Mapping an uncertainty zone between interpolated types of a categorical variable

Article *in* Computers & Geosciences · March 2012

DOI: 10.1016/j.cageo.2011.09.005

8 authors, including:

Xiaomin Mao
China Agricultural University

Katsuaki Koike
Kyoto University

Alvaro Penteado Crósta
University of Campinas

Wang Chunying
North China University of Water Conservancy…

Computers & Geosciences 40 (2012) 146–152

Contents lists available at SciVerse ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Mapping an uncertainty zone between interpolated types of a categorical variable

J.K. Yamamoto [a], X.M. Mao [b,*], K. Koike [c], A.P. Crosta [d], P.M.B. Landim [e], H.Z. Hu [b], C.Y. Wang [b], L.Q. Yao [b]

[a] Department of Environmental and Sedimentary Geology, University of Sao Paulo, Sao Paulo, Brazil
[b] College of Water Conservancy and Civil Engineering, China Agricultural University, No. 17 Qinghua East Road, Beijing 100083, China
[c] Graduate School of Engineering, Kyoto University, Kyoto, Japan
[d] Department of Geology and Natural Resources, University of Campinas, Campinas, Brazil
[e] Department of Applied Geology, State University of Sao Paulo, Rio Claro, Brazil

ARTICLE INFO

Article history:
Received 11 December 2010
Received in revised form
10 September 2011
Accepted 12 September 2011
Available online 22 September 2011

Keywords:
Categorical variable
Indicator function
Indicator kriging
Interpolation variance

ABSTRACT

Categorical data cannot be interpolated directly because they are outcomes of discrete random variables. Thus, types of categorical variables are transformed into indicator functions that can be handled by interpolation methods. Interpolated indicator values are then backtransformed to the original types of categorical variables. However, aspects such as variability and uncertainty of interpolated values of categorical data have never been considered. In this paper we show that the interpolation variance can be used to map an uncertainty zone around boundaries between types of categorical variables. Moreover, it is shown that the interpolation variance is a component of the total variance of the categorical variables, as measured by the coefficient of unalikeability.

1. Introduction

Collecting field samples at certain locations, and then using the information gathered in a field campaign to produce maps showing the spatial distribution of analyzed variables is a common procedure in Earth sciences. Quite often the nature of the variables being analyzed is categorical. For example, a soil survey usually comprises the collection of soil samples and the assignment of soil types to the samples. In this case, soil types are categorical variables or, in statistical terms, discrete random variables.

In order to produce maps based on categorical variables it is necessary to interpolate values for unsampled locations in between collection points. Even when we code categorical data as numbers, varying from 1 to the number of categories or types, we cannot use this information to interpolate a value for unsampled locations. This problem requires coding available information as indicator functions that can be interpolated and backtransformed to the original types of categorical variables (Koike and Matsuda, 2005; Teng and Koike, 2007; Leuangthong et al., 2008).

After transforming categorical data into indicator functions, the indicator kriging approach (Journel, 1983) can be used to estimate values for unsampled locations. This approach requires calculating and modeling a number of experimental semivariograms equal to the number of types of categorical variables (Leuangthong et al., 2008). However, when there are types of variables presenting low proportions this approach is almost impossible in practical terms. The indicator semivariograms in these cases are based on a few possible pairs and they might present statistical fluctuations.

Furthermore, after the interpolation and the generation of maps showing the distribution of the categorical types, the uncertainty zones that represent the boundaries between adjacent zones of different categorical types need to be properly established. This is usually done by defining buffer zones. However, these buffer zones are defined arbitrarily, using a constant or a variable distance, in a procedure known in geographic information systems (GIS) as "proximity analysis" (Star and Estes, 1990).

In this paper we propose the use of multiquadric equations (Hardy, 1971) that do not depend on semivariogram models. Thus, indicator functions will be interpolated for unsampled locations using multiquadric equations. Interpolated indicator values are backtransformed into original and mutually exclusive types of categorical variables (Teng and Koike, 2007). In this way

* Corresponding author. Tel.: +86 10 82373119; fax: +86 10 62737796.
E-mail addresses: jkyamamo@usp.br (J.K. Yamamoto), maoxiaomin@tsinghua.org.cn (X.M. Mao), koike.katsuaki.5x@kyoto-u.ac.jp (K. Koike), alvaro@ige.unicamp.br (A.P. Crosta), plandim@rc.unesp.br (P.M.B. Landim), haizhuhu@163.com (H.Z. Hu), wangchunying1987@yahoo.com (C.Y. Wang), liqiang85327@gmail.com (L.Q. Yao).

0098-3004/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2011.09.005

we are able to produce a map of interpolated types of the categorical variable. There is an uncertainty zone around these boundaries because the resulting interpolated map is based on sample data points. We propose the use of the interpolation variance (Yamamoto, 2000) for types of categorical variables for mapping uncertainty for the transition zones between interpolated types. By incorporating this concept into the resulting map the limitations of using arbitrary zones that result from spatial buffering can be overcome.

