

Indicator Kriging without Order Relation Violations

Raimon Tolosana-Delgado ·
Vera Pawlowsky-Glahn · Juan-Jose Egozcue

Received: 20 December 2006 / Accepted: 20 August 2007 / Published online: 19 February 2008
© The Author(s) 2008

Abstract Indicator kriging (IK) is a spatial interpolation technique aimed at estimating the conditional cumulative distribution function (ccdf) of a variable at an unsampled location. Obtained results form a discrete approximation to this ccdf, and its corresponding discrete probability density function (cpdf) should be a vector, where each component gives the probability of an occurrence of a class. Therefore, this vector must have positive components summing up to one, like in a composition in the simplex. This suggests a simplicial approach to IK, based on the algebraic-geometric structure of this sample space: simplicial IK actually works with log-odds. Interpolated log-odds can afterwards be easily re-expressed as the desired cpdf or ccdf. An alternative but equivalent approach may also be based on log-likelihoods. Both versions of the method avoid by construction all conventional IK standard drawbacks: estimates are always within the $(0, 1)$ interval and present no order-relation problems (either with kriging or co-kriging). Even the modeling of indicator structural functions is clarified.

Keywords Aitchison geometry · Ilr coordinates · Indicator variogram · Logistic regression

R. Tolosana-Delgado (✉)
Department of Sedimentology and Environmental Geology, University of Göttingen, Göttingen,
Germany
e-mail: raimon.tolosana@geo.uni-goettingen.de

V. Pawlowsky-Glahn
Department of Informatics and Applied Mathematics, University of Girona, Girona, Spain

J.J. Egozcue
Department of Applied Mathematics III, Technical University of Catalonia, Barcelona, Spain

1 Introduction

Indicator kriging (IK) is a geostatistical technique used to approximate the conditional cumulative distribution function (ccdf) at each point of a grid based on the correlation structure of indicator transformed data points (Journel 1983). The major drawback of IK is that it frequently yields inconsistent estimates, such as negative probabilities, probabilities larger than one, or a non-monotonic ccdf. In terms of its associated conditional probability density function (cpdf), such ccdf imply that some events at the estimated location are attached negative probabilities, or that the integral (the sum) of the cpdf is not one. Several methods have been developed to reduce the order relations violation a posteriori. This paper presents an alternative procedure, which by construction avoids such nonsense estimations, and is as straightforward as IK.

Existing strategies include: [a] indicator co-kriging with complementary variables, like probability kriging (Sullivan 1984) and cumulative distribution function of order statistics kriging (Juang et al. 1998); [b] better characterization of indicator cross-covariance structures, like indicator principal component kriging (Suro-Perez and Journel 1991), successive kriging of indicators (Vargas-Guzman and Dimitrakopoulos 2003) or the use of transition probabilities (Carle and Fogg 1996); and [c] the Disjunctive Kriging (Matheron 1976, DK) estimation of the cpdf. None of these methods completely eliminate the order relation problems (Carr and Mao 1993; Carle and Fogg 1996). Even DK, with its high theoretical complexity, usually yields negative estimates in the tails of the cpdf. Therefore, several posterior correction methods of IK results have been implemented, either through trans-Gaussian-type curves (Carr 1994; Bogaert 1999), or via a logistic regression model (Pardo-Igúzquiza and Dowd 2005). From another field of spatial statistics, the Bayesian-Maximum Entropy (BME) alternative to IK (Christakos 1990; Bogaert 2002) is based on assuming a joint multi-dimensional multinomial model for the desired random function, and estimates the parameter (in fact, a multidimensional array of probabilities) by a log-linear model. To the authors' knowledge, BME is the only alternative to IK working with a logistic scale for probabilities (though implicitly assumed), and the only one which by construction does never present order violations. This is achieved at the price of extensive computations and the loss of the intuitive character of kriging techniques.

Attending to the fact that the discrete estimate of the cpdf must have positive components summing up to one, Tolosana-Delgado et al. (2005) identify the sample space of probability vectors, or parameter vectors of a multinomial distribution, with the simplex (the sample space of compositions), and propose to interpolate them using any kriging method devised for data from this constrained space (Pawlowsky 1989; Pawlowsky-Glahn and Olea 2004). The present paper presents this simplicial indicator kriging in depth. From a practical point of view, simplicial IK combines the easy interpretation and implementation of IK with the capability of directly providing valid results. As shown in the scheme of Fig. 1, the method is equivalent to applying conventional linear geostatistics to the coordinates of probability vectors in a given orthonormal basis of the simplex. Such coordinates are, in fact, (generalized) log-odds (Lindley 1964), thus presenting a conceptual link with well-grounded, classical

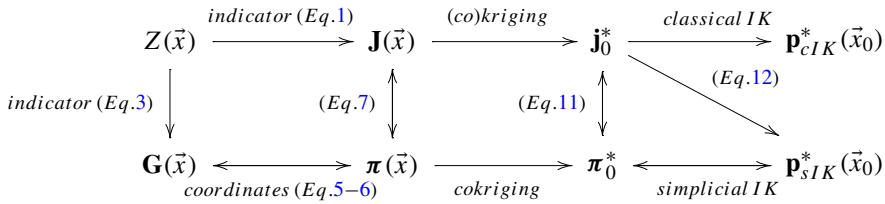


Fig. 1 Flux diagram of indicator kriging techniques and their relationships

tools of statistical analysis of probability density functions (information measures), as well as with some of the cited methodologies (Bogaert 2002; Pardo-Igúzquiza and Dowd 2005).

The present paper presents first this simplicial IK in its most straightforward form. Afterwards, the relationship between classical IK and the simplicial alternative is proven (Fig. 1). This helps both simplifying the computations of the simplicial path (lower one) and clarifying the conditions imposed to indicator covariances in the classical path (upper one). Finally, typical properties of the estimator are reviewed, and an interpretation within the framework of information theory and compositional geometry is put forward. A simulated example illustrates the differences and similitudes between the classical IK and the proposed simplicial alternative. Proofs are collected in the Appendix.

