posteriori"* using post-classification smoothing algorithms (e.g., Deutsch 1998), it would be more rigorous to use the spatial information during the classification.

Geostatistics (Journel and Huijbregts 1978; Goovaerts 1997) provides a set of statistical tools for incorporating the spatial and temporal coordinates of observations in data processing. Of particular interest is indicator kriging (Goovaerts 1994), an interpolation algorithm that allows one to estimate the probability of occurrence of classes on the basis of surrounding observations. Remotely sensed secondary information can be incorporated through a calibration procedure (see Goovaerts 1997, p.328). This procedure has been applied to different fields such as mapping of rock types using seismic imagery as secondary data or mapping of soil types (Oberthür et al. 1999), but despite its great potential, has seldom been used in remote sensing.

This paper aims to present a methodology to combine probabilities of occurrence provided by spectral classification, which ignores data locations, and the ones estimated by indicator kriging, which ignores spectral information. The technique is applied to the stratification of stream environments in the Lamar River, Yellowstone National Park, using hyperspectral data analyzed in other papers in this special issue (Jacquez and Maruca; Griffith; Lagona; Marcus; all this issue). The proposed classification is compared to results obtained using only spectral information (spectral classifier) or ground observations (indicator kriging).

2 Theory

Let $\{s_l, l = 1, \ldots, L\}$ be a set of L mutually exclusive classes (e.g. land cover classes or habitats) observed at n locations $\mathbf{u}_\alpha$ over a field $\mathscr{D}$. Spectral information is available at these same locations and takes the form of a vector of K variables Z_k (e.g. spectral bands or principal components computed from spectral data), $\{z_k(\mathbf{u}_\alpha), \alpha = 1, \ldots, n, k = 1, \ldots, K\}$. The objective is to assign to a single class s_l any location $\mathbf{u}$ within $\mathscr{D}$ where a vector of K spectral data is available. Typically these locations are on a grid and correspond to the pixels of the image.

A pixel $\mathbf{u}$ is classified into a given class $s_{l'}$ according to maximum likelihood (ML) criterion, which means that $s_{l'}$ is the class with the largest probability of occurrence at $\mathbf{u}$:

$$s(\mathbf{u}) = s_{l'} \quad \text{if} \quad p(\mathbf{u}; s_{l'}|B) > p(\mathbf{u}; s_l|B) \quad \forall\, l \neq l' \tag{1}$$

where $p(\mathbf{u}; s_{l'}|B)$ is the probability of occurrence of $s_{l'}$ at $\mathbf{u}$ given the information B available, and it is often referred to as "posterior" probability. Depending on the type of information B, there are different ways to estimate $p(\mathbf{u}; s_l|B)$.

2.1 Spectral classification

Consider the common situation where only the vector of spectral data at $\mathbf{u}$ is used, $B_z = \{z_k(\mathbf{u}), k = 1, \ldots, K\}$. The posterior probability of occurrence of any class s_l is estimated by applying Bayes' theorem:

$$p(\mathbf{u}; s_l|B_z) = \text{Prob}\{S(\mathbf{u}) = s_l|B_z\} = \frac{p_l \cdot f(B_z|s_l)}{\sum_{l=1}^{L} p_l \cdot f(B_z|s_l)} \quad l = 1, \ldots, L \tag{2}$$

where p_l is the "prior" probability of occurrence of class s_l at $\mathbf{u}$, and $f(B_z|s_l)$ is the conditional density of spectral data Z_k given the class s_l. Several choices exist for the prior probabilities p_l. For example, they can all be considered equal, $p_l = 1/L$ for all l, which means all L classes have *a priori* the same probability to prevail at $\mathbf{u}$. If the training data set is assumed to be representative of the proportion of different classes within the study area, the priors can be set proportional to the sample size, that is p_l is the proportion of n training data belonging to s_l.

There is also a variety of techniques available for estimating the densities $f(B_z|s_l)$. When the distribution of spectral data within each class is assumed to be multivariate normal, a parametric method based on multivariate normal distribution theory can be used to derive a linear or quadratic discriminant function. An alternative is to adopt a non-parametric method (e.g. kernel method or k-nearest-neighbor method) that does not make any assumption about the distributions (Everitt and Dunn 2001).