2. Mapping an uncertainty zone between interpolated types of a categorical variable

2.1. Categorical variables

Categorical variables come from observations in which certain qualitative characteristics are recognized such as color, texture, and pattern. Variables measured on a nominal scale (Stevens, 1946) and on ordinal scale (Stevens, 1946) are called categorical variables. Ordinal scales assign numbers representing the rank order of certain characteristics (Stevens, 1946). For example, sediments can be described as fine, medium, and coarse, depending on the grain size. Variables measured on an ordinal scale can be analyzed as categorical variables as well.

Outcomes of discrete random variables cannot be combined directly to give the values at unsampled locations. However, some functions of these categorical variables can be used to estimate linearly the value at a given location or domain (Rivoirard, 1994). These functions, known as indicator functions, are used to indicate a type present within a categorical variable.

2.2. The indicator function

Given a categorical variable with K types, the indicator function for the kth type is defined as

$$I(x_i;k)=\begin{cases}1 \text{ if type } k \text{ is present at location } x_i\\ 0 \text{ if type } k \text{ is not present at location } x_i\end{cases}. \tag{1}$$

The indicator variable is also known as an all-or-nothing variable (Journel and Huijbregts, 1978), because within K types of a categorical variable just one type k will have a value equal to one and all other equal to zero.

The mean of the indicator variable can be calculated as

$$E\left[I(x;k)\right]=\frac{f_k}{N}=p_k, \tag{2}$$

where p_k is the proportion of type k and $N=\sum_k f_k$ is the total number within the domain.

The variance of the indicator variable is

$$Var[I(x;k)]=E[I^2(x;k)]-(E[I(x;k)])^2=p_k-p_k^2=p_k(1-p_k). \tag{3}$$

Note that $E[I^2(x;k)]=E[I(x;k)]$. The variance is therefore the proportion of type k times the proportion of types different to k.

A single indicator variable that has two possible outcomes, one or zero, follows the Bernoulli distribution (Kader and Perry, 2007). Let p_1 be the proportion of ones and p_2 the proportion of zeroes then we have the mean equal to p_1 and the variance equal to p_1p_2 (Kader and Perry, 2007).

When we have K types within a categorical variable, then we have K indicator variables that follow a categorical distribution as a generalization of the Bernoulli distribution. Since indicator variables are mutually exclusive and exhaustive $\sum_k^K i(x;k)=1$ (Leuangthong et al., 2008) the categorical distribution is a special case of the multinomial distribution.

In expression (3) we can calculate the variance for kth indicator function. The global variance for all K types is given by the coefficient of unalikeability proposed by Kader and Perry (2007)

$$\mu_2=\sum_k^K p_k(1-p_k). \tag{4}$$

According to Kader and Perry (2007), the coefficient of unalikeability gives the proportion of possible comparisons that are unalike.

Now K indicator variables replace a categorical variable with K types. Indicator variables can be combined linearly to obtain estimated values at unsampled locations.

2.3. Indicator kriging

Indicator kriging is the most common interpolator to estimate every category type k at an unsampled location x_o as follows:

$$i_{IK}^*(x_o;k)=\sum_{i=1}^n \lambda_i i(x_i;k). \tag{5}$$

For example, if k=A it means that we are estimating the probability that the category type is A at location x_o

$$i_{IK}^*(x_o;k)=P(x_o;k=\mathrm{A}).$$

We can also calculate the uncertainty associated with the indicator kriging estimate (4) as

$$s_o^2(x_o;k)=\sum_{i=1}^n \lambda_i[i(x_i;k)-i_{IK}^*(x_o;k)]^2. \tag{6}$$

This is none other than the interpolation variance proposed by Yamamoto (2000). Rewriting this expression we obtain

$$s_o^2(x_o;k)=\sum_{i=1}^n \lambda_i i^2(x_i;k)-(i_{IK}^*(x_o;k))^2.$$

Since $\sum_{i=1}^n \lambda_i i^2(x_i;k)=\sum_{i=1}^n \lambda_i i(x_i;k)$ the interpolation variance can be written as

$$s_o^2=i_{IK}^*(x_o;k)-(i_{IK}^*(x_o;k))^2=i_{IK}^*(x_o;k)(1-i_{IK}^*(x_o;k)).$$

For example, for k=A the interpolation variance is

$$s_o^2(x_o;k=\mathrm{A})=P(x_o;k=\mathrm{A})P(x_o;k\neq\mathrm{A}), \tag{7}$$

and, therefore, the variance is equal to the product of probabilities that the category type at location x_o is A and that the category type is different than A.

The indicator kriging approach requires K indicator semivariogram models (Leuangthong et al., 2008). This is very difficult because some types can present just few data points and consequently few pairs presenting large statistical fluctuations. Thus, instead of indicator kriging we can apply multiquadric equations for interpolation of indicator variables at unsampled locations.

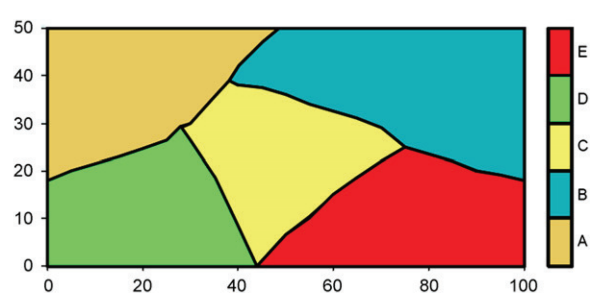

Fig. 1. Exhaustive data set showing a categorical variable with 5 types.