2 Notation and Preliminary Definitions

2.1 Indicator Functions

Let $\vec{x} \in \mathcal{D} \subset \mathbb{R}^p$ be a point in a domain \mathcal{D} of the space-time real space, with $p \in \{1, 2, 3, 4\}$, and $Z(\vec{x})$ a regionalized variable with range $A = [a_0, a_D]$. Let $\{a_1, \dots, a_{D-1}\}$ be a set of thresholds, leading to a partition $A = \bigcup_{i=1}^D A_i$, with $A_i = [a_{i-1}, a_i)$. Then one may either define a set of cumulative indicator, $\mathbf{I}(\vec{x}) = (I_1, \dots, I_D)$, or disjunctive vector transformations, $\mathbf{J}(\vec{x}) = (J_1, \dots, J_D)$, as

$$I_i(\vec{x}) = \begin{cases} 1, & Z(\vec{x}) < a_i, \\ 0, & Z(\vec{x}) \geq a_i, \end{cases} \quad J_i(\vec{x}) = \begin{cases} 1, & Z(\vec{x}) \in A_i, \\ 0, & \text{otherwise,} \end{cases} \quad i \in \{1, 2, \dots, D\}. \quad (1)$$

It is well known that these two vectors satisfy $\mathbf{I} = \mathbf{L} \cdot \mathbf{J}$ and $\mathbf{J} = \mathbf{T} \cdot \mathbf{I}$ with

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{T} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ -1 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & -1 & 1 \end{pmatrix}. \quad (2)$$

Using the same set of thresholds, one can also define a generalized indicator function,

$$G_i(\vec{x}) = \begin{cases} 1 - b, & Z(\vec{x}) \in A_i, \\ \frac{b}{D-1}, & \text{otherwise,} \end{cases} \quad i \in \{1, 2, \dots, D\}, \quad (3)$$

where b is a parameter fixed by the analyst (its role is discussed below).

Let $z(\vec{x}_1), z(\vec{x}_2), \dots, z(\vec{x}_N)$ be an observed sample of $Z(\vec{x})$ at sampled locations $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N$, and \vec{x}_0 an unsampled location. The application of the indicator transforms (1), resp. (3), to the sample yields pseudo-observations $\vec{j}_1, \vec{j}_2, \dots, \vec{j}_N$, resp. $\vec{g}_1, \vec{g}_2, \dots, \vec{g}_N$.

Recall that, when applying IK, one looks for the conditional expectation of $\mathbf{J}(\vec{x}_0)$ given $\vec{j}_1, \vec{j}_2, \dots, \vec{j}_N$, in the hope that this will give valuable information on the conditional distribution of $Z(\vec{x}_0)$ given z_1, z_2, \dots, z_N (Journel 1983). This is feasible because, independently of the original distribution of Z , the transformed \mathbf{J} follows a multinomial distribution of a single trial and probability vector \mathbf{p} . This applies both to the stationary, marginal pdf of the random field $\mathbf{J}(\vec{x})$ ($p_i = \Pr[Z \in A_i]$), and to the local pdf of $\mathbf{J}(\vec{x}_0)$ conditional on the available data ($p_i = \Pr[Z \in A_i | z_1, \dots, z_N]$).

2.2 Compositional Data Analysis Concepts

Independent of any spatial issues, the probability parameter vector of a multinomial distribution $\mathbf{p} = (p_1, p_2, \dots, p_D)$ has positive components ($p_i \geq 0$) summing up to one ($p_1 + p_2 + \dots + p_D = 1$). This is the standard definition of a D -part compositional datum, and its sample space is called the D -part simplex, denoted by \mathbb{S}^D (Aitchison 1986). If the components $(p'_1, p'_2, \dots, p'_D)$ do not fulfill the sum constraint, the closure operation $\mathcal{C}(\cdot)$ may be applied

$$\mathbf{p} = \mathcal{C}(\mathbf{p}') =: \frac{1}{p'_1 + \dots + p'_D} (p'_1, p'_2, \dots, p'_D).$$

Both vectors \mathbf{p}' and \mathbf{p} give the relative likelihood of occurrence of each class A_1, \dots, A_D . The closure only changes its total sum, but not the odds of any two categories. This points to the fact that the relevant information for probability vectors is a relative one. This is the reason why they have often been analyzed on a logarithmic-like scale, with log-linear models, logistic regression, or Shannon information measure.

The simplex of D parts (\mathbb{S}^D) admits an Euclidean space structure (Billheimer et al. 2001; Pawlowsky-Glahn and Egozcue 2001) with the following operations. The Abelian group operation, called perturbation and equivalent to Bayes updating, is given by $\mathbf{p} \oplus \mathbf{q} = \mathcal{C}(p_1 q_1, \dots, p_D q_D)^t$. The scalar multiplication, or powering, is given by $\alpha \odot \mathbf{p} = \mathcal{C}(p_1^\alpha, \dots, p_D^\alpha)^t$. The inner product and associated distance are given by

$$\langle \mathbf{p}, \mathbf{q} \rangle_a = \frac{1}{D} \sum_{i < j} \ln \frac{p_i}{p_j} \ln \frac{q_i}{q_j}, \quad d_a^2(\mathbf{p}, \mathbf{q}) = \frac{1}{D} \sum_{i < j} \left(\ln \frac{p_i}{p_j} - \ln \frac{q_i}{q_j} \right)^2, \quad (4)$$

where $\mathbf{p}, \mathbf{q} \in \mathbb{S}^D$, $\alpha \in \mathbb{R}$, and the superindex t marks transposition. In this structure, known as Aitchison geometry, orthonormal coordinates may be computed with

$$\boldsymbol{\pi} = \boldsymbol{\Psi} \cdot \ln \mathbf{p}, \quad (5)$$

where Ψ is a $(D - 1) \times D$ matrix with rows forming a set of orthonormal vectors (as stated in Property 1 in the Appendix). The vector \mathbf{p} can be recovered by applying the coordinates to the basis, giving

$$\mathbf{p} = \mathcal{C}(\exp(\Psi^t \cdot \boldsymbol{\pi})). \quad (6)$$

It is useful to notice here that each and every possible coordinate may be expressed as a (generalized) log-odd (Lindley 1964), also known as log-contrast (Aitchison 1984), like $\pi_i = \ln((\prod_u p_{iu}^{\psi_{iu}})/(\prod_d p_{id}^{\psi_{id}}))$, where the sum of the powers affecting parts in the numerator $\sum_u \psi_{iu}$ must equal the sum of those in the denominator $\sum_d \psi_{id}$. The Aitchison geometry of the simplex captures all the relative information of probability vectors.

According to Pawlowsky-Glahn (2003), whenever a meaningful Euclidean geometry is available, one should represent observations in any orthonormal basis according to this geometry and apply any desired statistical method to the coordinates in that basis. Results describing objects of the original Euclidean space may then be applied to the basis used to obtain final estimates. Geostatistics is not an exception (Tolosana-Delgado 2006; Tolosana-Delgado and Pawlowsky-Glahn 2007).

