AN ALGORITHM FOR SPATIAL DATA CLASSIFICATION AND AUTOMATIC MAPPING BASED ON "SPIN" CORRELATIONS

Milan Žukovič
Department of Mineral Resources Engineering
Technical University of Crete
Chania 73100, Greece
Email: mzukovic@mred.tuc.gr

Dionissios T. Hristopulos
Department of Mineral Resources Engineering
Technical University of Crete
Chania 73100, Greece
Email: dionisi@mred.tuc.gr

KEYWORDS

spatial statistics, non-Gaussian distribution, Ising model, optimization, simulations

ABSTRACT

We employ the Ising model from statistical physics in the problem of spatial data classification. We use a multiple-class discretization of the sample values. The proposed algorithm predicts the class identity at unmeasured points based on Monte Carlo simulations that are conditional on the observed data (sample). The algorithm aims to minimize the deviation between the normalized correlation energies of the sample and the entire domain. A hierarchical scheme is used, in which points predicted to belong in lower-level classes retain their identity in the inference of the higher-level classes. The method is non-parametric and thus suitable for application to non-Gaussian data. The method is investigated using real data of surface elevation over a large part of the territory of North America. The effects of the ratio of training to prediction points, the number of classes, and the initial conditions are investigated. Potential extensions of the model are also discussed.

INTRODUCTION

The nowadays availability of large quantity of remotely-sensed geo-referenced data, such as land cover, terrain elevation, population, meteorological variables, atmospheric pollution, etc., creates increasing demands for efficient processing and analysis. The information contained in the data can provide reliable policy and management tools for application in diverse areas, such as ecosystem management, environmental policy, the design of real-time hazard warning systems, and various other decision making tasks. However, the data may be derived from different sources, measured by various systems using different methodologies and space-time resolutions. Furthermore, the data coverage is often incomplete due to different reasons, such as limited resources (material, human, technical), equipment limitations (detection level, resolution), meteorologi-

cal conditions (observations hindered by clouds). This situation leads to the missing data problem. Therefore, in order to use standard tools for the analysis of space-time data, there is a need to fill the data gaps and unify the resolution. These tasks involve down-scaling (refining) of the data with sparse resolution. They can be performed by means of well established interpolation and classification techniques (Atkinson et al., 1999). However, considering the ever-increasing size of spatial data, some classical methods used in geostatistics (e.g., kriging), e.g. (Wackernagel, 2003), can be impractical due to the high computational complexity involved. Furthermore, methods like kriging usually assume a jointly Gaussian distribution, an assumption that is often not justified by the data. In addition, application of such methods typically requires considerable human (subjective) input (Diggle et al., 2007).

A recently proposed approach focuses on the modelling of spatial correlations by means of short-range interactions, inspired from models of statistical physics (Hristopulos, 2003; Hristopulos and Elogne, 2007). These works focused on the development of computationally efficient spatial dependence models that are applicable to gridded and scattered Gaussian data. The objective of the present study is to extend the scope of the interaction-based representation to non-Gaussian data. Herein, we formulate a classification model inspired from the Ising spin model used in statistical physics and demonstrate its application on remotely sensed earth surface elevation data. The proposed model is computationally fast, it is non-parametric (i.e., it makes no assumption on the data distribution), and it requires minimal input by the user. Hence, it is suitable for the automatic classification and mapping of large, non-Gaussian data.

DEFINITION OF THE PROBLEM

Let us consider a set of sampling points $G_s = \{\vec{r}_i\}$, where $\vec{r}_i = (x_i, y_i) \in \mathbb{R}^2$ and $i = 1, ..., N$. These points are assumed to be scattered on a rectangular grid $\tilde{G}$ of size $N_G = L_x \times L_y$, where L_x and L_y are respectively the horizontal and vertical dimensions of the rectangle (in terms of the unit length) and $N < N_G$. Let z_i be a value attributed