2.4. Multiquadric equations

Multiquadric equations can be used to interpolate an unsampled location using an equivalent expression as (5)

$$i^{*}_{MQ}(x_o;k) = \sum_{i=1}^{n} w_i i(x_i;k). \tag{8}$$

Table 1
Areas and proportions of categories for the exhaustive data set.

Type	Calculated area	%	Number of pixels	%
A	978.03	19.56	981	19.62
B	1283.08	25.66	1279	25.58
C	890.37	17.81	890	17.80
D	892.08	17.84	891	17.82
E	956.32	19.13	959	19.18
Total	4999.88	100.0	5000	100.0

Table 2
Distribution of types within stratified random samples.

Sample	Types					Total	%
	A	B	C	D	E		
Sample10	2	3	1	2	2	10	0.20
Sample21	4	6	4	5	2	21	0.42
Sample32	8	8	7	4	5	32	0.64
Sample50	8	13	10	10	9	50	1.00
Sample72	15	18	14	12	13	72	1.44
Sample98	19	23	16	20	20	98	1.96

Actually, this expression results from developing a dual system introduced by Hardy (1977) to compare covariance and multiquadric methods. The weights of multiquadric equations come from the solution of a system of linear equations as follows:

$$\begin{cases} \sum_j w_j \phi(x_j - x_i) + \mu = \phi(x_o - x_i) \text{ for } i = 1,\ldots,n \\ \sum_j w_j = 1 \end{cases}.$$

This is none other than the ordinary kriging system in which the semivariogram function is replaced by the basis function (ϕ).

There are several options (Kansa, 1990; Beatson and Newsam, 1992) for the basis function (ϕ), but the most used are

Linear	$\phi(x) = \lvert x \rvert$
Cubic	$\phi(x) = \lvert x \rvert^3$
Generalized multiquadric	$\phi(x) = (c + \lvert x \rvert^2)^{(2k+1)/2}$, for $k = -1, 0, \ldots$
Splines	$\phi(x) = \lvert x \rvert^2 \log \lvert x \rvert$
Gaussian	$\phi(x) = \exp(-c\lvert x \rvert^2)$,

where $\lvert \bullet \rvert$ denotes norm of a vector in $\mathbf{R}^n$ and c is a positive constant.

For the generalized multiquadric, for k equal to -1 we have the reciprocal multiquadric kernel and when k is equal to zero we obtain the original multiquadric kernel (Hardy, 1971). We have used the multiquadric kernel with the arbitrary constant c equal to zero because, according to Hardy (1971), this constant might be appropriate but not necessarily optimal.

Among available radial basis functions the multiquadric kernel is the most widely used, because it always produces a nonsingular matrix of coefficients. Moreover, according to Franke (1982), the multiquadric equations, when compared to other interpolation