3 Simplicial Indicator Kriging (I): Interpolating Log-Odds

Assume first that the studied random field is just $\mathbf{p}(\vec{x})$. One has thus observations of the vectors of probabilities $\mathbf{p}(\vec{x}_n) = \mathbf{p}_n$ at several locations \vec{x}_n , $n = 1, 2, \dots, N$. In this case, given that \mathbf{p}_n are space-dependent compositions, they might be interpolated with any method devised for this kind of constrained data (Pawlowsky 1989; Pawlowsky-Glahn and Olea 2004). However, such probability vectors are actually not observed. Instead, one estimates them using observations of $Z(\vec{x})$, a regionalized variable. Classical IK estimates the \mathbf{p}_n by \mathbf{j}_n (see (1)). But the presence of zeroes in this vector precludes the computation of log-odds, and thus the application of the Aitchison geometry.

Whenever the analyst admits a certain degree of uncertainty, probability vectors \mathbf{p}_n may be estimated by \mathbf{g}_n (see (3)), depending on a parameter b . This parameter accounts for the (subjective) uncertainty regarding the true value of $Z(\vec{x}_n)$: observing $z(\vec{x}_n) \in A_i$ means that $\Pr[Z(\vec{x}_n) \in A_i] = 1 - b$. By using the generalized indicator transform, one obtains a probability vector with no zeroes, which fully respects the log-odd geometry. Thus, one may compute for every location the vector of coordinates $\boldsymbol{\pi}(\vec{x}_0)$ (5) and treat the obtained coordinates with conventional geostatistics. Finally, the kriging results, denoted $\boldsymbol{\pi}_K$, might be applied to the basis in use, represented by the matrix Ψ (6), to recover the true kriged probability vector, \mathbf{p}_K . Three objections might appear at this point:

Basis dependence. Do results obtained with this methodology depend on the choice of the orthonormal basis? (i.e. of the matrix Ψ). The next sections show that the whole co-kriging procedure is independent of it.

The uncertainty parameter. Regarding the value of b , one can show that it does not influence the covariance estimation and co-kriging procedures. It is nevertheless an

important parameter in the obtention of the final prediction for $\mathbf{p}(\vec{x})$: for the moment, consider it a small value, say 0.05 or 0.1, depending on the actual number of classes involved and the degree of uncertainty admitted for $Z(\vec{x})$ (both factors favor a higher value of b). It is not reasonable to take $b > (D - 1)/D$, as this would imply that the observed class is less probable than the non-observed ones. A thorough discussion of the role of b is given afterwards.

Adequacy of probability estimates through indicator functions. The third issue relates to the adequacy of the generalized indicator function (3) as an estimator at the sampled locations \vec{x}_n for the true discrete probability density function, $\mathbf{p}(\vec{x}_n)$. Note that $\mathbf{G}(\vec{x})$ (3) has only two outcomes ($b/(D - 1)$ or $1 - b$). In fact, this estimator does not take into account proximity effects between classes. For example, if $Z(\vec{x})$ falls in the fifth class, then the fourth or the six will not receive more probability than the first class. This is strictly valid only if each class can be equally confounded with all the other classes, for instance, when classes are a priori equiprobable and they have no ordering. The same lack of ordering information is present in the disjunctive indicators \mathbf{J} (1), as its underlying model is multinomial, but it also affects the classical cumulative indicators \mathbf{I} (as proven afterwards). One might find several alternatives, like (a) putting some arbitrary values, decreasing as classes become far apart, (b) using a suitable probability model with a small dispersion to compute the probability of each class, or (c) estimating each probability of conditional classification $\Pr[z(\vec{x}_n) \in A_j | z(\vec{x}_n) \in A_i]$ with the “nugget effect” of transition probabilities (Carle and Fogg 1996). These options are left for further research. The rest of the paper deals with generalized indicators, to keep the parallelism with classical indicator kriging.

4 Details on Geostatistics for Probability Vectors

4.1 Data Representation

Several relationships can be established between the cumulative indicators \mathbf{I} , the disjunctive ones \mathbf{J} , and the coordinates π_n (5) of the generalized indicator (3). In addition to the well-known relation $\mathbf{I} = \mathbf{L} \cdot \mathbf{J}$ (2), it is easy to prove (Appendix, Proposition 1) that

$$\boldsymbol{\pi} = \beta \cdot \boldsymbol{\Psi} \cdot \mathbf{J} \quad \text{and} \quad \mathbf{J} = \frac{1}{\beta} \cdot \boldsymbol{\Psi}^t \cdot \boldsymbol{\pi} + \frac{1}{D} \cdot \mathbf{1}, \quad \text{where } \beta = \ln \frac{(1-b)(D-1)}{b}. \quad (7)$$

Therefore, any of the vectors \mathbf{I} , \mathbf{J} , \mathbf{G} , and $\boldsymbol{\pi}$ are linked through linear, full-rank transformations.

4.2 Structural Analysis

Using (7) and Proposition 2 (see Appendix), both the mean and covariance structure of the coordinates $\boldsymbol{\pi}(\vec{x})$ are linearly related to the mean and covariance structure of

disjunctive indicators $\mathbf{J}(\vec{x})$

$$\bar{\pi} = \beta \cdot \Psi \cdot \bar{\mathbf{J}} \quad \text{and} \quad \bar{\mathbf{J}} = \frac{1}{\beta} \cdot \Psi^t \cdot \bar{\pi} + \frac{1}{D} \cdot \mathbf{1}, \tag{8}$$

$$\mathbf{C}^\pi(\vec{h}) = \beta^2 \cdot \Psi \cdot \mathbf{C}^J(\vec{h}) \cdot \Psi^t \quad \text{and} \quad \mathbf{C}^J(\vec{h}) = \frac{1}{\beta^2} \cdot \Psi^t \cdot \mathbf{C}^\pi(\vec{h}) \cdot \Psi. \tag{9}$$

In these expressions, $\bar{\mathbf{J}}$, $\mathbf{C}^J(\cdot)$, $\bar{\pi}$, and $\mathbf{C}^\pi(\cdot)$ denote, respectively, means and covariances of \mathbf{J} and π . The same can be said about the relationship cumulative-disjunctive indicators, giving

$$\bar{\mathbf{I}} = \mathbf{L} \cdot \bar{\mathbf{J}} \quad \text{and} \quad \mathbf{C}^I(\vec{h}) = \mathbf{L} \cdot \mathbf{C}^J(\vec{h}) \cdot \mathbf{L}^t. \tag{10}$$

All these equations are naturally satisfied by the experimental means and covariances, at all lag distances. If the fitted models also fulfill them, then they may be called compatible. These relationships offer a way to clarify the modeling of indicator structural functions. Recall that the disjunctive indicator covariance $\mathbf{C}^J(\vec{h})$ is bound to sum up to zero by rows and by columns for each \vec{h} , due to the constant sum constraint of \mathbf{J} . Also, $\mathbf{C}^J(\vec{0}) = \text{diag}(\bar{\mathbf{J}}) - \bar{\mathbf{J}} \cdot \bar{\mathbf{J}}^t$, because it must correspond to a valid covariance matrix of a multinomial vector (Bogaert 2002). According to this author, these extra conditions, together with the classical conditions on positive-definiteness of the covariance structure, are enough to ensure a valid model for $\mathbf{C}^J(\vec{h})$. In contrast, conditions for a valid covariance structure $\mathbf{C}^I(\vec{h})$ of cumulative indicators are not known, apart from a conjecture by Matheron (1993).