In this paper, the spectral classification has been performed using a parametric approach that is implemented in the SAS software (SAS Institute Inc. 1989) as follows:

1. Computation for each class s_l of the variance–covariance matrix between the K spectral variables, $\mathbf{C}_l$.
2. Derivation of the inverse of each matrix $\mathbf{C}_l$, $\mathbf{C}_l^{-1}$, and computation of the natural log of the determinant, $\ln(|\mathbf{C}_l|)$.
3. Computation of the generalized squared distance between the "unclassified" location $\mathbf{u}$ and each class s_l as:

$$d^2(\mathbf{u}; s_l) = [\mathbf{Z}(\mathbf{u}) - \mathbf{m}_l]^T \cdot \mathbf{C}_l^{-1} \cdot [\mathbf{Z}(\mathbf{u}) - \mathbf{m}_l] + \ln(|\mathbf{C}_l|) - 2\ln[p_l] \tag{3}$$

where $\mathbf{Z}(\mathbf{u})$ is the vector of spectral data at $\mathbf{u}$: $\mathbf{Z}(\mathbf{u}) = [z_1(\mathbf{u}), z_2(\mathbf{u}), \ldots, z_K(\mathbf{u})]$, $\mathbf{m}_l$ is the vector of means of the K variables for the class s_l: $\mathbf{m}_l = [m_{1|l}, m_{2|l}, \ldots, m_{K|l}]$.

4. Computation of the posterior probability for the location $\mathbf{u}$ to belong to any given class s_l as:
$$p(\mathbf{u}; s_l|B_z) = \frac{\exp\{-0.5 \cdot d^2(\mathbf{u}; s_l)\}}{\sum_{l=1}^{L} \exp\{-0.5 \cdot d^2(\mathbf{u}; s_l)\}} \quad (4)$$
5. Classification of location $\mathbf{u}$ into class s_l with the largest posterior probability of occurrence.

2.2 Indicator kriging (IK)

Assume there is no spectral data and the only information available for classifying $\mathbf{u}$ are n_c surrounding ground data, $\{s(\mathbf{u}_\alpha), \alpha = 1, \ldots, n_c\}$, with $n_c \leq n$. This information is hereafter denoted B_{geo}. The indicator approach for modeling probabilities $p(\mathbf{u}; s_l|B_{geo})$ is based on a prior coding of ground data into local prior probability values:

$$\text{Prob}\{s(\mathbf{u}_\alpha) = s_l| \text{ local information at } \mathbf{u}_\alpha\} \quad (5)$$

If there is no uncertainty about the class prevailing at sampled locations $\mathbf{u}_\alpha$ (no field classification error), prior probabilities are one for the class observed and zero for $(L-1)$ other classes. The information at $\mathbf{u}_\alpha$ consists then of a vector of L indicator values $i(\mathbf{u}_\alpha; s_l)$ where:

$$i(\mathbf{u}_\alpha; s_l) = \begin{cases} 1 & \text{if } s(\mathbf{u}_\alpha) = s_l \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

The global proportion of class s_l, denoted p_l, is simply the arithmetical average of the corresponding indicator data:

$$p_l = \frac{1}{n} \sum_{\alpha=1}^{n} i(\mathbf{u}_\alpha; s_l)$$

A measure of transition frequency between classes s_l and $s_{l'}$ as a function of the separation vector $\mathbf{h}$ is the indicator cross semivariogram that is computed from indicator data as:

$$\gamma_I(\mathbf{h}; s_l, s_{l'}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [i(\mathbf{u}_\alpha; s_l) - i(\mathbf{u}_\alpha + \mathbf{h}; s_l)][i(\mathbf{u}_\alpha; s_{l'}) - i(\mathbf{u}_\alpha + \mathbf{h}; s_{l'})] \quad (7)$$

where $N(\mathbf{h})$ is the number of data pairs separated by a vector $\mathbf{h}$. Substituting s_l for $s_{l'}$ in expression (7) yields the indicator direct semivariogram $\gamma_I(\mathbf{h}; s_l, s_l)$ which characterizes the spatial continuity of the class s_l.

