Proceedings of IAMG'05: GIS and Spatial Analysis, Vol.1, 651-656

A Compositional Approach to Indicator Kriging

R. Tolosana-Delgado[1], V. Pawlowsky-Glahn[1],
J.J. Egozcue[2] , K. G. van den Boogaart[3]

[1]*Dept. Informàtica i Matemàtica Aplicada, Universitat de Girona, E-17071 Girona (Spain);*
E-mail:raimon.tolosana@cdg.es
[2]*Dept. Matemàtica Aplicada III, U. Politècnica de Catalunya, E-08034 Barcelona (Spain);*
[3]*Institut für Mathematik und Informatik, Ernst-Moritz-Arndt-Universität Greifswald,*
D-17487 Greifswald (Germany)

1. Abstract

Instead of using indicator kriging to estimate discrete versions of cumulative distribution functions, with its order-relation problems, we suggest a compositional approach to the estimation of probability functions. A preliminary Bayesian treatment is needed to obtain estimates of the pdf (without zeroes) at sampled locations, and afterwards an isometric log-ratio transform provides unbounded real scores, which are finally kriged.

2. Introduction

Estimation of probability functions (pdf)—or their cumulative versions (cdf)—of regionalized variables is a common problem in geostatistics. Disjunctive kriging (DK) and indicator co-kriging (IK) are the most-used estimation techniques. These techniques aim to estimate the target pdf or cdf, conditional to observed data, by minimizing some square error difference between observed frequencies and estimated probabilities. To our understanding, this difference criterion is the cause of the so-called *order relation problems*: these methods frequently yield impossible probability estimates, such as negative probabilities. Here we present an alternative kriging technique, where the fit of the model is measured in terms of squared distances in the simplex, since relative frequencies and probabilities can be viewed as compositions. The methodology consists in two steps. First, the range of the regionalized variable is partitioned into D classes, and each observed value is translated into a probability vector following Bayesian techniques. This vector replaces the raw zero-one observation of a class given an opportunity to assess observation errors. This is then represented by log-ratio coordinates of the simplex, which gives a D-1 real vector without any constraint. The second step is the co-kriging of this coordinate vector of "observed" probabilities. Hypothesis about the Gaussian character of such a vector random function are now consistent and classical co-kriging techniques give a prediction of its coordinates at unsampled locations. These predictions are back-transformed into probability vectors, which are viewed as estimates of the sought discretized pdf of the regionalized variable. To test the reliability of this approach, we use a simulated example of a Gaussian random function, and then apply our technique to predict the conditional distribution at new locations.

3. Basics of Indicator Kriging

Let Z(x) be a real random function in a spatial domain, with image a set $A=[a_0, a_D]$. Let $\{x_1, x_2, ..., x_N\}$ be a set of sampled locations, and $\{z_1, z_2, ..., z_N\}$ be the values observed at these locations. We want to estimate, at an unsampled location x_0, the conditional probability distribution of Z(x_0). Assuming the random function to have a joint normal distribution, this conditional distribution is a normal distribution defined by the simple kriging (SK) predictor and its variance. If this joint normality assumption is not admissible, other techniques must be used, e.g. *indicator kriging* (Journel, 1983, IK). This technique estimates a discrete version of the cdf, by interpolating the step indicator transforms **I**(x) defined by a set of cutoffs $\{a_1, a_2, ..., a_{D-1}\}$. Alternatively, we can estimate a discrete version of the pdf by interpolating the equivalent set of disjunctive functions **J**(x), as *disjunctive kriging* (Matheron, 1976, DK) does. These two functions are defined by components as

$$I_i(x) = \begin{cases} 0, & \mathrm{Z}(x) < a_i; \\ 1, & \mathrm{Z}(x) \geq a_i. \end{cases} \quad J_i(x) = \begin{cases} 1, & a_{i-1} \leq \mathrm{Z}(x) < a_i; \\ 0, & \text{otherwise}. \end{cases} \tag{1}$$

The vectors **I**(x) and **J**(x) can be respectively viewed as degenerated cdfs and pdfs. Indeed, their expectation corresponds to the conditional functions they aim to estimate, thus the expected *difference between prediction and true value* is zero. Also, when these **I**(x) and **J**(x) are considered as random functions, a variographic analysis of these functions yields estimates of the equivalent marginal cdf or pdf, which correspond with their global means. In other words, knowledge of the variogram implies knowledge of the mean, and SK is then fully applicable. However, IK results are occasionally non-sense, as probabilities must be positive, and increasing—for **I**(x)—or sum up to one—for **J**(x)—. Both IK and DK present these problems.