On the contrary, the covariance structure $\mathbf{C}^\pi(\vec{h})$ is almost free from these limitations. Any model fitted to the covariance structure of coordinates will be consistent with a covariance structure $\mathbf{C}^J(\vec{h})$ summing up to zero both by rows and by columns, thanks to the pre- and post-multiplication by Ψ in (9). Provided that the model fitted to $\mathbf{C}^\pi(\vec{h})$ is positive definite, one has only to tune the value of β and the nugget effects, in order to fit the variance of a multinomial variable for $\vec{h} = \vec{0}$. Even when working within a standard IK framework, the covariance system used may be easily ensured to be a valid one by computing the experimental covariance of the coordinate function, by fitting a valid model to the resulting $(D - 1) \times (D - 1)$ covariance system, and by computing the compatible model for $\mathbf{J}(\vec{x})$ through (9), or even for $\mathbf{I}(\vec{x})$ through (10). These derived models for $\mathbf{C}^I(\vec{h})$ or $\mathbf{C}^J(\vec{h})$ are singular, if one does not erase a component of the indicators (the last for $\mathbf{I}(\vec{x})$, which is constant and equal to 1, or any one for $\mathbf{J}(\vec{x})$, since each can be recovered as one minus the sum of the others). Another possible, though sub-optimal, option is to work with all disjunctive indicators, but ignore the cross-covariances (effectively applying kriging instead of co-kriging).

4.3 Simple Co-kriging

The data set of π and \mathbf{J} (Appendix, Proposition 1), as well as their expectations (8) and covariances (9), keep a compatible linear relationship between them; this is actually an invertible relation. Furthermore, this relation is also fulfilled by the

respective co-kriging predictors. If \mathbf{j}_0^* and $\boldsymbol{\pi}_0^*$ represent these predictors obtained for an unsampled location \bar{x}_0 , then

$$\boldsymbol{\pi}_0^* = \beta \cdot \boldsymbol{\Psi} \cdot \mathbf{j}_0^* \quad \text{and} \quad \mathbf{j}_0^* = \frac{1}{\beta} \cdot \boldsymbol{\Psi}' \cdot \boldsymbol{\pi}_0^* + \frac{1}{D} \cdot \mathbf{1}. \quad (11)$$

A proof is provided in the [Appendix](#) (Proposition 3), although these expressions are logically implied by the linearity of the co-kriging predictor. Obviously, the same can be said about the disjunctive and the cumulative indicator predictions, thanks to their relation through (2) and (10). One gets exactly the same result by estimating $\mathbf{J}(\bar{x}_0)$ and accumulating the results in the adequate order, or by kriging $\mathbf{I}(\bar{x}_0)$: where \mathbf{i}_0^* presents an order inversion, \mathbf{j}_0^* has a negative component, and vice versa. This implies that a full classical co-kriging does not take any advantage of the ordering of the classes, as \mathbf{J} does not convey this information.

5 Simplicial Indicator Kriging (II): Interpolating Log-Likelihoods

From the interpolated values $\boldsymbol{\pi}_0^*$, \mathbf{i}_0^* , or \mathbf{j}_0^* , one must obtain an estimate for $\mathbf{p}(\bar{x}_0)$. The classical IK takes this prediction as $\mathbf{p}_{cIK}^*(\bar{x}_0) = \mathbf{j}_0^*$, after applying the necessary corrections. For instance, in the widespread GSLIB (Deutsch and Journel 1998), routine `ordrel` corrects predictions \mathbf{i}_0^* in a way equivalent to forcing each component of \mathbf{j}_0^* to satisfy $0 \leq j_k^* \leq 1$ and modifying the last components to ensure that $1 = \sum_{k=1}^D j_k^*$.

The simplicial alternative presented here applies the predicted coordinates $\boldsymbol{\pi}_0^*$ to the basis (6), thus yielding a compositional vector. No further correction is needed, as it naturally satisfies the conditions of the sample space (positive component summing up to one). The estimated cdf is always ordered and bounded between 0 and 1. According to Fig. 1, this can be further simplified ([Appendix](#), Proposition 4) to

$$\mathbf{p}_{sIK}^*(\bar{x}_0) = \mathcal{C}(\exp(\beta \cdot \mathbf{j}_0^*)), \quad (12)$$

where $\beta = \ln((1 - b)(D - 1)/b)$. Recall that b is fixed as a probability of error in the determination of the categories (3).

The practical implications of this simplification are very important. First, old results obtained with the classical IK can be recasted straightforwardly to simplicial IK estimates, which by definition will not present order violations. Second, existing software routines for IK estimation can be reused to obtain simplicial IK programs with small changes. Third, if modeling the $\frac{(D-1)(D-2)}{2}$ auto- and cross-covariances needed for a full simplicial indicator cokriging is not affordable (as is usually the case, due to the excessive fluctuations of empirical variograms and covariance functions), one might consider kriging instead of co-kriging. Independently interpolating the disjunctive indicators $\mathbf{j}(\bar{x})$ and plugging them into (12), one obtains a valid estimate of a probability vector, though in the process some information will be lost (as kriging is suboptimal against cokriging).

5.1 Properties

The most important property of simplicial IK estimator in any of its two forms, through coordinates or with (12), is its consistency. It always results in a positive set of D values which sum up to one, automatically satisfying the conditions of a valid cpdf, without the need of any further correction procedure. But the final form of (12) implies two interesting stability properties.

- The choice of a matrix Ψ (thus, of a basis of the simplex) does not affect the final prediction obtained for $\mathbf{p}(\vec{x})$, as it simply does not appear in (12). However, this is only strictly true if a full covariance system (with auto- and cross- covariances) has been used in the co-kriging procedure. If kriging was applied to each coordinate instead of co-kriging, some differences could be observable. For this reason, if cokriging $\pi(\vec{x})$ is not applicable, one might prefer to interpolate $\mathbf{J}(\vec{x})$ and use (12). Since each component J_k of $\mathbf{J}(\vec{x})$ is taken as non-dependant on the other, their interpolation will be interpreted as the log-likelihood of $Z(x_0) \in A_k$ (see discussion below).
- The value β , thus the value b from (3), plays no role in the kriging procedure (neither for interpolations of \mathbf{j}_0^* nor of π_0^*) and simply “scales up” the final result.