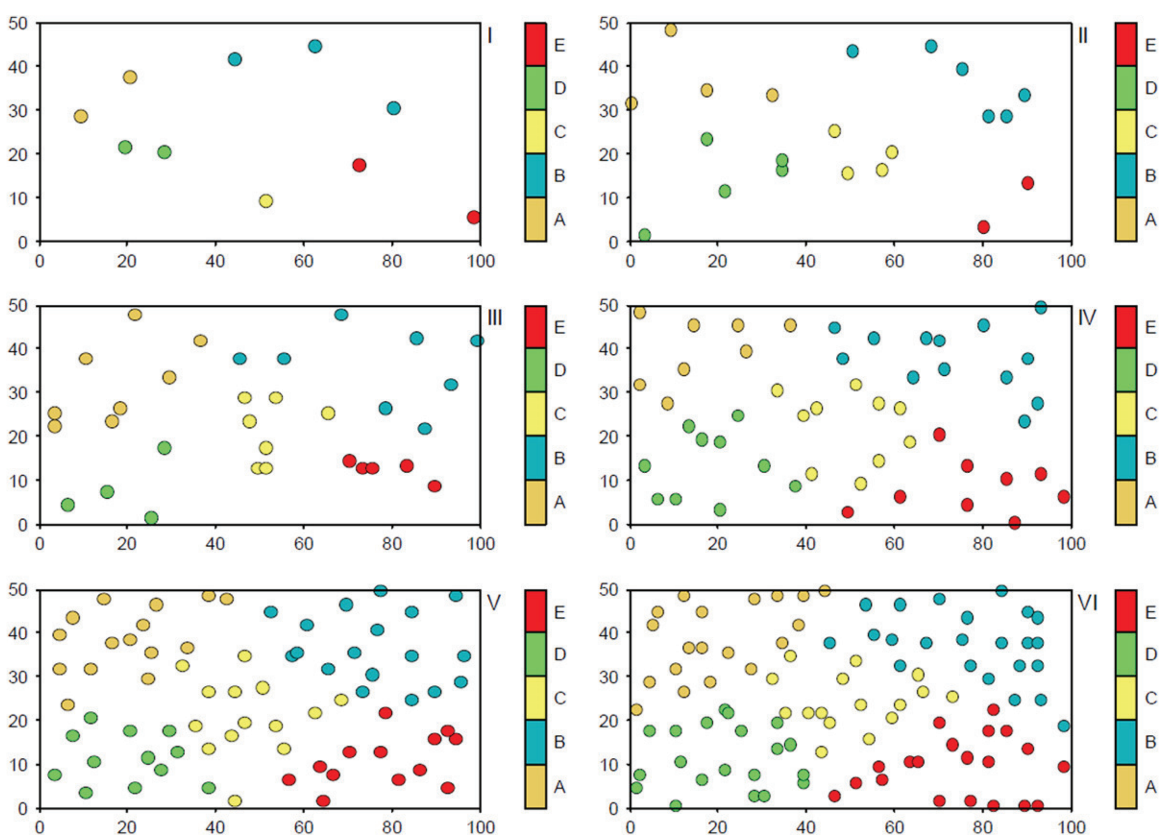

Fig. 2. Location maps for stratified random samples: I=10, II=21, III=32, IV=50, V=72, and VI=98 data points (after Yamamoto et al., 2011).

methods, give the most accurate results in terms of fitting and visual smoothness. Splines and cubic kernels call for coordinate transformation (usually normalization) because they rise rapidly with $|x|$. In this regard, the performance of multiquadric equations is sensitive to scaling; i.e., it is important to the condition number whether distances are expressed in centimeters or meters (Kansa, 1990). In this paper we will consider just the original multiquadric kernel as proposed by Hardy (1971), that does not take into account any anisotropies.

Since Eq. (8) is a weighted average formula we can calculate the uncertainty associated with the interpolated value as follows:

$$s_o^2(x_o;k) = \sum_{i=1}^{n} w_i[i(x_i;k) - i^*_{MQ}(x_o;k)]^2. \tag{9}$$

Evidently this expression has the same interpretation as given in (7).

2.5. Negative weights

Weighted average formulas (5) and (8) can result in negative values when some weights are negative. When negative results occur the sum of indicator estimates at an unsampled location will not be equal to one and, consequently, will not be jointly exhaustive such that $\sum_k^K i^*(x_o;k) = 1$ (Leuangthong et al., 2008). Teng and Koike (2007) report that they got indicator values in the range -0.03 to 1.02 for a case of 3D geological modeling of rock types in a geothermal region using borehole column data.