to the point $\vec{r}_i$. Then, the set $Z_s = \{z_i \in \mathbb{R}\}$ represent the sample of the process. Let $G_p = \{\vec{r}_p\}$ be the set of prediction points where $p = 1, \ldots, P$, such that $\tilde{G} = G_s \cup G_p$. For many applications Z_s can be represented as a realization of a continuous underlying random field $Z(\vec{r}_i)$. In the following we discretize the continuous distribution using a number of classes, $\mathcal{C}_q$, $q = 1, \ldots, N_c$. The classes are defined with respect to a set of threshold levels t_k, $k = 1, \ldots, N_c + 1$, where $t_1 = \min(z_1, \ldots, z_N)$ and $t_{N_c+1} = \max(z_1, \ldots, z_N)$. Each class $\mathcal{C}_q$ corresponds to an interval as follows: $\mathcal{C}_q = (t_q, t_{q+1}]$ for $q = 2, \ldots, N_c - 1$, $\mathcal{C}_1 = (-\infty, t_2]$, and $\mathcal{C}_{N_c} = (t_{N_c}, \infty)$. We define the *indicator field* $I_Z(\vec{r})$ to take integer values $q = 1, \ldots, N_c$ equal to the appropriate class index. In particular, $I_Z(\vec{r}) = q$ implies that $z(\vec{r}) \in \mathcal{C}_q$. The prediction problem is then posed as a classification problem, i.e., each point in G_p is assigned a class label. In order to estimate the class identity of Z at the prediction points, we use the well-studied Ising model, e.g. (McCoy et al., 1973). Once all the prediction points have been assigned to a class, a map of the process Z can be generated consisting of equivalent class (isolevel) contours.

THE ISING MODEL

Let us assume a set of variables s_i^q ("spins"), which can take the value $s_i^q = 1$ ("spin-up") or $s_i^q = -1$ ("spin-down"). The Ising model considers pairwise interactions between the spins, expressed by the following Hamiltonian:

$$H = -\sum_{i,j} J_{ij}\, s_i\, s_j - \sum_i h_i s_i. \tag{1}$$

The first term corresponds to the "spin-spin exchange" interaction energy. The coupling strength J_{ij} controls the strength as well as the type of the interaction: if $J_{ij} > 0$ it is "ferromagnetic" or if $J_{ij} < 0$ it is "antiferromagnetic". The second term corresponds to the symmetry-breaking bias, which is caused by the presence of a site dependent "external field" h_i. The latter controls the overall distribution of the "spin" values between 1 and -1 (the magnetization). The coupling strength is usually considered to be uniform and its range limited to nearest neighbors. However, the model can be generalized to include also non-uniform, longer-range couplings.

The probability density function of the spin configurations is expressed following the Gibbs formalism as:

$$f = \frac{e^{-H/k_B T}}{Z[H]}, \tag{2}$$

where k_B is the Boltzmann's constant and T the temperature. The partition function $Z[H]$ is a normalization factor obtained by summation of $e^{-H/k_B T}$ over all possible "spin

configurations". Hence, it is only a function of the model parameters J_{ij} and h_i, not of a particular configuration. In the forward problem, the coupling strength and the polarizing field are known and one is interested in the most probable spin configurations. In the inverse problem, the "spins" are known from the sampled values, and the estimation focuses on inferring the model parameters (e.g., by means of the maximum likelihood method). Following that, the optimal prediction of the "spin" values at unsampled locations can be performed by maximizing the conditional (on the data) probability f (or, equivalently by minimizing H) with respect to the unknown values. Unfortunately, the normalizing constant Z is in most cases intractable by analytical means, while its numerical evaluation is a computational bottleneck. Possible approaches to circumvent this problem, such as the maximum pseudo-likelihood approach (Besag, 1975) or various Markov Chain Monte Carlo estimation techniques (Chen et al., 1997), can be either very inaccurate or prohibitively slow respectively.

THE MULTILEVEL CLASSIFICATION ALGORITHM

Non-Parametric Nearest-Neighbor Model

In the following we use the ideas motivating the Ising Hamiltonian (Eq. 1). We propose a hierarchical scheme in which the sample, G_s^q and prediction, G_p^q grids are progressively updated for increasing class index q. For the lowest class $G_s^1 = G_s$, $G_p^1 = G_p$, where G_s and G_p are the initially defined grids. Let N_q denote the number of known spins at level q. The discretization is binary with respect to each threshold value, i.e., $s_{i_q}^q = -1$ if $z_{i_q} \leqslant t_{q+1}$ and $s_{i_q}^q = 1$ if $z_i > t_{q+1}$ for $i_q = 1, \ldots, N_q$. For $q > 1$ the sample (prediction) grid is augmented (diminished) by the points $\vec{r}_l$ for which $s_l^{q-1} = -1$. It follows that $N_1 = N$ and $N_{q>1} \geqslant N$. The set $S_s^q = \{s_{i_q}^q\}$, $q = 1, \ldots, N_c$; $i_q = 1, \ldots, N_q$ includes all the spin values for class index q. Let $\tilde{S}^q = S_s^q \cup S_p^q$ denote the "spin" values at the prediction grid $\tilde{G}$. In this framework, the Ising model can be used to represent spatial interactions between the spins $\tilde{S}^q$ with respect to the specific threshold.