But simplicial IK satisfies other standard properties of estimators inherited from standard kriging techniques. The fact that co-kriging estimators of $\pi(\vec{x}_0)$ are BLUE in real space (the space of coordinates) guarantees that the estimators expressed as compositions $\mathbf{p}_{sIK}^*(\vec{x}_0)$ are BLUE in the simplex with respect to the Aitchison geometry (Pawlowsky-Glahn and Egozcue 2002). In particular, the Aitchison distance (4) between the true but unknown discrete probability function $\mathbf{p}(\vec{x}_0)$ and its estimate is minimized by the simplicial IK.

Finally, assuming the random function $\pi(x)$ has a jointly normal distribution, its simple kriging prediction, together with the kriging variance-covariance matrix, gives the parameters of the true distribution of $\pi(x_0)$ conditional on the observed data set. This implies that the conditional distribution of $\mathbf{p}(x_0)$ is a Normal on the simplex (Mateu-Figueras et al. 2003). This result opens the door to a possible hierarchical approach to IK, which is left for further investigation.

6 Discussion: The Role of β

The discussion on the role of β in (12) requires some additional concepts from the Aitchison geometry of the simplex (Egozcue and Pawlowsky-Glahn 2005, 2006). Some quick calculations show that the Aitchison norm, $\|\mathbf{x}\|_a = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle_a}$, of the indicator transformed vectors at a given location is $\|\mathbf{J}\|_a = \infty$ and $\|\mathbf{G}\|_a = \beta/\sqrt{2}$, regardless of the value of the regionalized variable Z . Actually, the following three approximations occur simultaneously: ($\mathbf{G} \rightarrow \mathbf{J}$), ($\beta \rightarrow +\infty$), and ($b \rightarrow 0$). A probability vector is more informative (the uncertainty on the value of Z is lower) as a function of this norm, i.e. of the value β . Now, if one takes (12) and applies a logarithm to both sides, one gets the idea that $\ln \mathbf{p}_0^* = \beta \cdot \mathbf{j}_0^*$ plus a constant. This constant comes from the closure and is irrelevant if the left side of the equality is interpreted

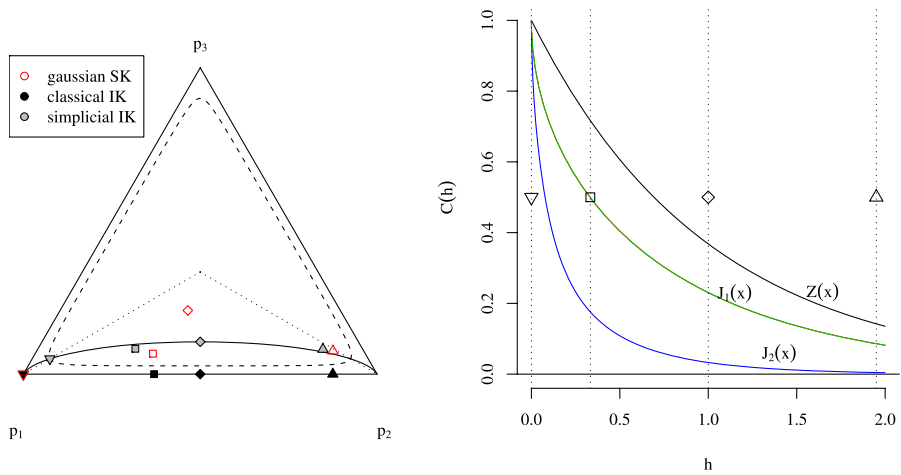
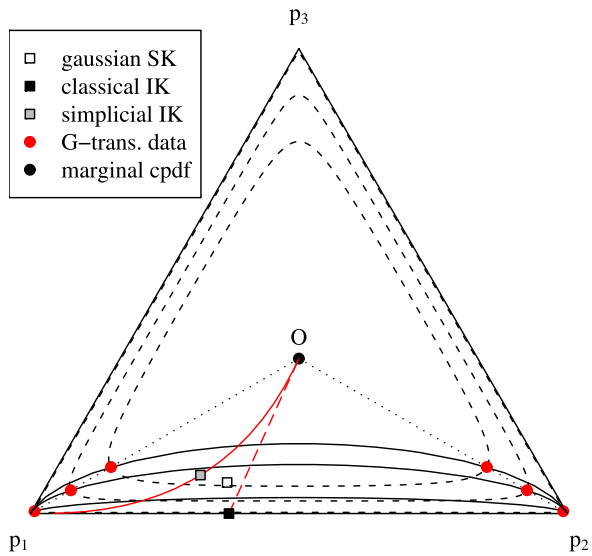


Fig. 2 (Left) Predictions obtained with simple kriging of the underlying Gaussian field $Z(x)$ (empty symbols), with disjunctive indicator kriging (black symbols), and with simplicial indicator kriging (gray symbols), obtained with $b = 0.1$, at four different locations: $x_0 = 0$ (triangle pointing down), $x_0 = 0.33$ (square), $x_0 = 1$ (diamond), and $x_0 = 1.95$ (triangle pointing up). The two observed data were placed at $x_1 = 0$ and $x_2 = 2$. (Right) Covariance functions of $Z(x)$, of the J_1 and J_3 disjunctive indicators, and of the J_2 disjunctive indicator (from top to bottom; note that $C_1^J = C_3^J$); symbols represent the location of the predicted distributions of the left plot

as the vector of log-likelihoods of each class. Leaving β aside, simplicial IK may be seen as a (co)-kriging technique on a normalized log-likelihood vector, which is equal to \mathbf{J} . The obtained results are scaled by β independently of the geostatistical procedure, giving to the resulting probability vector its informative power. The value of β acts thus as “units”. Fixing its value, the analyst has the opportunity to introduce his own assessment (necessarily subjective) on the magnitude of the uncertainty affecting $Z(\vec{x})$. It could even be “tuned” differently at each interpolated location, although this option is left for further research.