4. Estimation of Multinomial Probability Vectors

Let us reinterpret the objects already introduced, ignoring the spatial dependency: Z is a random variable with image a set $A=[a_0, a_D)$, and $\{a_1, a_2, ..., a_{D-1}\}$ is a set of cutoffs used to define the partition $A=\cup A_i$, with $A_i=[a_{i-1}, a_i)$. Then, Eq. (1) defines a categorical random vector $\mathbf{J}=(J_i)$, with D components, such that $J_i=1$ if and only if $\mathrm{Z} \in A_i$, else $J_i=0$. This categorical vector follows a multinomial model: $\mathbf{J} \sim M(n=1, \mathbf{p})$ where $\mathbf{p}=(p_i)$ is a multinomial vector of probabilities, with $p_i = \Pr[J_i=1] = \Pr[\mathrm{Z} \in A_i]$. Note that $\mathbf{p}$ is a composition, thus its sample space is S^D, the D-part simplex (Aitchison, 1986).

Our first goal is the estimation of the vector $\mathbf{p}$ from a single observation of **J**. Assume that we observed category k, thus $J_k=1$. Classical IK applies the frequentist estimation formula *favourable/total cases*, and obtains Eq. (1) as estimator. Instead, we use a Bayesian approach. First, we encode all available information on $\mathbf{p}$ in a prior distribution, which might be of Dirichlet type (Haas and Formery, 2002)—including the uniform in S^D, if all categories are equally probable—or normal in the simplex (Mateu-Figueras et al., 2003). Then, it is updated by the likelihood of the sample, to obtain a posterior distribution—respectively a Dirichlet or an Aitchison's A distribution (Aitchison, 1986)—. Finally, a representative value $\mathbf{p}^*$ is extracted from this posterior. Changing the prior distribution and the loss function of the Bayesian estimation, we obtain different results. However, if the prior model treats all categories equally, all estimators are like

$$p_i^* = \begin{cases} a & i = k \\ \left(\dfrac{1-a}{D-1}\right) & i \neq k \end{cases} \tag{2}$$

which is a subjective assessment of the probability of actually having the i-th category in a place where the k-th category was observed. The value a is then the largest probability in $\mathbf{p}^*$, and for coherence, it should satisfy $a > 1/D$. Note that Eq. (2) does not take into account the order of the categories, as the underlying model is a multinomial one.

5. Kriging of Multinomial Probability Vectors

The sample space of both $\mathbf{p}(x)$ and its estimator $\mathbf{p}^*(x)$ is the D-part simplex S^D, since these vectors are compositions. It is then natural to follow Pawlowsky-Glahn and Olea (2004) in the geostatistical characterization of the vector $\mathbf{p}(x)$, considered as a multivariate random function. These authors suggested to apply Aitchison (1986) log-ratio transformations, interpolate the transformed values with a suitable kriging technique, and back-transform the results to obtain a composition. We select the isometric log-ratio transformation (Egozcue et al., 2003, ilr), defined in general as

$$\pi = \mathrm{ilr}(\mathbf{p}) = \Psi \cdot \ln(\mathbf{p}), \text{ with } \Psi^t \cdot \Psi = \mathbf{I}_D - \tfrac{1}{D}\mathbf{1} \text{ and } \Psi \cdot \Psi^t = \mathbf{I}_{D-1} \tag{3}$$