The probability $p(\mathbf{u}; s_l|B_{geo})$ is the conditional expectation of the indicator random function $I(\mathbf{u}; s_l)$, where $p(\mathbf{u}; s_l|B_{geo}) = E[I(\mathbf{u}; s_l)|B_{geo}]$. Thus, it can be estimated by (co)kriging using neighboring indicator data $i(\mathbf{u}_\alpha; s_l)$. The indicator cokriging (coIK) algorithm which utilizes indicator data related to any class is theoretically the best because it accounts for transition probabilities between classes. Its implementation, however, requires joint modeling of $L(L+1)/2$ indicator (cross) semivariograms and solving large and possibly unstable cokriging systems. Moreover, Goovaerts (1994)

showed that coIK may not provide more accurate results than the straightforward indicator kriging (IK) which involves modeling only L direct indicator semi variograms $\gamma_I(\mathbf{h}; s_l, s_l)$. In this paper, conditional probabilities were estimated using the IK algorithm which utilizes only indicator data on the class s_l being considered:

$$p^*(\mathbf{u}; s_l|B_{geo}) = \sum_{\alpha=1}^{n_c} \lambda(\mathbf{u}_\alpha; s_l).i(\mathbf{u}_\alpha; s_l) \tag{8}$$

The weights $\lambda(\mathbf{u}_\alpha; s_l)$ are obtained by solving the following ordinary kriging system of $(n_c + 1)$ equations (Goovaerts 1997):

$$\begin{aligned} &\sum_{\beta=1}^{n_c} \lambda(\mathbf{u}_\beta; s_l)\ \gamma_I(\mathbf{u}_\alpha - \mathbf{u}_\beta; s_l, s_l) - \mu_l = \gamma_I(\mathbf{u}_\alpha - \mathbf{u}; s_l, s_l) \\ &\quad \text{for all } \alpha = 1 \text{ to } n_c \\ &\sum_{\beta=1}^{n_c} \lambda(\mathbf{u}_\beta; s_l) = 1 \end{aligned} \tag{9}$$

At each location $\mathbf{u}$, the set of L probabilities $\{p^*(\mathbf{u}; s_l|B_{geo}), l = 1, \ldots, L\}$ must satisfy the two following order relations:

$$0 \leq p^*(\mathbf{u}; s_l|B_{geo}) \leq 1 \tag{10}$$

$$\sum_{l=1}^{L} p^*(\mathbf{u}; s_l|B_{geo}) = 1 \tag{11}$$

The first condition may not be satisfied because kriging weights can be negative, therefore the kriging estimate is a non-convex linear combination of the conditioning data. Following Deutsch and Journel (1998), the first constraint (10) is met by resetting faulty probabilities to the nearest bound, 0 or 1. Once this correction has been performed, each value $p^*(\mathbf{u}; s_l|B_{geo})$ is standardized by the sum $\sum_{l=1}^{L} p^*(\mathbf{u}; s_l|B_{geo})$ to meet the second condition (11). An alternative to this *ad hoc* correction of estimated probabilities is to incorporate these conditions directly into the kriging system, a technique known as compositional kriging (De Gruijter et al. 1997).

2.3 A mixed approach

The key idea is to find a way to account for both spectral information and surrounding ground data into the classification procedure. In other words, how could we derive probabilities of occurrence $p(\mathbf{u}; s_l|B)$ where $B = \{z_k(\mathbf{u}), k = 1, \ldots, K\} \cup \{s(\mathbf{u}_\alpha), \alpha = 1, \ldots, n_c\}$?

Multivariate geostatistical algorithms, such as soft cokriging or simple kriging with soft prior means (Goovaerts 1997, p.328; Oberthür et al. 1999), could be used to incorporate both spectral and spatial information into the estimation of $p(\mathbf{u}; s_l|B)$. These techniques are, however, demanding in terms of semivariogram inference and modeling, and are not implemented in most GIS packages. A simpler approach is here adopted whereby the prior probabilities p_l in Bayes' expression (2) are replaced by posterior probabilities estimated using indicator kriging:

$$p(\mathbf{u}; s_l|B) = \frac{p^*(\mathbf{u}; s_l|B_{geo}) \cdot f(B_z|s_l)}{\sum_{l=1}^{L} p^*(\mathbf{u}; s_l|B_{geo}) \cdot f(B_z|s_l)} \quad (12)$$

This approach amounts to assuming that the prior probability of occurrence of a class s_l is not the same everywhere but depends on the pixel location $\mathbf{u}$. For example, that probability will be large if data belonging to class s_l are close geographically. Once probabilities (12) have been derived, the maximum likelihood criterion is applied.