A small synthetic example may serve as illustration (Fig. 2). Assume $Z(x)$ is a Gaussian random field, with zero mean and covariance $C(h) = \exp(-|h|)$ (with an effective range around 3 units). The range of this variable is split into three equally-probable categories by the intervals $A_1 = (-\infty, -a)$, $A_2 = [-a, a)$, and $A_3 = [a, +\infty)$, with $a = 0.43072$. The central dot of Fig. 2, left, represents the corresponding non-conditional, marginal discretized pdf $\mathbf{p}^0 = (1, 1, 1)/3$. Given that $Z(x)$ is Gaussian, one can compute the corresponding covariances for the cumulative indicators \mathbf{I} (Suro-Perez and Journel 1991). Due to the invertibility of (2), these can be used to derive covariances for the disjunctive indicators \mathbf{J} (Fig. 2, right). Two observations are available: $z(x = 0) = 0.85$ and $z(x = 2) = 0.00$ giving, respectively, $\mathbf{j}_1 = (1, 0, 0)$ and $\mathbf{j}_2 = (0, 1, 0)$. Kriging $Z(x)$ and each component of $\mathbf{J}(x)$ with their corresponding covariances, one obtains three kinds of estimation of the conditional pdf: a simple kriging (SK) prediction of $Z(x_0)$ (giving the true conditional distribution), the classical and the simplicial IK predictions. These predictions are compared in Fig. 2 (left): IK in its classical form gives predictions on a side of the ternary plot (working with only two samples). Similarly, simplicial IK predictions also fall on

Fig. 3 Discrete probability densities for the 3-category synthetic example (see text for details). Legend: cpdf obtained with SK of the underlying Gaussian field (*empty square*), with classical IK (\mathbf{j}_0^* , *black square*), or with simplicial IK (\mathbf{p}_{sIK}^* , *gray square*); available data transformed through \mathbf{G} (circles near p_1 and p_2 vertices, for values of $b = 0.01, 0.1,$ and 0.2); non-conditional pdf (central circle, marked with O)



a trend (only exactly with cokriging), corresponding to an Aitchison compositional line (Pawlowsky-Glahn and Egozcue 2002): i.e., the set of all mixtures of the available “data” $\mathbf{g}_1 = \mathcal{C}(\exp(\beta \mathbf{j}_1))$ and $\mathbf{g}_2 = \mathcal{C}(\exp(\beta \mathbf{j}_2))$, with $\beta = \ln\left(\frac{(D-1)(1-b)}{b}\right)$. In this case, a value of $b = 0.1$ was taken. Finally, the Aitchison circle of radius $r = \beta/\sqrt{2}$ is also portrayed (dashed triangle with smooth vertices).

Focusing on the prediction for $x_0 = 0.33$, Fig. 3 shows the effect of varying b . The dots close to the vertices of the ternary diagram represent the resulting \mathbf{g}_1 and \mathbf{g}_2 for three different values of b (or β): 0.2 (2.08, innermost), 0.1 (2.89, intermediate), and 0.01 (5.29, outermost and visually almost coincident with the borders of the triangle). As references, the figure includes the corresponding Aitchison circles (dashed curves: all points on one of these circles give a pdf with the same informative power), as well as the compositional lines between each pair of \mathbf{g}_n (solid lines: all points on one of these trends correspond to possible simplicial indicator cokriging predictions for a fixed value of b). Finally, the compositional line from the center of the plot (point O) to p_1 gives the set of cpdfs predicted with simplicial IK at $x_0 = 0.33$, for all possible values of b between 0 (on the vertex) and $\frac{D-1}{D}$ (on the center O), and the dashed line is its tangent at the center. This tangent is, in fact, a vector representation of the predicted \mathbf{j}_0^* .

7 Illustration: Simulated Example

In this section, a more complex simulation exercise was conducted, to compare performance of IK in the simplicial and in the classical approaches. A single realization was drawn (with LU decomposition) from a zero-mean, unit-variance Gaussian random function $Z(\vec{x})$, with an exponential covariance model of effective range 9 units (u). The simulation contained 19 “observations” $\{Z(\vec{x}_n), n = 1, 2, \dots, 19\}$

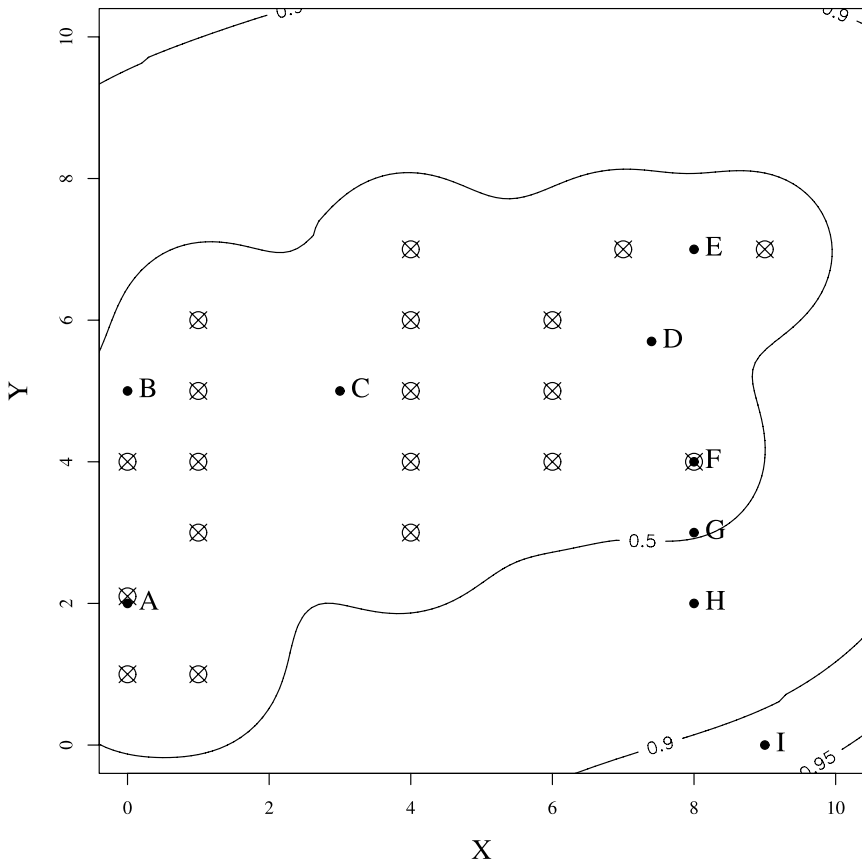


Fig. 4 Data (marked with crossed circles) and locations to be predicted (marked with dots and alphabetic characters). The contour lines mark the levels of 0.5, 0.90, and 0.95 kriging variance obtained with simple kriging of $Z(x)$, the original Gaussian random function

within a square of $10 \times 10 u^2$ (Fig. 4). Figure 4 shows also the 9 locations where the probability distribution will be estimated, portraying several scenarios of spatial relationship with the observations. Again, the non-conditional distribution of the random function at any location is known: $Z(\vec{x}_n) \sim \mathcal{N}(0, 1)$. Its range may be split into 10 a priori equally-probable classes (Table 1). Applying the disjunctive indicator transform (1), the simulated values of $Z(\vec{x}_n)$ are transformed into “observed” vectors \mathbf{j}_n (of 10 components), for each $n = 1, \dots, 19$.