where $\boldsymbol{\pi}$ represents the vector of ilr-transformed scores, $\boldsymbol{\Psi}$ a matrix with D-1 rows and D columns, satisfying the kind of orthogonality conditions of Eq. (3), and $\mathbf{I}_D$ the D-dimen-sional identity matrix, and $\mathbf{1}$ a $D\times D$ square matrix with all elements equal to 1. The vector of transformed scores $\boldsymbol{\pi}$ has D-1 unbounded real values, suitable to be treated with all classical statistical methods, including standard variography and co-kriging techniques, using any existing software. In particular, given that the full vector is available at all sampled locations, the matrix co-kriging notation of Myers (1982) is highly useful. Results, denoted by $\boldsymbol{\pi}^*$, will be back-transformed into the simplex using the inverse ilr,

$$\mathbf{p}^* = \mathrm{C}\left(\exp\left(\Psi^t \cdot \pi^*\right)\right) \tag{4}$$

where $\mathrm{C}(\cdot)$ represents the closure operation, which divides all components by their total sum, thus forcing the result to sum up to one. Note that, due to this operation and the properties of the exponential, the final results $\mathbf{p}^*$ obtained with Eq.(4) will always be valid multinomial probabilities, with positive components summing up to one. In other words, this technique *never* presents problems like those of IK or DK.

Pawlowsky-Glahn (2003) provides us with tools to understand the properties of this estimator (Eq. 4) in the non-regionalized case. The simplex has an Euclidean space structure, in which Eq. (3) is the expression of the so-called ilr coordinates with respect to a basis of the simplex characterized by the matrix $\boldsymbol{\Psi}$. In such an Euclidean space, we may define a characteristic measure and a normal distribution completely embedded in the simplex (Mateu-Figueras et al., 2003). Using these tools, it can be shown that the estimator of Eq. (4) minimizes the Aitchison distance $d_A(\mathbf{p}^*(x_0),\mathbf{p}(x_0))$ between the prediction $\mathbf{p}^*(x_0)$ and the true value $\mathbf{p}(x_0)$, characterizing the distribution of $\mathbf{J}(x_0)$,

$$\mathbf{p}^* = \arg\min d_A\left(\mathbf{p}^*, \mathbf{p}\right) = \arg\min \sqrt{\frac{1}{D}\sum_{i<j}\left(\ln\frac{p_i^*}{p_j^*} - \ln\frac{p_i}{p_j}\right)^2}$$

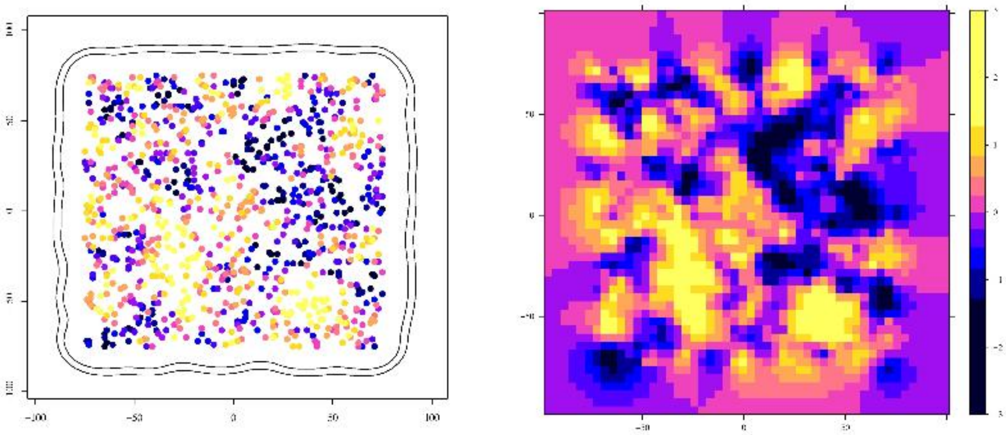
Also, assuming the random function $\mathbf{p}(x)$ to have a joint normal distribution on the simplex, SK prediction of the ilr coordinates $\boldsymbol{\pi}(x)$ and their error variance-covariance matrix give the parameters of the true distribution of $\mathbf{p}(x_0)$ conditional on the observed data set. This distribution follows a normal model on S^D,

$$\mathbf{p}(x_0) \sim N_{S^D}\left(\pi^*, \sigma_{SK}^2\right) \Leftrightarrow \pi(x_0) \sim N^{D-1}\left(\pi^*, \sigma_{SK}^2\right)$$

which is a classical normal distribution defined on the ilr coordinates $\boldsymbol{\pi}(x_0)$ themselves. Also, ordinary or universal kriging represent valid approximations to this conditional distribution up to the same extent they are for a conventional Gaussian random function. Finally, it can be shown that the estimator $\mathbf{p}^*$ and this conditional distribution do not depend on the chosen basis in the simplex S^D.