Although most classification algorithms don't provide densities $f(B_z|s_l)$, the approach can be easily implemented using a two-step procedure: 1) the posterior probabilities (2) are estimated using Bayes' theorem with $p_l = 1/L$, 2) these probabilities are then used instead of densities $f(B_z|s_l)$ in equation (12). Because densities and posterior probabilities are proportional if $p_l = 1/L$, this shortcut yields the same results as the theoretical expression (12). Another option is to implement the procedure described in Section 2.1, using the following generalized squared distances for the computation of posterior probabilities (4):

$$d^2(\mathbf{u}; s_l) = [\mathbf{Z}(\mathbf{u}) - \mathbf{m}_l]^T \cdot \mathbf{C}_l^{-1} \cdot [\mathbf{Z}(\mathbf{u}) - \mathbf{m}_l] + ln(|\mathbf{C}_l|) - 2\ln[p^*(\mathbf{u}; s_l|B_{geo})]$$

3 Case study

The three types of classification are applied to the mapping of morphologic units in the Lamar River in Northeast Yellowstone (see Fig. 1 and detailed description of data in Jacquez et al. this issue). Remote sensing information consists of a hyperspectral (128 bands) and high spatial resolution (pixel size = 1 m^2) imagery. Hyperspectral data have been processed using principal component analysis. The analysis is conducted on the first 25 components which were found to provide the best maximum likelihood classification in a prior study (Marcus, this issue).

Four morphologic units, corresponding to different in-stream habitats, are distinguished: eddy drop zones, riffles, pools and glides. Field mapping provides what is considered as "ground truth data" for 11,552 pixels of size 1 m^2. A random subset of 519 pixels, including 129 to 131 data per unit, forms the training data set. The remaining 11,033 pixels are used for validation.

3.1 Indicator kriging

Each of the 519 training sites was coded into a vector of four indicators $[i(\mathbf{u}_\alpha; s_l), l = 1, 2, 3, 4]$, and four experimental indicator semivariograms were computed as:

$$\hat{\gamma}_I(\mathbf{h}; s_l) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [i(\mathbf{u}_\alpha; s_l) - i(\mathbf{u}_\alpha + \mathbf{h}; s_l)]^2 \quad (13)$$

where $N(\mathbf{h})$ is the number of data pairs within a given class of distance. Typically the Euclidian distance between observations is computed, which becomes meaningless in a meandering river as these distances could relate to

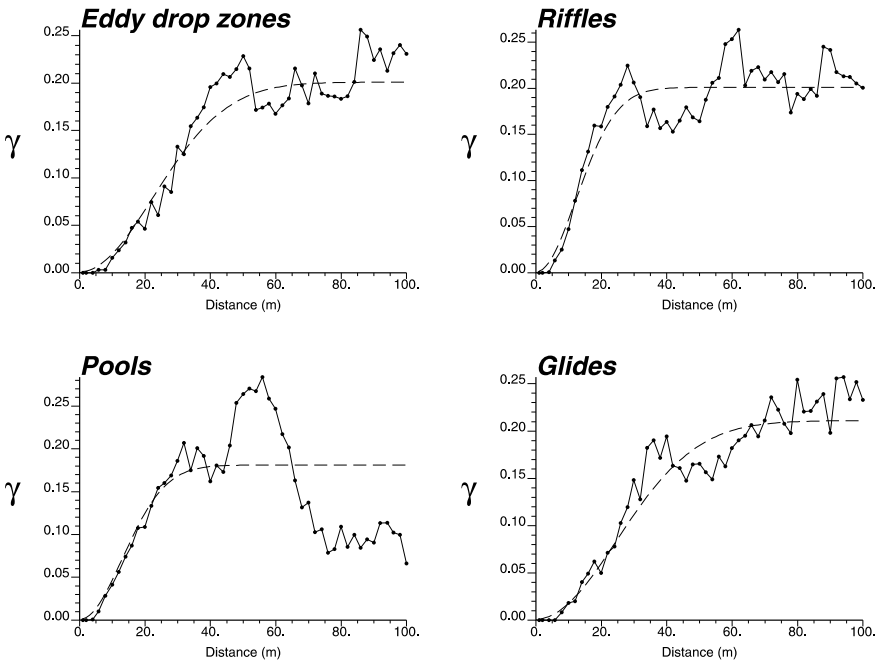

Fig. 1. Experimental indicator semivariograms of four morphologic units, with the model fitted