In this study, we restrict the scope of the Ising model to the simplest energy functional: we set the polarizing field uniformly to zero, i.e., $h_i = 0, i = 1, ..., N$ and limit the exchange interactions to uniform strength, "ferromagnetic," i.e. $J_{ij} = J > 0$, with nearest neighbor coupling range. The choice of zero polarizing field does not allow to explicitly control the volume fraction of "up" versus "down" spins. As explained below, this is achieved in the simulations by selecting the initial "spin" values so as to reflect the "up-down" distribution of the sample.

The choices made above help to avoid the parameter inference problem and suggest a non-parametric approach.

This approach utilizes a cost function, $U(S_p^q|S_s^q)$, that describes the deviation (the error) between a suitably normalized energy of the simulated distribution, $\tilde{C}^q$, and its counterpart, C_s^q, for the sample distribution.

$$U(S_p^q|S_s^q) = \begin{cases} \left(1 - \tilde{C}^q/C_s^q\right)^2, & \text{for } C_s^q \neq 0, \\[2ex] (\tilde{C}^q)^2, & \text{for } C_s^q = 0, \end{cases} \tag{3}$$

where $C_s^q = \langle s_i^q s_j^q \rangle_{G_s^q}$ is the spin pair correlation of the $q-$level sample and $\tilde{C}^q = \langle s_i^q s_j^q \rangle_{\tilde{G}}$ is the spin pair correlation over the entire grid.

Given the above, the classification problem of estimating S_p^q is reduced to finding the optimal configuration $\hat{S}_p^q$ corresponding to the minimum of the cost function (3) at a fixed temperature T:

$$\hat{S}_p^q = \arg\min_{S_p^q} U(S_p^q|S_s^q). \tag{4}$$

Hierarchical Classification Strategy
We propose a hierarchical algorithm that proceeds sequentially by increasing the class index. The binary discretization and the classification are first performed with respect to the first (lowest) class and then repeated sequentially for the remaining classes. The "gaps" in the prediction grid, G_p are gradually filled as the algorithm proceeds through consecutive levels. At each level, all the locations identified as having -1 spin values at the lower levels, are used as input (sample data) for the current classification. The reduced prediction grid, G_p^q for the class index q contains P^q points so that for $q > q'$ it holds that $P^q \leqslant P^{q'}$ and $P^1 = P$. The classes $\mathcal{C}_q$ can be defined as desired and do not need to be uniform. The algorithm uses the *rejection ratio*, which is defined by:

$$\rho = \frac{\text{number of rejected states}}{\text{number of simulated states}}.$$

The algorithm consists of the following steps:

Algorithm

(i) *Initialize the indicator field on the entire grid by means of $\hat{I}_Z(\tilde{G}) = \text{NaN}$*
(ii) *set the class index to $q = 1$*
(iii) while [1] $q \leqslant N_c - 1$ *discretize Z_s with respect to t_{q+1} to obtain S_s^q*
(iv) *given the data S_s^q, calculate the sample correlation energy C_s^q*
(v) *assign initial values to the spins at G_p^q, i.e., generate $\hat{S}_p^{q\,(0)}$*
(vi) *calculate the initial values of the simulated correlation $\tilde{C}^{q\,(0)}$ and the cost function $U^{(0)}$; initialize the simulated*

states index $i = 1$
(vii) *initialize the rejection ratio $\rho \to 0$; and the rejected states index $i_r = 0$*
(viii) while [2] $\rho < 1$ *repeat the following updating steps:*
  (a) generate a new state $\hat{S}_p^{q\,(i+1)}$ by perturbing $\hat{S}_p^{q\,(i)}$
  (b) calculate $\tilde{C}^{q\,(i+1)}$ and $U^{(i+1)}$
  (c) if $U^{(i+1)} < U^{(i)}$ accept the new state; $i_r \to 0$.
    else keep the "old" state; $i_r \to i_r + 1$; endif
  (d) $\rho \to i_r/P^q$; $i \to i + 1$;
end [2]
(ix) *assign the -1 "spins" to the q level, i.e., $\hat{I}_Z(\{\vec{r}_i\}) = q$ if $\hat{S}^{(i_{\max})}(\{\vec{r}_i\}) = -1, \{\vec{r}_i\} \in \tilde{G}$*
(x) *increase class index, $q \to q + 1$, return to step (iii)*
end [1]
(xi) *for $q = N_c$ set $\hat{I}_Z(\{\vec{r}_i\}) = N_c \; \forall \vec{r}_i \; (i = 1, \dots, N_G)$ such that $\hat{I}_Z(\{\vec{r}_i\}) = \text{NaN}$.*