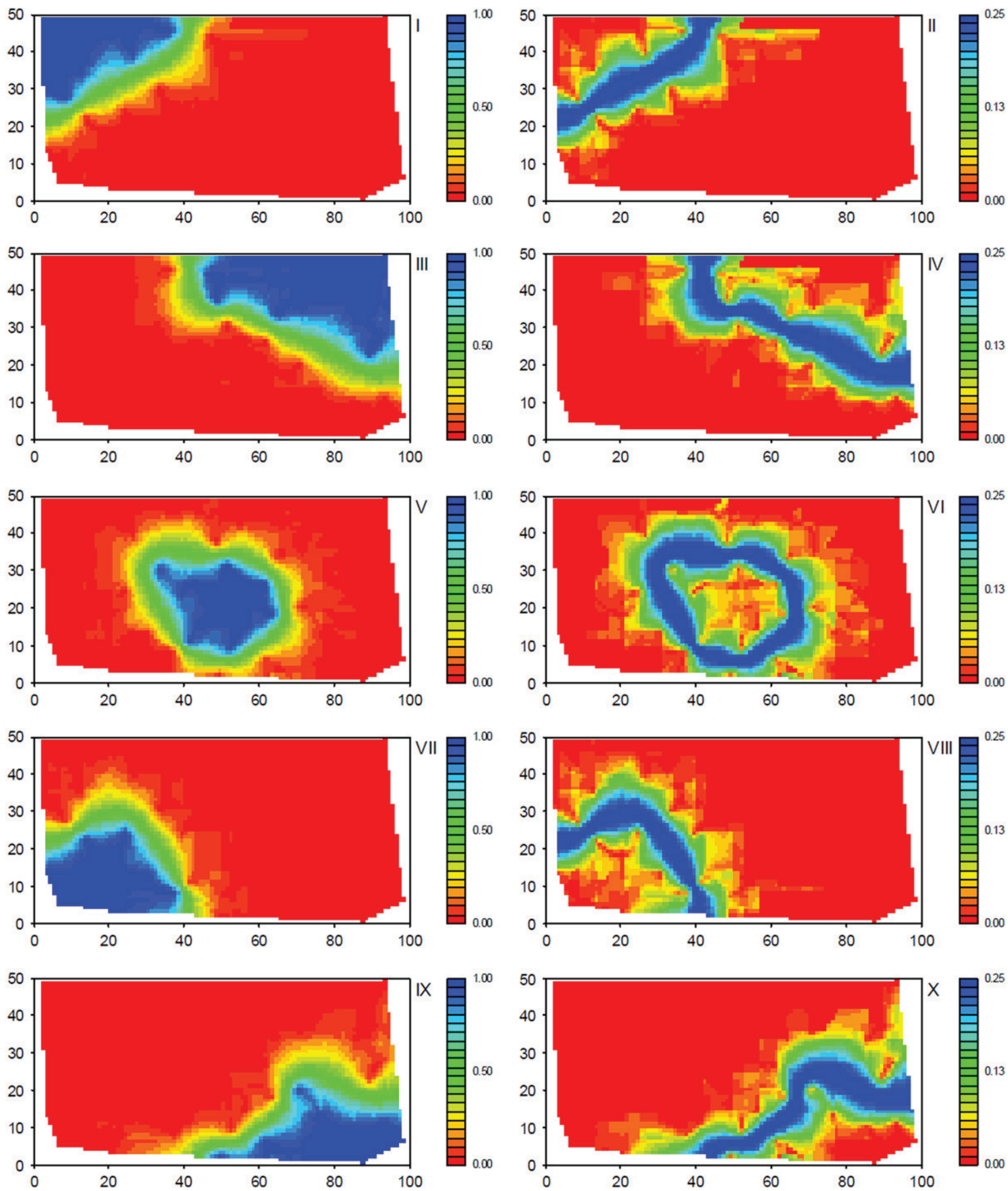

Fig. 3. Maps of interpolated probabilities on the left-hand side (Eq. (8)) and associated uncertainties on the right-hand side (Eq. (9)) for Sample50: I and II for type k=A; III and IV for type k=B; V and VI for type k=C; VII and VIII for type k=D; IX and X for type k=E.

Negative weights must be corrected to avoid these problems. Moreover, calculation of the interpolation variance is done with positive weights (Yamamoto, 2000).

The algorithm proposed by Rao and Journel (1997) was employed for correcting for negative weights. According to this algorithm, a constant equal to the modulus of largest negative weight is added to all weights and then recalculated to give a sum equal to one.

2.6. Backtransforming interpolated indicators to original type

Since all calculations have been made on indicator variables (Eq. (5) or (8)) we have to backtransform them to the original type of a categorical variable. In this paper we have considered the suggestion given by Teng and Koike (2007) in which the greater indicator is interpreted as the most likely type

$$i^{*}(x_o;k_{\max}) = \max(i^{*}(x_o;k), k=1,\ldots,K), \tag{10}$$

here $k_{\max}$ is the most likely type assigned to location x_o.

3. Materials and methods

We tested the procedure for interpolation of categorical variables departing from an exhaustive data set shown in Fig. 1. This map was discretized into 100 by 50 points. Areas and number of pixels inside each category are given in Table 1.

From this data set six stratified random samples with variable sizes have been drawn (Table 2 and Fig. 2). Sample size varies from 10 to 98 data points representing just 0.20–1.96% of the data set. The way the uncertainty zone varies with sample size will be shown next.

These samples will be used to interpolate at unsampled locations within the convex hull for each data set (Yamamoto, 1997). After obtaining interpolated maps we use them to make inferences about the population represented by the exhaustive data set. Statistical inferences based on samples are subject to some uncertainty that depends on the sample size. Actually, we demonstrate how reliable are interpolated maps according to the width of uncertainty zones. Unsampled locations will be interpolated based on multiquadric equations and nearest-neighbor information. An unsampled location will be interpolated using a total of 12 points taking the three closest points per quadrant. The indicator kriging will not be considered in this case since it is very difficult to compute a semivariogram for each type of the categorical variable presenting few data points, as we can see in Table 2.

4. Results and discussion

Before presenting the final maps displaying interpolated types and associated uncertainty zones, we discuss the procedure that results in a map with most likely types and the uncertainty zone. For illustration purposes we have chosen the sample with 50 data points. Fig. 3 shows probabilities maps on the left-hand side (Eq. (8)) and uncertainties maps on the right-hand side (Eq. (9)). In these maps it is clear that the uncertainty zone is close to the boundary between types of the categorical variable. The maximum uncertainty is equal to 0.25 and it occurs when two types have interpolated indicator functions presenting values equal to 0.5.

We can also examine the relationship between probabilities and associated uncertainties by plotting points on a scattergram. For all pairs probabilities versus uncertainties the scattergrams are exactly equal to the one shown in Fig. 4.

The quadratic function is described as

$$S_o^2(x_o;k=\mathrm{A}) = I_{MQ}^{*}(x_o;k=\mathrm{A}) - (I_{MQ}^{*}(x_o;k=\mathrm{A}))^2, \tag{11}$$

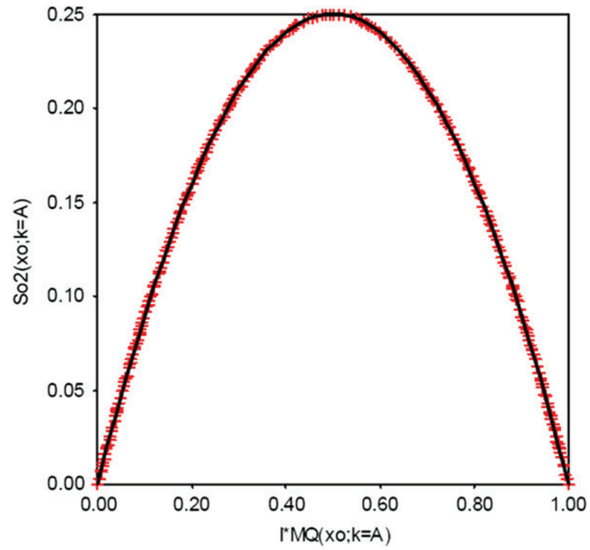

Fig. 4. Scattergram and a quadratic function fitted to all data for type A of the Sample50.