points over intervening land. Euclidian distances were used in this study and the results were sufficiently good to suggest that more complex distance measures were not necessary. An alternative to the use of Cartesian coordinates is to measure "x-axis" stream distances along the center line or thalweg (deepest channel) of the stream, and "y-axis" coordinates from either the center line or one of the banks (see for example Ladd et al. 1998). Another option is to transform coordinates through the generation of a grid within the river that "straightens out" the domain of analysis ensuring that distances are measured within the river (Barabás et al. 2001).

Figure 1 shows the experimental indicator semivariograms with the model fitted. Zero values for first lags indicate that clustered training sites belong to the same units: there is no transition between units up to a separation distance of 3–4 m. This high spatial continuity of morphologic units stresses the potential benefit of spatial information for classification. Parabolic behavior near the origin of semivariograms has been captured using a Gaussian model, and a small nugget effect (discontinuity at the origin) was added to avoid numerical instabilities in spatial interpolation (Goovaerts 1997). For each validation site the probability of occurrence of the four units was computed using indicator kriging and the sixteen closest observations (i.e. $n_c = 16$).

3.2 Classification procedures

The spectral classification was conducted using the DISCRIM procedure of the SAS software (SAS Institute Inc. 1989). Best results were obtained for a

parametric approach that assumes hyperspectral data to follow a multivariate normal distribution within each unit. Within-unit covariance matrices were used instead of a pooled covariance matrix. A discriminant function was developed from the first 25 principal components at 519 training sites, and used to classify the 11,033 validation sites.

The second type of classification was performed by applying the ML criterion to probabilities estimated by indicator kriging.

Probabilities estimated by indicator kriging and provided by the discriminant function were combined according to equation (12), and a ML criterion was then applied.

Figure 2 shows the training sites, the validation data, and the three classifications for a section of the River where all units are present. The validation data illustrates the high spatial continuity of units that was already detected on the indicator semivariograms of Fig. 1. Because it ignores pixel locations, the spectral classifier creates a stratification that is very noisy, in particular in the bottom left cluster of riffle pixels (light gray dots). The indicator kriging based approach yields a much smoother classification that displays striking similarities with the reference map. A few misclassified pixels can be found at the bottom of the right-hand side cluster of glides pixels. These pixels are wrongly allocated to eddy drop zone (edz) because of the proximity of many edz data (black pixels) that overwhelm the single glide field observation (open circle) found in this part of the section. This "local" undersampling of glide pixels can be compensated by incorporating spectral data through a mixed classification approach: all pixels of this specific unit are now correctly classified. Although the mixed classification is less noisy than the spectral one, it still displays several isolated pixels in the bottom left cluster of riffle pixels.

This example nicely illustrates the features of the three classification procedures. Visually, the IK-based approach leads to the most accurate classification, which can be explained by the high spatial continuity of morphologic units coupled with a dense network of ground data.

3.3 Performance comparison

Tables 1–3 show, for each classification procedure, the error matrix computed from 11,033 validation sites. For all four units, the two methods that account for pixel coordinates yield a larger number of correctly classified pixels (diagonal elements of the matrix) than the spectral classifier of Table 1. Best results are systematically produced by the IK approach that ignores spectral information.

To summarize and quantify the information contained in error matrices, several statistics were computed (see Table 4). Here, e_{ij} is the element of the ith row and jth column in the error matrix, which represents the number of validation pixels from the jth unit that are classified into the ith unit. The marginal row totals, e_{+j}, represent the total numbers of ground truth pixels within each unit, while the marginal column totals, e_{i+}, represent the total number of pixels classified within each unit. The following statistics are derived for each ith unit:

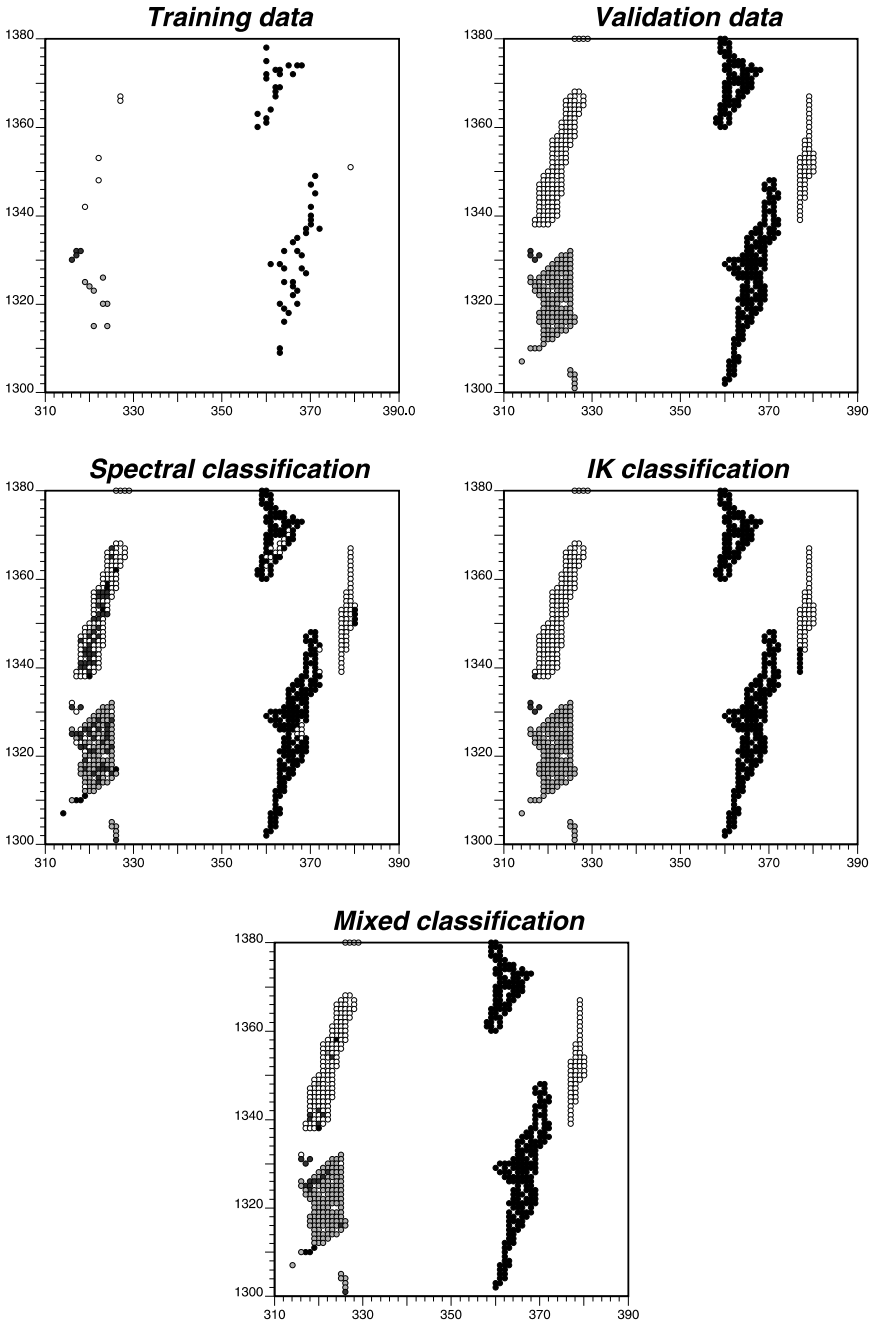

Fig. 2. Results of the three classification procedures for a section of the stream. Top graphs shows the training data available in that section and the field map (reference stratification). The color code is: eddy drop zones (*black*), riffles (*light gray*), glides (*white*), and pools (*drak gray*)

Table 1. Error matrix for maximum likelihood supervised classification using 1 meter imagery, principal components 1 through 25

Classified pixels	**Ground truth pixels**				
	eddy drop zones	riffles	glides	pools	Total
eddy drop zones	**563**	135	140	2	840
riffles	23	**2663**	160	4	2850
glides	78	392	**6111**	12	6593
pools	1	307	308	**134**	750
Total	665	3497	6719	152	11033

Table 2. Error matrix for indicator kriging based classification using 519 validation data