This algorithm returns an indicator field $\hat{I}_Z = I_Z(G_s) \cup \hat{I}_Z(G_p)$, which consists of the original sample discretization and the class estimates at G_p. The indicator values at the sampling sites are exactly reproduced, because the initial state respects these values and the updating steps skip over the sites in S_S^q. Below we refer to $I_Z(G_s)$ as the training set. The generation of new states in step (viii) is realized using single-spin Metropolis updating at zero temperature. The algorithm may sweep through the lattice several times, until the rejection ratio exceeds the threshold value. The $T = 0$ constraint means that there is no stochastic selection of unfavorable spins. Hence, the candidate "spin" for the update is flipped unconditionally only if it lowers the cost function. This is called a "greedy" Monte Carlo algorithm (Papadimitriou et al., 1982) and guarantees convergence, which is usually very fast. In comparison, in simulated annealing T is slowly lowered starting from an initial high-temperature state. This approach is much slower computationally, but the resulting configuration is less sensitive to the initial state $\hat{S}_p^{q\,(0)}$. The sensitivity of the greedy algorithm is known to be especially pronounced in high-dimensional spaces with non-convex energies. In such cases, the greedy algorithm is more likely to get stuck in local minima, instead of converging to the global one. However, this is not a concern in the current problem. In fact, targeting the global minimum of the cost fnction U strongly emphasizes the sample correlation energy per "spin" pair C_s^q, ignoring that the latter is influenced by sample-to-sample fluctuations.

The initial configuration can be selected in a number of ways. Since the proposed model aims to provide a fast and automatic classification mechanism, the initial configuration should minimize the relaxation path (in state space) to the equilibrium. It should also be selected preferably with little or no user intervention. Assuming a certain degree of spatial continuity, common in geospatial data sets, $\hat{S}_p^{q\,(0)}$

is determined based on the "augmented sample" states S_s^q in the immediate neighborhood of the individual prediction points. The neighborhood of $\vec{r}_p$ is determined by an $m \times m$ stencil (where $m = 2l + 1$) centered at $\vec{r}_p$. Then, $\hat{s}_p^{q\,(0)}$ is assigned by majority rule, based on the prevailing value of its neighbors from S_s^q that are situated inside the stencil. The stencil size is chosen automatically, reflecting the local sampling density and the "spin" values distribution. Namely, it is adaptively set to the smallest size, that contains a finite number of sampling points with one value (1 or -1) prevailing.

The updates are accepted unconditionally as long as they lower the cost function. The grid is swept randomly, point by point. The simulation terminates, for a given class index, if one complete sweep (i.e., P_q tried spin flips) through G_p^q does not produce a single successful update. This scheme implies an overall decreasing number of steps with increasing q in line with the size reduction of G_p^q. The parametrization required by the algorithm thus involves only the definition of the class intervals. The number of classes depends on the study objective: if the goal is to determine exceedance levels a binary classification is sufficient. For environmental monitoring and decision making purposes a moderate number (e.g., eight) of classes is often sufficient.

DATA AND SIMULATIONS

The data used in this study represent the surface elevation, $Z(\vec{r})$, on a 5-minute latitude/longitude grid, over a part of the territory of North America (approx. $80 - 110°$ W, $55 - 40°$ N). The data form a rectangle of 400×200 pixels (NOAA, 1988). Elevation values are in whole meters, representing the elevation of the center of each cell, with a resolution of 1 meter. Some summary statistics: number $= N_G = 80\,000$, $z_{\min} = 1$, $z_{\max} = 3790$, $\bar{z} = 774.41$, $z_{0.50} = 441$, $\sigma_z = 713.17$. The value of the skewness coefficient is 1.37 and of the kurtosis coefficient 4.07. As evidenced from the above statistics, the data are non-Gaussian and positively skewed. The histograms, corresponding to the respective class widths considered in the study, will be presented later.

From the complete data set, we generate a sample Z_s of size $N = (1 - p)N_G$ by randomly removing $P = pN_G$ values (validation points). For different degrees of thinning, $p = 0.33, 0.5$ and 0.66, we generate 100 different configurations of removed points. The values at the prediction points are then classified using the multilevel classification algorithm. In the classification performance evaluation, the indicator values at the prediction points $I_Z(G_p)$ will be compared with the estimates $\hat{I}_Z(G_p)$, obtained after removing the points G_p from the data. We use two resolutions and respective class intervals. First, a resolu-

tion of 500 m segregates the data into eight classes $C_q = [500(q - 1), 500q]$, $q = 1, \ldots, 8$. Second, a finer resolution of 250 m resulting in $N_c = 15$ classes corresponding to the intervals $C_q = [250(q - 1), 250q)$, $q = 1, \ldots, 15$ is used.