Table 3
Proportions of types of the categorical variable within different samples.

Sample	k=A	k=B	k=C	k=D	k=E	M
Sample10	0.08710	0.33965	0.15228	0.16171	0.25926	1995
Sample21	0.20845	0.21283	0.23774	0.25918	0.08179	3629
Sample32	0.24036	0.28582	0.21957	0.15334	0.10092	3663
Sample50	0.16509	0.26443	0.20824	0.17973	0.18251	4326
Sample72	0.19185	0.26887	0.22317	0.16012	0.16299	4156
Sample98	0.18516	0.22745	0.17835	0.19854	0.21050	4544

Table 4
Coefficients of unalikeability computed using Eq. (4).

Sample10	Sample21	Sample32	Sample50	Sample72	Sample98
0.76049	0.78086	0.77863	0.79385	0.79261	0.79844

which is exactly the formula for computation of the interpolation variance, reasserting the reliability of the interpolation variance.

Proportions of each type k of a categorical variable (Table 3) can be calculated as

$$p_k = \frac{\sum_{j=1}^{M} i_{MQ}^{*}(x_{o_j};k)}{M},$$

where M is the number of interpolated locations.

Coefficients of unalikeability (Eq. (4)) can be calculated directly from proportions (Table 3) and are shown in Table 4.

The global variance can also be calculated from probabilities and variances

$$S^2 = \frac{1}{M}\sum_{k=1}^{K}\left\{\sum_{j=1}^{M}(i_{MQ}^{*}(x_{o_j};k)-p_k)^2\right\} + \frac{1}{M}\sum_{k=1}^{K}\left\{\sum_{j=1}^{M}S_o^2(x_{o_j};k)\right\}. \tag{12}$$

here the first term on the right-hand side means the variance between pixels and global proportions, and the second term means the variance within pixels, which is given by the interpolation variance. Moreover, if we apply indicator kriging the traditional kriging variance would not be used to compute the global variance. Components of global variance are presented in Table 5.

Therefore, the coefficient of unalikeability is none other than the global variance. It shows that there is no difference between statistics for nonspatial and spatial data. Expression (12) comes from the calculation of the interpolation variance for an OK block estimate, as shown by Yamamoto (2000). Later Yamamoto (2001) showed that this expression could also be used for computation of the global variance of a mineral deposit. Both papers give a mathematical proof that the interpolation variance is a component of the global variance as computed after Eq. (12).

Now we must integrate all information into a single map, as illustrated in Fig. 5. In this map all interpolated indicator functions are transformed back into original qualitative types according to Eq. (10). In addition, it is necessary to decide which uncertainty level we want to display.

Table 5
Components of global variance after Eq. (12).

Sample	First term	Second term	Global variance
Sample10	0.29348	0.46701	0.76049
Sample21	0.41851	0.36235	0.78086
Sample32	0.46755	0.31108	0.77863
Sample50	0.55092	0.24293	0.79385
Sample72	0.57532	0.21729	0.79261
Sample98	0.61709	0.18135	0.79844

In this map we also draw the uncertainty zone given by interpolation variance greater than or equal to 0.20 and $i^*(x_o; k_{\max}) < 0.6$ (Eq. (10)). The value for the interpolation variance for mapping the uncertainty zone will depend on the number and spatial distribution of types of the categorical variable and the number and distribution of sample data.

Fig. 5 shows uncertainty zones around boundaries between different types of the categorical variable. It shows how reliable is the approach given by the interpolation variance as uncertainty measurement. The width of the uncertainty zone depends on the interpolation variance and this is the first approach that gives a calculated uncertainty zone instead of an arbitrary zone defined by the user, such as buffering in geographical information systems (GIS). Fig. 5 shows that the smaller the sample size, the greater the uncertainty zone. The sample representing 1% (Sample50) of the exhaustive data set shows a reasonable match between interpolated and real boundaries. On the other hand, the smaller sample (10 data points or 0.02% of the exhaustive data set) shows large errors, and this is the trade-off for working with small samples.