Classified pixels	**Ground truth pixels**				
	eddy drop zones	riffles	glides	pools	Total
eddy drop zones	**665**	31	63	0	759
riffles	0	**3329**	44	0	3373
glides	0	117	**6579**	0	6696
pools	0	20	33	**152**	205
Total	665	3497	6719	152	11033

Table 3. Error matrix for mixed (spectral + spatial) classification using 1 m imagery, principal components 1 through 25

Classified pixels	**Ground truth pixels**				
	eddy drop zones	riffles	glides	pools	Total
eddy drop zones	**636**	130	107	5	878
riffles	11	**3100**	57	3	3171
glides	18	171	**6469**	1	6659
pools	0	96	86	**143**	325
Total	665	3497	6719	152	11033

1. producer's accuracy $= e_{ii}/e_{+j}$
2. users' accuracy $= e_{ii}/e_{i+}$
3. Kappa statistic $= [N \cdot e_{ii} - e_{+j} \cdot e_{i+}]/[N \cdot e_{i+} - e_{+j} \cdot e_{i+}]$

The overall producer's accuracy and Kappa statistic are also computed.

All statistics confirm the superiority of the IK-based classification over the purely spectral one, with the mixed approach being a close second. Using only pixel coordinates increases the overall producer's accuracy from 85.8% to 97.2%, while the Kappa statistic rises from 0.74 to 0.95. Improvement is particularly substantial for the pools category. It's noteworthy that the producer's accuracy is 100% for the two units (pools and eddy drop zones) that are the most important from an environmental monitoring perspective (Marcus, this issue).

The impact of ground data density on classification results was investigated by repeating the analysis using, respectively, only 64 and 32 training

sites per unit, which represents a 50% and 75% reduction in sample size. Table 5 shows that using 32 training pixels per unit greatly affects the quality of hyperspectral data calibration, with a total accuracy down to 60.7% for the mixed approach and 56.9% for the spectral classifier. IK-based classification is much more robust with respect to data sparsity, producing a reasonable total accuracy of 86.8% when only 128 training data are available.

The lack of benefit of using hyperspectral information here can be explained by the high spatial continuity of classes and the dense network of ground observations. The high continuity of the field classification might

Table 4. Summary statistics for the three types of classification

Statistics	eddy drop zones	riffles	glides	pools	Overall
Spectral classification					
Producer's accuracy	84.7%	76.2%	91.0%	88.2%	85.8%
User's accuracy	67.0%	93.4%	92.7%	17.9%	
Kappa statistic	0.65	0.90	0.81	0.17	0.74
IK-based classification					
Producer's accuracy	100.0%	95.2%	97.9%	100.0%	97.2%
User's accuracy	87.6%	98.7%	98.3%	74.1%	
Kappa statistic	0.87	0.98	0.96	0.74	0.95
Mixed classification					
Producer's accuracy	95.6%	88.6%	96.3%	94.1%	93.8%
User's accuracy	72.4%	97.8%	97.1%	44.0%	
Kappa statistic	0.71	0.97	0.93	0.43	0.88

Table 5. Impact of the number n_t of training data on the accuracy of 3 types of classification

Statistics	n_t	eddy drop zones	riffles	glides	pools	Overall
Spectral classification						
Producer's accuracy	256	88.0%	77.2%	87.8%	72.4%	84.2%
	128	92.5%	56.6%	53.7%	46.1%	56.9%
User's accuracy	256	54.3%	89.3%	93.9%	16.9%	
	128	18.1%	69.2%	88.4%	10.2%	
Kappa statistic	256	0.51	0.84	0.84	0.16	0.72
	128	0.13	0.55	0.70	0.09	0.36
IK-based classification						
Producer's accuracy	256	100.0%	90.6%	92.8%	100.0%	92.6%
	128	96.1%	83.4%	87.3%	100.0%	86.8%
User's accuracy	256	80.7%	92.7%	97.4%	38.6%	
	128	68.1%	86.7%	94.3%	29.9%	
Kappa statistic	256	0.79	0.89	0.93	0.38	0.86
	128	0.66	0.81	0.85	0.29	0.76
Mixed classification						
Producer's accuracy	256	96.4%	85.2%	91.8%	92.1%	90.0%
	128	95.2%	60.0%	57.7%	62.5%	60.8%
User's accuracy	256	58.6%	95.2%	97.3%	30.2%	
	128	17.5%	76.8%	93.2%	17.9%	
Kappa statistic	256	0.56	0.93	0.93	0.29	0.82
	128	0.12	0.66	0.83	0.17	0.42