In order to evaluate the classification performance, we calculate the misclassification rate as a fraction of misclassified pixels, defined by

$$F = \frac{1}{P} \sum_{p=1}^{P} \left[1 - \delta\big(I_Z(\vec{r}_p), \hat{I}_Z(\vec{r}_p)\big) \right], \tag{5}$$

where $I_Z(\vec{r}_p)$ is the true value at the validation points, $\hat{I}_Z(\vec{r}_p)$ is the classification estimate and $\delta(I, I') = 1$ if $I = I'$, $\delta(I, I') = 0$ if $I \neq I'$.

The computations are performed in the Matlab® programming environment on a desktop computer with 1.93 GB of RAM and a the processor Intel®Core ™2 CPU 6320 @ 1.86 GHz 1.86 GHz.

RESULTS

Resolution 500 m - 8 classes

The isolevel map and the histogram of the original complete data, discretized according to the vector of thresholds corresponding to this resolution, are shown in Fig. 1. In the map, the elevations in the range of $0 \leqslant Z < 500$ m (first class) dominate, covering about 55% of the area, while those above 3500 m correspond only to 0.1%. Fig. 1(b) presents the histogram of the class values for the original data (left histogram bar) as well as the histograms of the data reconstructed by the classification for the three degrees of thinning. The histograms resulting by application of the classification algorithm to the training sets with $p = 0.33, 0.5$ and 0.66, are shown by the respective bars $2 - 4$ (moving from left to right). The histograms represent mean values obtained from 100 realizations. The match between the distributions of the original and the reconstructed data is excellent. As mentioned above, this was achieved without explicit control of the data volume fraction ($h_i = 0$ in Eq. 1).

Fig. 2 helps to visualize the classification results in terms of reconstructed maps. The isolevel maps on the left are obtained from a single realization (the first from the set of one hundred). The maps on the right represent the standard deviations of the class indices obtained from the 100 realizations. We observe that the reconstructed maps provide a close visual match to the original map, shown in Fig. 1(a), especially at lower p. Misclassification mainly occurs along the borders between different classes. The effect of class borders is better demonstrated in the plots of the right column, which show the spatial distribution of the standard deviations of the estimates. Note that the
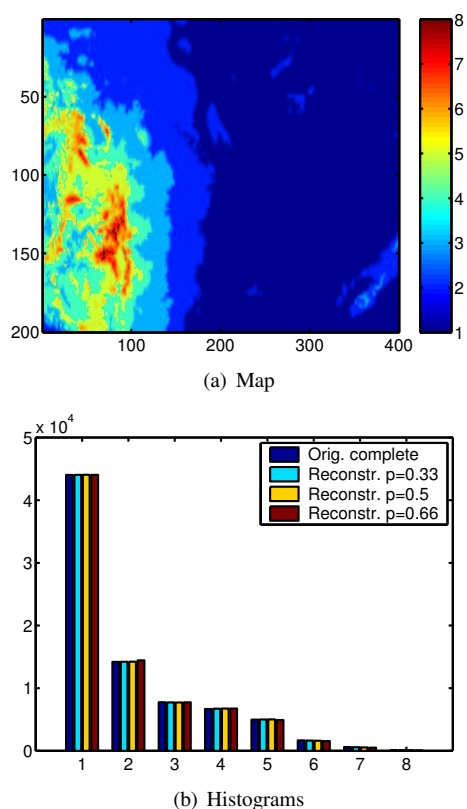
(a) Map

(b) Histograms

Figure 1: (a) 8-class isolevel map of the original complete data with the resolution 500 m. (b) Group of histograms. From left to right: the original complete data and the reconstructed ones from the thinned data with $p = 0.33, 0.5$ and 0.66, respectively. The histograms of the reconstructed data represent mean values, obtained from 100 realizations.