Since we have departed from an exhaustive data set, we can now compare interpolated types with actual types. The results are presented in Table 6, divided into certainty and uncertainty zones. A summary is shown in Table 7, showing that the size of the uncertainty zone decreases with increasing sample size. Moreover, if we sum up the matches in both, certainty and uncertainty zones (Table 7), we realize that the matches are very good

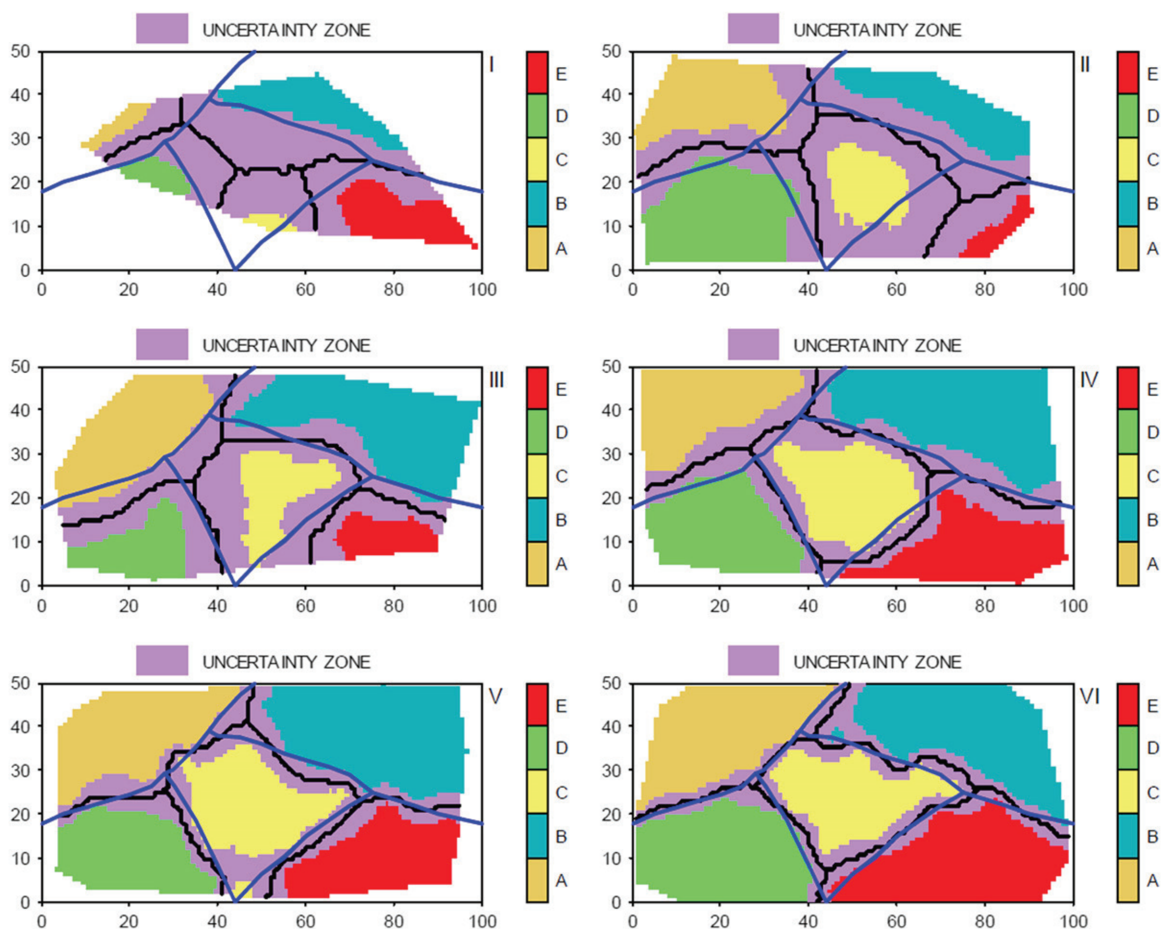

Fig. 5. Maps of interpolated indicator functions for types A, B, C, D, and E and uncertainty zones for interpolation variance greater than or equal to 0.20 and $i^*(x_o; k_{\max}) < 0.6$ (Eq. (10)). Thick black lines are boundaries between different types of the categorical variable and thick blue lines are the boundaries as seen on the exhaustive data set (Fig. 1). Legend: I = sample10, II = sample21, III = sample32, IV = sample50, V = sample72, and VI = sample98. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 6
Interpolated types according to certainty and uncertainty zones.

S	C	Certainty zone						Uncertainty zone					
		A	B	C	D	E	∑	A	B	C	D	E	∑
10	Match	91	353	56	109	381	990	63	20	156	4	62	305
	Do not	22	0	91	0	18	131	60	1	490	0	18	569
21	Match	571	540	388	717	147	2363	87	80	215	40	107	529
	Do not	48	0	71	0	137	256	58	9	211	5	198	481
32	Match	592	899	500	419	249	2659	13	53	201	109	87	463
	Do not	0	3	42	101	80	226	4	18	139	59	95	315
50	Match	617	1014	648	625	656	3560	102	68	137	39	79	425
	Do not	59	0	11	0	40	110	82	4	92	6	47	231
72	Match	682	964	729	582	594	3551	37	47	77	53	67	281
	Do not	7	22	49	17	22	117	12	61	34	48	52	207
98	Match	751	925	646	795	843	3960	38	65	99	38	34	274
	Do not	1	43	64	0	15	123	16	58	81	16	16	187

S: sample; C: comparison between actual and estimated types.