6. Case Study

To assess the goodness of the proposed method, we have simulated a zero-mean Gaussian random field Z(x), with an exponential variogram of c = 1 (sill) and a = 30u (effective range). The data set $\{x_1, x_2,\ldots, x_N\}$ contains N=1000 samples located at random in a 1u grid of 150×150 u^2 (Fig. 1A). The conditional distribution of the nodes of a 4u grid of 200×200 u^2 is obtained using SK (Fig. 1B). Applying a set of cuttoffs (Table 1, and legend of Fig. 1B) to this data, we compute the observations of the vector random function $\mathbf{J}(x)$, represented in Fig. 1A as a categorical variable. To estimate the multinomial probability vector $\mathbf{p}(x_n)$ at each one of these sampled locations we follow the Bayesian approach described in section 4. Given that the categories are equally-distributed (Table 1), we may use Eq. (2), and we decide to consider the parameter a=0.95, a classical value in statistics. In other words, after observing category k at a given location, we attach a confidence of 95% to it, and a (5/9)% to the other categories.



A B

Fig. 1. maps of data set (A) and SK predictions (B) of the case study. The color scale corresponds to the categories defined in Table 1, and it is valid for both maps. In (A), two lines mark the levels of 0.90 and 0.95 of kriging variance.

Table 1 Cutoffs of definition of the categories.

I	1	2	3	4	5	6	7	8	9	10
a_i	-1.28	-0.84	-0.52	-0.25	0	0.25	0.54	0.84	1.28	$+\infty$
$\Pr[Z<a_i]$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0

An arbitrary coordinate system was chosen, variograms were computed (Fig. 2), and SK was applied to predict these coordinates. The predicted coordinates at each unsampled location x_0 were applied Eq. (4) to obtain the predicted multinomial vectors $\mathbf{p}^*(x_0)$ approximating the sought distribution of $Z(x_0)$. Fig. 3 compares some predicted probability functions and discrete versions of their true values—obtained with direct SK of $Z(x_0)$—. Note that predicted distributions are in general more uncertain—*less informative*—and the conditional mean of the predictions smoother than the truth.

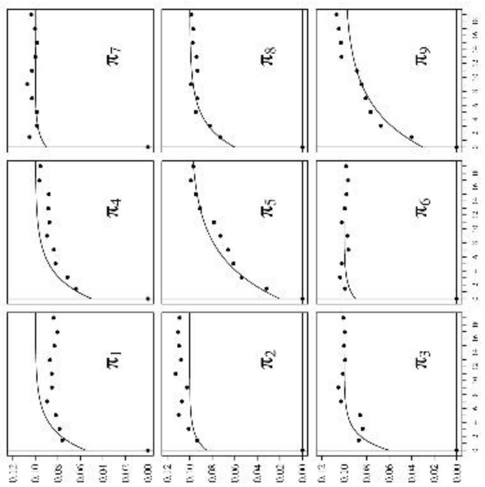


Fig. 2. Experimental variograms of the 9 coordinates, and used models, all of exponential type.

7. Final Considerations

An isometric logistic kriging technique can be applied to estimate discrete probability density functions if a certain degree of uncertainty is accepted when estimating the probability distribution at sampled locations. The obtained predictor minimizes the compositional Aitchison distance between the discrete distribution and its prediction at unsampled locations. This has some connections with information concepts, which has not been explored here. The obtained predictions are *by construction* valid probabilities, positive and summing up to one, thus this technique overcomes the main flaws of IK. Our results—being estimates—are less informative than the true conditional distribution.

8. Acknowledgements

This work was elaborated during a research stage funded by the projects A/04/33586 from the German academic exchange office (DAAD) and 2004-BE-00147 from the Catalan universitary grant agency (AGAUR), and by a Student Grant of the International Association for Mathematical Geology.