(a) p=0.33 Reconstructed map   (b) p=0.33 Standard deviations

(c) p=0.5 Reconstructed map   (d) p=0.5 Standard deviations

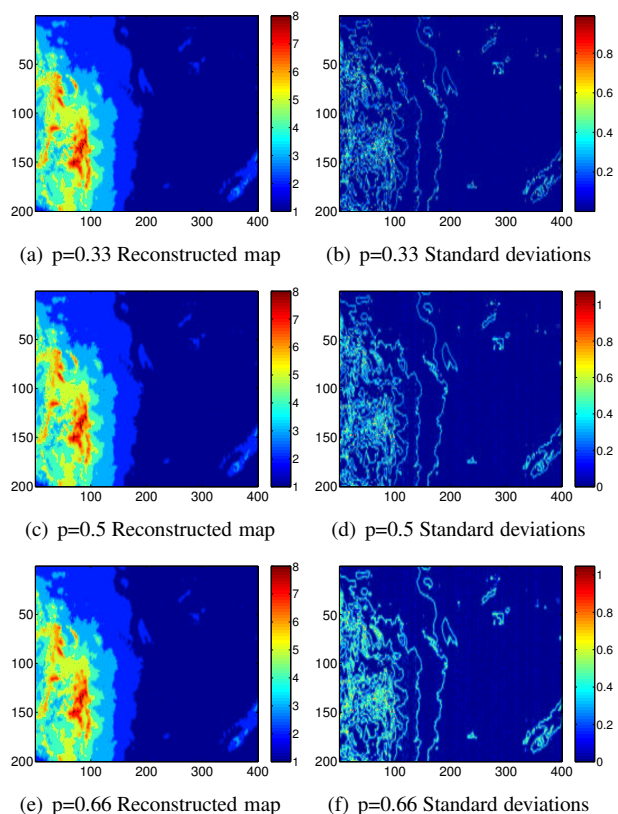(e) p=0.66 Reconstructed map   (f) p=0.66 Standard deviations

Figure 2: Examples of class maps obtained from individual realizations (a,c,e) and standard deviations of the class indices calculated from 100 different realizations (b,d,f). The data are reconstructed from samples with thinning degrees $p = 0.33, 0.5$ and 0.66, respectively. An 8-level classification scheme is used.

sampling points locations G_s (vary from realization to realization) bring zero contribution by construction, due to honoring the values at G_s. The spatial patterns in the original map, however, are reconstructed reasonably well even at $p = 0.66$.

Table 1 presents quantitative measures of the classification performance. It includes the mean value of the misclassification rate $\langle F^* \rangle$ and its standard deviation STD_{F^*}. The optimization's computational performance is measured in terms of mean number of Monte Carlo steps needed to achieve the optimum $\langle N_{MC} \rangle$, the optimization CPU time $\langle T_{cpu} \rangle$, and the cost function mean value $\langle U^* \rangle$ at termination. As expected, the misclassification rate increases with p.

The increase of $\langle N_{MC} \rangle$ with p is due to the fact that $\langle N_{MC} \rangle$ is a measure of the "spin" system relaxation time, while an increasing p translates into higher N and, thus, also larger state-space size 2^N. Since N_q decreases with

increasing q, due to the progressive filling of the gaps, the Metropolis sampler in general speeds up as q increases. The relaxation time is shortened by proper choice of the initial state. To give an approximate estimate, 2.08×10^5 Monte Carlo steps at $0.66 \times 400 \times 200$ prediction points correspond to roughly 4 grid sweeps. There are inter-level differences in the value of $U(S_p^q | S_s^q)$. Nevertheless, even at $p = 0.66 \max(U^*) < 7 \times 10^{-4}$. The average CPU time needed for the optimization at $p = 0.33, 0.5$ and 0.66 is approximately 6, 8, and 14 seconds, respectively.
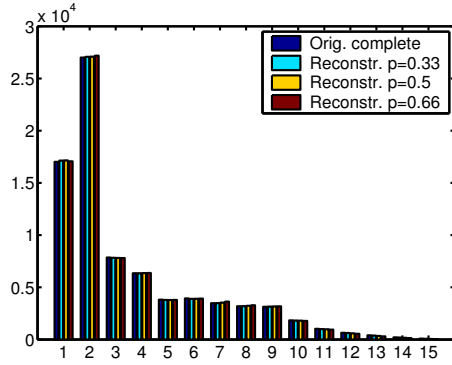
Resolution 250 m - 15 classes

Next, we repeat the classification experiment using a resolution of 250 m. The isolevel map 3(a) has higher resolution than the eight-level one in 1(a). The most and least represented classes are the second and the last ones, which contain approximately 34% and 0.1% of the values, respectively. Due to the higher resolution an increase in the mis-

Table 1: Eight-level classification performance measures: mean values of the misclassification rate $\langle F^* \rangle$ and its standard deviation STD_{F^*}, mean number of Monte Carlo steps $\langle N_{MC} \rangle$, optimization CPU time $\langle T_{cpu} \rangle$, mean value of cost function at termination $\langle U^* \rangle$. The averaging is performed over 100 realizations.