Table 7
Comparison between actual and estimated types split into certainty and uncertainty zones.

Sample	Total estimated	Certainty zone (in %)			Uncertainty zone (in %)		
		Estimated	Match	Do not	Estimated	Match	Do not
10	1995	56.19	49.62	6.57	43.81	15.29	28.52
21	3629	72.17	65.11	7.05	27.83	14.58	13.25
32	3663	78.76	72.59	6.17	21.24	12.64	8.60
50	4326	84.84	81.61	2.54	15.16	9.82	5.34
72	4156	88.26	85.44	2.82	11.74	6.76	4.98
98	4544	89.85	87.15	2.71	10.15	6.03	4.11

considering the sample size. For example, for the sample with just 32 data points we have a match of 85.23% between interpolated and actual types. We can also compute the error by summing up the nonmatching types in Table 7. The error decreases from 35.09% for Sample10 to 6.82% for Sample98. For the sample with 50 data points the error is 7.88%, a good result considering that this sample represents just 1% of the exhaustive data set. Actually, it proves the effectiveness of Eq. (10), proposed by Teng and Koike (2007).

5. Conclusions

This paper presents the use of multiquadric equations for interpolation of indicator functions at unsampled locations. The interpolation variance was used for mapping the uncertainty zone between types of a categorical variable. Moreover, it was shown there is a connection between the coefficient of unalikeability and the interpolation variance.

This approach presents a great potential for application in many fields that make use of categorical variables, for example, mapping faults from elevations of a layer based on a few drill holes; mapping the uncertainty between ore and waste; mapping soil types; etc.

Acknowledgments

J.K. Yamamoto, A.P. Crósta, and P.M.B. Landim acknowledge CNPq–Brazilian National Council for Scientific and Technological Development for supporting this research. The authors thank the anonymous reviewer for his valuable comments and constructive criticism that contributed to improve the manuscript.

References

Beatson, R.K., Newsam, G.N., 1992. Fast evaluation of radial basis functions: I. Computers & Mathematics with Applications 24 (12), 7–19.
Franke, R., 1982. Scattered data interpolation: test of some methods. Mathematics of Computation 38, 181–200.
Hardy, R.L., 1971. Multiquadric equations of topography and other irregular surfaces. Journal of Geophysical Research 76, 1905–1915.
Hardy, R.L., 1977. Least square prediction. Photogrammetric Engineering & Remote Sensing 43, 475–492.
Journel, A.G., 1983. Nonparametric estimation of spatial distributions. Mathematical Geology 15, 445–468.
Journel, A.G., Huijbregts, Ch.J., 1978. Mining Geostatistics. Academic Press, London 600 pp.
Kader, G.D., Perry, M., 2007. Variability for categorical data. Journal of Statistics Education 15 (2), 17 〈http://www.amstat.org/publications/jse/v15n2/kader.html〉(accessed on 22 June 2010).
Kansa, E.J., 1990. Multiquadrics – a scattered data approximation scheme with applications to computational fluid-dynamics – I. Computers & Mathematics with Applications 19, 127–145.
Koike, K., Matsuda, S., 2005. Spatial modeling of discontinuous geologic attributes with geotechnical applications. Engineering Geology 78, 143–161.
Leuangthong, O., Khan, K.D., Deutsch, C.V., 2008. Solved Problems in Geostatistics. John Wiley & Sons, Hoboken 208 pp.
Rao, S.E., Journel, A.G., 1997. Deriving conditional distributions from ordinary kriging. In: Baafi, E.Y., Schofiled, N.A. (Eds.), Geostatistics Wollongong'96, vol. I. , Kluwer Academic Publishers, Dordrecht, pp. 92–102.
Rivoirard, J., 1994. Introduction to Disjunctive Kriging and Non-linear Geostatistics. Oxford University Press, New York 181 pp.
Star, J., Estes, J., 1990. Geographic Information Systems: An Introduction. Prentice-Hall, Englewood Cliffs 303 pp.
Stevens, S.S., 1946. On the theory of scales of measurement. Science 103, 677–680.
Teng, Y., Koike, K., 2007. Three-dimensional imaging of a geothermal system using temperature and geological models derived from a well-log dataset. Geothermics 36, 518–538.
Yamamoto, J.K., 1997. Convex_Hull—a Pascal program for determining the convex hull for planar sets. Computers & Geosciences 23, 725–738.
Yamamoto, J.K., 2000. An alternative measure of the reliability of ordinary kriging estimates. Mathematical Geology 32, 489–509.
Yamamoto, J.K., 2001. Computation of global estimation variance in mineral deposits. In: Xie, H., Wang, Y., Jiang, Y. (Eds.), Computer Applications in the Minerals Industries, A.A. Balkema Publishers, Lisse, pp. 61–65.
Yamamoto, J.K., Kikuda, A.T., Koike, K., Mao, X., 2011. The indicator approach for measuring the correlation between categorical variables. In: Marschallinger, R., Zobi, F. (Eds.), Proceedings of the IAMG 2011 Conference, Salzburg. 〈www.iamg2011.at/IAMG_2011_program_web.pdf〉.