At each of the 9 unsampled data locations, the true cpdf is computed applying simple kriging to the simulated data set, using its true exponential covariance. Figure 4 presents the area where interpolation is meaningful (for example, with a kriging variance lower than 90% or 95% of the sill of the exponential variogram). This true cpdf of $Z(\vec{x}_0)$ (a Gaussian) is discretized in the same classes of Table 1. This is done because IK techniques estimate this discrete version of the true cpdf, therefore it is better suited for comparison purposes. Given that the joint distribution of $Z(\vec{x})$ is Gaussian and that its true covariance is known, the true covariance structure of its

Table 1 Definition of classes by intervals and central values (medians)

Class	Interval		\bar{z}_i
A ₁	−∞	−1.282	−1.645
A ₂	−1.282	−0.842	−1.036
A ₃	−0.842	−0.524	−0.674
A ₄	−0.524	−0.253	−0.385
A ₅	−0.253	0.000	−0.126
A ₆	0.000	0.253	0.126
A ₇	0.253	0.524	0.385
A ₈	0.524	0.842	0.674
A ₉	0.842	1.282	1.036
A ₁₀	1.282	+∞	1.645

indicator transformations may be computed, using a relation between the distribution of bi-normal variables and that of their cumulative indicator transformations given by Suro-Perez and Journel (1991). The compatible autocovariances for the disjunctive indicators may be obtained inverting (2) and are plotted in Fig. 5. At each of the 9 unsampled locations, simple kriging is applied to estimate each component of $\mathbf{J}(\vec{x}_0)$. Only auto-covariances are used (Fig. 5). Kriging (and not co-kriging) is applied, to resemble what is really done in standard IK practice. From the kriged values \mathbf{j}_0^* (Table 2), the two IK estimations are computed.

Classical IK directly takes the kriged values $\mathbf{p}_{cIK}^*(\vec{x}_0) = \mathbf{j}_0^*$ as a discrete estimate of the cpdf of $Z(\vec{x}_0)$, after correcting them to avoid negative probabilities and forcing their sum to 1. The correction proposed by Deutsch and Journel (1998) in Table 2 affects preferentially the upper tail classes (as it is applied on the ccdf), leaving one wondering why the lower tail estimates are more reliable than the upper tail ones.

Simplicial IK uses the transformation (12) to get $\mathbf{p}_{sIK}^*(\vec{x}_0)$ as an estimate of the cpdf of $Z(\vec{x}_0)$. If one considers that observing $z(\vec{x}_n) \in A_i$ means that truly $Z(\vec{x}_n) \in A_i$ with a 90% probability, then $b = 0.1$ in (3), and thus a value of $\beta = 4.39$ was taken in (12). The proposed estimator actually interprets the result of $\mathbf{j}^*(\vec{x}_0)$ as the non-normalized log-likelihoods of each category, and (12) simply transforms them to closed probabilities.

A comparison of the true distribution with its classical and simplicial IK estimations, as cumulative functions (Fig. 6), reveals the following issues:

- As expected, both IK predictors (as well as the cpdf estimated with results from simple kriging of Z) tend to the non-conditional marginal distribution (as seen with the series of locations F to I) when estimating the cpdf at a location far away from the data.
- Both IK predictors tend to give more disperse distributions than the true ones. This is directly related to the underlying model (multinomial in IK, normal in SK) or the non-parametric character of both IK techniques. Good examples are locations A and C.
- Simplicial IK predictions \mathbf{p}_{sIK}^* cannot have a zero in any component, and in the lower tail they tend to be greater than \mathbf{p}_{cIK}^* (locations A and D). The extreme case is location F, which coincides with a datum. Given that simplicial IK is also

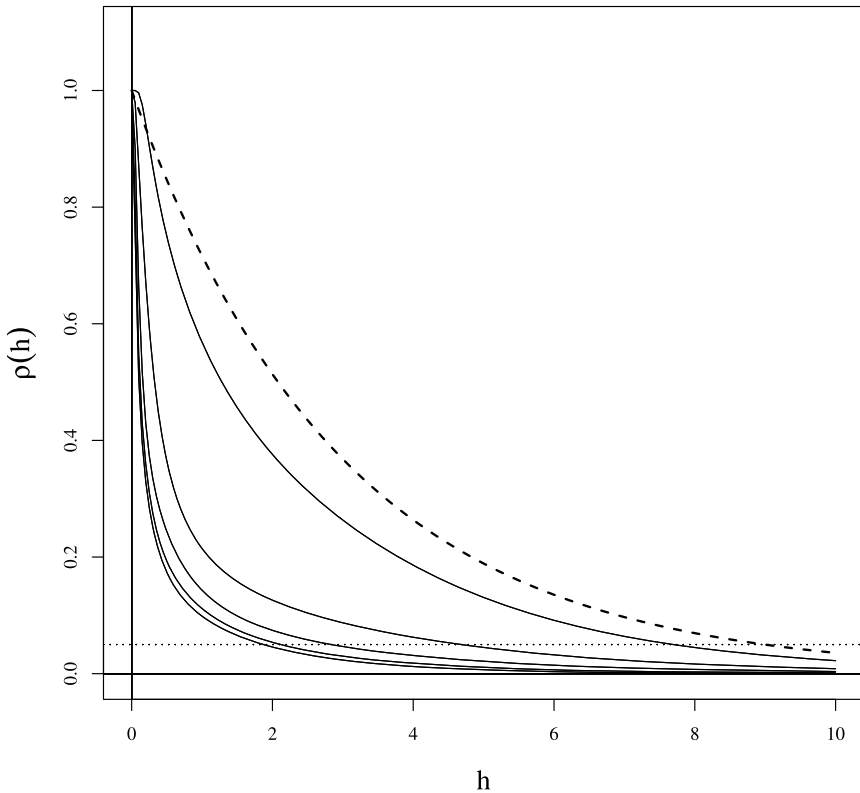


Fig. 5 Auto-correlograms of the Gaussian random function ($Z(x)$, dashed line) and of the components of \mathbf{J} (solid lines): curves (from top to bottom) correspond to the correlograms of j_1 (or j_{10}), j_2 (or j_9), j_3 (or j_8), j_4 (or j_7), and j_5 (or j_6). The dotted line marks the 0.05 correlation level, showing the effective range of each correlogram. Note that the classes near the median are much less continuous than the tail classes

an exact interpolator, the final estimation of kriging is exactly the result of (3): $(1 - b)$ in the observed category, and b evenly distributed among the others.

- Classical IK, on the contrary, gives higher probability to central classes (locations B, D, and G) than simplicial IK.
- Finally, upper tails of \mathbf{p}_{cIK}^* are trimmed in those cases where the total sum of \mathbf{J}_0^* is far from one (Table 2). On the other side, simplicial IK offers always a reliable estimation of these tails, as seen in Fig. 6. It is interesting to note that at locations G and H, the corrections of classical IK only affect the upper tail, while the main source of distortion from the true distribution is the lower tail.