8 classes	$p = 0.33$	$p = 0.5$	$p = 0.66$
$\langle F^* \rangle$ [%]	5.27	5.79	6.48
STD_{F^*}	0.12	0.10	0.12
$\langle N_{MC} \rangle$	9.10e+4	1.35e+5	2.08e+5
$\langle T_{cpu} \rangle$ [s]	5.55	7.95	13.60
$\langle U^* \rangle$	8.84e-5	1.44e-4	1.83e-4
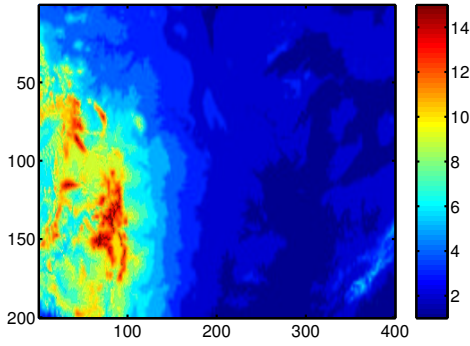
(a) Map

(b) Histograms

Figure 3: (a) 15-class isolevel map of the original complete data with the resolution 250 m. (b) Group of histograms. From left to right: the original complete data and the reconstructed ones from the thinned data with $p = 0.33, 0.5$ and 0.66, respectively. The histograms of the reconstructed data represent mean values, obtained from 100 realizations.

classification rate is expected. Nevertheless, as we can see in Fig. 4, both the class distributions and the visual patterns are recovered quite well in all cases. The statistics of the

(a) p=0.33 Reconstructed map

(b) p=0.33 Standard deviations

(c) p=0.5 Reconstructed map

(d) p=0.5 Standard deviations

(e) p=0.66 Reconstructed map
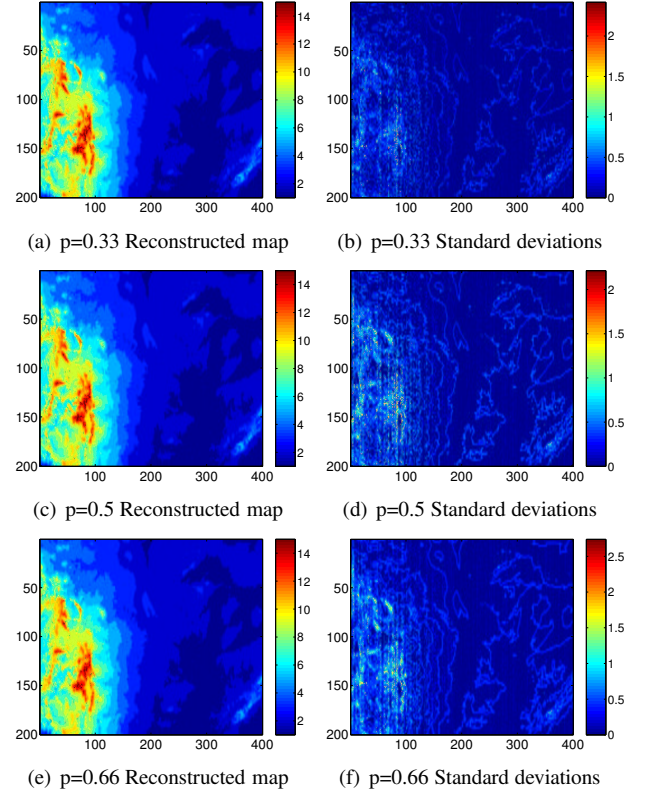

(f) p=0.66 Standard deviations

Figure 4: Examples of class maps obtained from individual realizations (a,c,e) and standard deviations of the class indices calculated from 100 different realizations (b,d,f). The data are reconstructed from samples with thinning degrees $p = 0.33, 0.5$ and 0.66, respectively. An 15-level classification scheme is used.

15-level optimization are given in Table 2.

CONCLUSIONS

We presented a novel classification model suitable for automatic mapping of large spatial data sets and validated it on the real world surface elevation data. The model is inspired from statistical physics and is based on minimizing distance between the normalized discrete correlation energies of the sample and the prediction domain data. It is computationally fast, universal with regard to the data distribution (makes no assumptions), and is almost automatic. The only parametrization in the proposed approach involves only the number of classes.

We have also used a deterministic updating scheme (results not shown here), in which the lattice is swept through sequentially. This scheme converges faster than the stochastic updating. For comparison, using the deterministic updating for the eight-level classification with thinning degree $p = 0.33$ we obtain $\langle T_{cpu} \rangle = 4.11$ sec,

Table 2: Fifteen-level classification performance measures. The same quantities are tabulated as in Table 1.