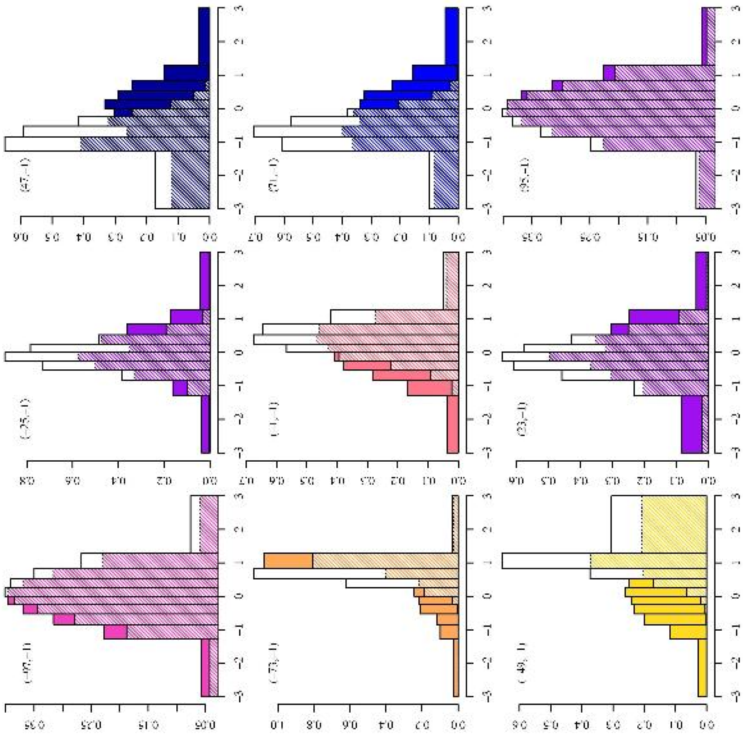


Fig. 3. Estimated histograms (coloured), and equivalent discrete versions of the true probability functions (white), with colours from the truly observed category (legend in Fig. 1). The coordinates of the sampling locations are shown in the left upper corner of each plot.

9. References

Aitchison, J.,1986. *The Statistical Analysis of Compositional Data*, Monographs on Statistics and Applied Probability. Chapman Hall Ltd., London (UK).

Egozcue, J.J., V. Pawlowsky-Glahn, G. Mateu-Figueras, and C. Barceló-Vidal, 2003. Isometric logratio transformations for compositional data analysis. *Math. Geol.*, 35(3), 279-300.

Haas, A. and Ph. Formery ,2002. Uncertainties in facies proportion estimation I. Theoretical framework: the Dirichlet distribution. *Math. Geol.* 34(6), 679-702.

Journel, A. ,1983. Nonparametric estimation of spatial distributions, *Math. Geol.* 15(3), 445-468.

Mateu-Figueras, G., V. Pawlowsky-Glahn, and C. Barceló-Vidal, 2003. Distributions on the simplex. In Thió-Henestrosa and Martín-Fernández (2003).

Matheron, G., 1976. A simple substitute for the conditional expectation: the disjunctive kriging, p. 221-236. In: Guarascio, David, and Huijbregts (eds) *Advanced Geostatistics in the Mining Industry*, vol. 24 of *NATO Advances Study Institute Series; Series C: Mathematical and Physical Sciences*, D. Reidel Publishing Company, Dordrecht (NL).

Myers, D.E. ,1982. Matrix formulation of co-kriging. *Math. Geol.* 14(3) 249-257.

Pawlowsky-Glahn, V., 2003. Statistical modelling on coordinates. In Thió-Henestrosa and Martín-Fernández (2003).

Pawlowsky-Glahn, V. and R.A. Olea, 2004. *Geostatistical Analysis of Compositional Data*, Studies in Mathematical Geology (7). Oxford University Press.

Thió-Henestrosa, S. and J.A. Martín-Fernández, editors, 2003. *Compositional Data Analysis Workshop - CoDaWork'03, Proceedings*. Universitat de Girona.