however be spurious and reflect the tendency of field teams to lump many small-scale areas into a single uniform polygon to save time and map more areas. In practice, a given unit (e.g., a riffle) would have small areas that are more like a glide, others that are more like an eddy drop zone. The "speckled" patterns displayed by the spectral classification on Fig. 2, where individual pixels appear in the middle of units, might then be real but these speckles are categorized as misclassifications because of the mapping of units as homogeneous classes. Also, indicator kriging has likely been favored by the fact that transition zones between units, which exhibit short-scale variability and would benefit the most from spectral information, have been removed from the validation data set (Marcus, this issue). Another limitation of the validation set is that it concerns a short reach that has been very densely sampled. In practice, these data would be used to calibrate a classification of the remaining watershed where the lack of ground data would handicap the application of indicator kriging, since it relies on nearby field observations to estimate the probability of occurrence of various classes.

4 Conclusions

Indicator geostatistics provides tools for modeling the spatial distribution of categorical variables and estimating probabilities of occurrence of classes based on surrounding observations. These probabilities can be fed directly into a maximum likelihood classification, or be first updated using spatially exhaustive hyperspectral data. This paper presents a straightforward procedure to combine kriging-based probabilities with probabilities derived from calibration of hyperspectral information.

Accounting for pixel locations greatly improves the accuracy of stratification of the stream into morphologic units. Surprisingly, best results are obtained when only ground truth data are accounted for; that is, when hyperspectral information is ignored. Outstanding performances of indicator kriging for this data set can, to some extent, be explained by: (i) the unusually dense network of ground observations, (ii) the removal of transition zones that typically display short-scale variability and would benefit the most from spectral information, and (iii) the high spatial continuity of the field classification which might be spurious. This should be further investigated by applying the different procedures to other data sets that display various patterns of spatial continuity and other sampling schemes. The potential benefit of more sophisticated geostatistical algorithms, such as soft cokriging, for merging both spatial and spectral-based probabilities should also be examined.

References

Barabás N, Goovaerts P, Adriaens P (2001) Geostatistical assessment and validation of uncertainty for three-dimensional dioxin data from sediments in an estuarine river. *Environmental Science & Technology* 35:3294–3301

De Gruijter JJ, Walvoort DJJ, van Gaans PFM (1997) Continuous soil maps – a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models. *Geoderma* 77:169–195

Deutsch CV (1998) Cleaning categorical variable (lithofacies) realizations with maximum a-posteriori selection. *Computers & Geosciences* 24:551–562

Deutsch CV, Journel AG (1998) *GSLIB: Geostatistical Software Library and User's Guide*. Oxford University Press, New York, 369 p

Everitt BS, Dunn G (2001) *Applied Multivariate Data Analysis*, Oxford University Press, New York, 352 p

Goovaerts P (1994) Comparison of CoIK, IK and mIK performances for modeling conditional probabilities of categorical variables. In: Dimitrakopoulos R (ed) *Geostatistics for the Next Century*. Kluwer, Dordrecht, pp 18–29

Goovaerts P (1997) *Geostatistics for Natural Resources Evaluation*. Oxford Univ. Press, New-York, 512 pp

Jacquez GM, Marcus WA, Aspinall RJ, Greiling DA (2002) *Exposure assessment using high spatial resolution hyperspectral (HSRH) imagery*. (In this issue) *Journal of Geographical Systems* 4:1–14

Journel AG, Huijbregts CJ (1978) *Mining Geostatistics*. Academic Press, New York, 600 p

Ladd S, Marcus WA, Cherry S (1998) Trace metal segregation within morphologic units. *Environmental Geology and Water Sciences* 36:195–206

Marcus WA (2002) *Mapping of stream microhabitats with hight spatial resolution hyperspectral imagery*. (In this issue.) *Journal of Geographical Systems* 4:113–126

Oberthür T, Goovaerts P, Dobermann A (1999) Mapping soil texture classes using field texturing, particle size distribution and local knowledge by both conventional and geostatistical methods. *European Journal of Soil Science* 50:457–479

SAS Institute Inc. (1989) *SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 1*. SAS Institute Inc., Cary, NC, 943 p