15 classes	$p = 0.33$	$p = 0.5$	$p = 0.66$
$\langle F^* \rangle$ [%]	11.93	12.93	14.23
STD_{F^*}	0.19	0.16	0.15
$\langle N_{MC} \rangle$	1.58e+5	3.16e+5	5.13e+5
$\langle T_{cpu} \rangle$ [s]	10.20	16.43	29.15
$\langle U^* \rangle$	9.07e-5	1.50e-4	1.97e-4

$\langle N_{MC} \rangle = 4.03 \times 10^4$ in contrast with $\langle T_{cpu} \rangle = 5.55$ sec and $\langle N_{MC} \rangle = 9.10 \times 10^4$ for the stochastic scheme. The misclassification rate is almost identical, $\langle F^* \rangle \approx 5.26\%$. The disadvantage of the deterministic scheme is that it generates certain artifacts. One of them is increased sensitivity to the initial conditions used, leading to high variability near the domain edge where the updating originates. Although we did not observe it at $p = 0.33$, it becomes more pronounced as p increases.

The model is demonstrated on a regular grid, however, the extension to irregularly spaced sampling points is straightforward. The interaction constant J_{ij} in Eq. (1) can be defined through a kernel function (such as the radial basis function), and the interaction neighborhood (nearest neighbors), for example, as pairs of points whose Voronoi cells have a common boundary. Further possible extensions could include incorporation of further-neighbor or/and "multi-spin" correlation energy in the Hamiltonian. Another possibility is to consider the q-state Potts model, instead of the sequential Ising one. Then the simulations could be performed simultaneously at all levels, however, as we also observed in our tests, the increased state space would result in much slower relaxation and hence much longer optimization times. The advantage would be in the initial state selection, since the "spins" of a range of states within the stencil would participate in the majority vote.

ACKNOWLEDGMENTS

REFERENCES

Atkinson P. M., Tate, N. J. (Eds.) (1999). *Advances in Remote Sensing and GIS Analysis*. John Wiley & Sons.

Diggle P. J., Ribeiro, P. J. (2007). *Model-based Geostatistics. Series: Springer Series in Statistics*. Springer.

McCoy B. M., Wu, T. T. (1973). *The Two-Dimensional Ising Model*. Harvard University Press, Cambridge Massachusetts.

Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician*, 24:179–195.

Chen, M. H., Shao, Q. M. (1997). On Monte Carlo methods for estimating ratios of normalizing constants. *Ann. Statist.*, 25:1563–1594.

Hristopulos, D. T. (2003). Spartan Gibbs random field models for geostatistical applications. *SIAM J. Scient. Comput.*, 24(6):2125–2162.

Hristopulos, D. T. and Elogne, S. N. (2007). Analytic properties and covariance functions for a new class of generalized Gibbs random fields. *IEEE Trans. Inform. Theor.*, 53(12):4467–4679.

Data Announcement 88-MGG-02, Digital relief of the Surface of the Earth. NOAA, National Geophysical Data Center, Boulder, Colorado, 1988.

Papadimitriou C. H., Steiglitz, K. (1982). *Combinatorial Optimization*. Prentice Hall.

Wackernagel H. (2003). *Multivariate Geostatistics*. Springer.

AUTHOR BIOGRAPHIES

MILAN ŽUKOVIČ was born in Svidnik, Slovakia and obtained his PhD degree in applied physics from Kyushu University, Japan in 2000. He pursued his research in the field of condensed matter physics at Kyushu University for 2 more years, before he assumed a position in automotive industry within Yazaki Corporation. Since October 2006, he has been involved in the research in modeling of spatial random fields as a Marie-Curie postdoc at Technical University of Crete, Greece. His e-mail address is : mzukovic@mred.tuc.gr.

DIONISSIOS HRISTOPULOS was born in Neo Faliro (Pireas), Greece. He holds a Diploma in Electrical Engineering (1985) from the National Technical University of Athens (Greece) and a PhD in Physics from Princeton University (1991). After two years of mandatory military service in Greece, he joined the Department of Environmental Sciences and Engineering at the University of North Carolina (Chapel Hill, USA). In 2000 he moved to the Pulp and Paper Research Institute of Canada to work in the field of Paper Physics. For research published with T. Uesaka, they were awarded the 2003 Johannes Van den Akker Prize for Advances in Paper Physics. Since 2002 he has been with the Department of Mineral Resources Engineering, at the Technical University of Crete. He recently became Professor of Geostatistics. His email is dionisi@mred.tuc.gr and his personal webpage at http://www.mred.tuc.gr/home /hristopoulos/dionisi.html.