Summarizing, simplicial IK, while keeping a performance globally comparable to that of classical IK, never gives order relation problems. This advantage is due to its very definition as a cokriging of log-odds, and not because of posterior corrections. Its optimality properties are not obscured by the correction process.

Table 2 IK predictions for disjunctive indicators, as obtained by kriging (upper part) and after correction (below). Corrections are marked with an asterisk

Uncorrected	A	B	C	D	E	F	G	H	I
j_1	0.03	0.29	0.01	0.02	0.02	0.00	0.04	0.06	0.08
j_2	0.03	0.05	0.04	0.05	0.06	0.00	0.07	0.08	0.09
j_3	0.05	0.13	0.09	0.07	0.08	0.00	0.08	0.09	0.09
j_4	0.07	0.11	0.09	0.08	0.08	0.00	0.09	0.09	0.10
j_5	0.06	0.11	0.09	0.09	0.09	0.00	0.09	0.10	0.10
j_6	0.06	0.08	0.08	0.08	0.09	0.00	0.09	0.10	0.10
j_7	0.06	0.09	0.17	0.13	0.11	0.00	0.10	0.10	0.10
j_8	0.50	0.12	0.16	0.15	0.17	0.00	0.11	0.11	0.11
j_9	0.07	0.09	0.10	0.07	0.07	0.00	0.09	0.10	0.10
j_{10}	0.00	0.02	0.02	0.49	0.43	1.00	0.58	0.41	0.22
Sum	0.94	1.09	0.85	1.22	1.18	1.00	1.33	1.22	1.09
Corrected	A	B	C	D	E	F	G	H	I
j_1	0.03	0.29	0.01	0.02	0.02	0.00	0.04	0.06	0.08
j_2	0.03	0.05	0.04	0.05	0.06	0.00	0.07	0.08	0.09
j_3	0.05	0.13	0.09	0.07	0.08	0.00	0.08	0.09	0.09
j_4	0.07	0.11	0.09	0.08	0.08	0.00	0.09	0.09	0.10
j_5	0.06	0.11	0.09	0.09	0.09	0.00	0.09	0.10	0.10
j_6	0.06	0.08	0.08	0.08	0.09	0.00	0.09	0.10	0.10
j_7	0.06	0.09	0.17	0.13	0.11	0.00	0.10	0.10	0.10
j_8	0.50	0.12	0.16	0.15	0.17	0.00	0.11	0.11	0.11
j_9	0.07	0.02*	0.10	0.07	0.07	0.00	0.09	0.10	0.10
j_{10}	0.07*	0.00*	0.17*	0.26*	0.23*	1.00	0.24*	0.17*	0.13*
Sum	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

8 Conclusions

A finite, discrete (approximation of a) probability density function is a composition, i.e. a vector with D positive components summing up to one. Probability vectors and compositions have as sample space the D -part simplex (\mathbb{S}^D), and this space can be given an Euclidean space structure compatible with a log-ratio scale.

These considerations are taken into account when estimating probabilities, by applying conventional linear geostatistical techniques to the coordinates of the indicator-transformed data in a chosen orthonormal basis (computed as a set of generalized log-odds). Interpolated results can be easily re-expressed as probability vectors by back-application to the basis. This technique is called simplicial indicator kriging, and its results are always valid multinomial probability vectors. Using common concepts from linear algebra, a clean solution to order relation problems of indicator kriging is obtained.

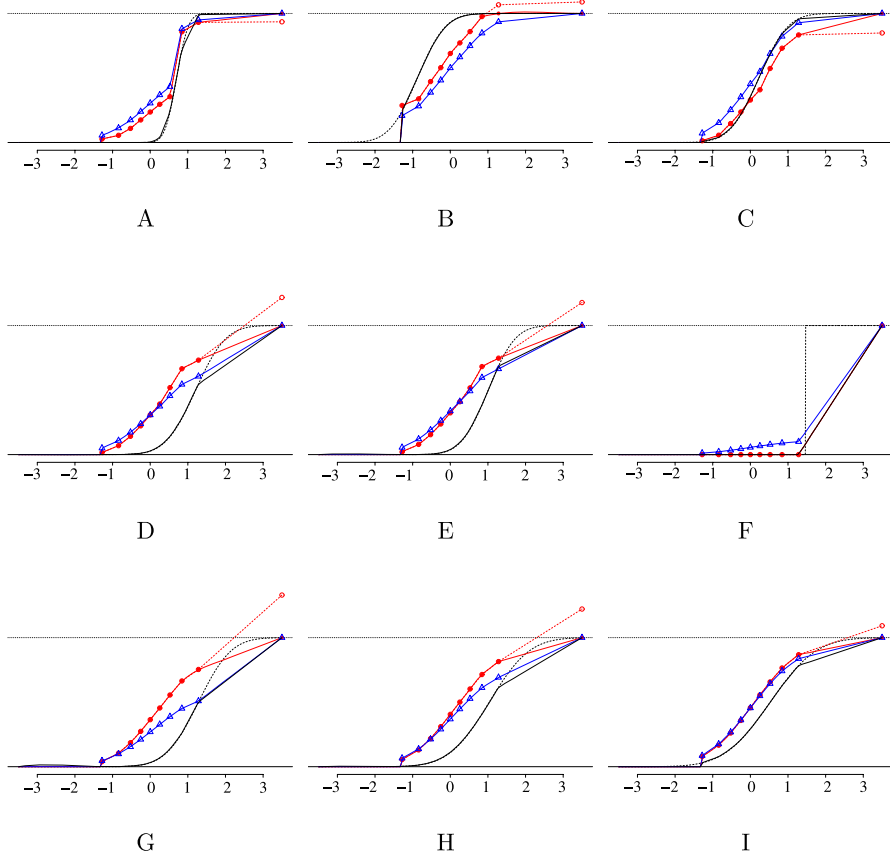


Fig. 6 Final cdf predictions obtained with the classical IK (filled circles) and simplicial IK (empty triangles) and true conditional cumulative probability functions (dashed line), together with its discrete version (thick line). The naive predictions of \mathbf{i}^* are also included (empty circles)

Under certain conditions (when classes are a priori equally probable), one can work with normalized log-likelihoods of each class, instead of log-odds. These normalized log-likelihoods are equivalent to classical disjunctive indicators, and may be interpolated using standard available software. To convert kriging results back to probability vectors, the analyst must first choose a normalizing value, which accounts for the odds of confusion between categories. This alternative procedure yields equal results to simplicial IK, but it also yields a valid estimation even when not using a full co-kriging procedure, just kriging each log-likelihood independently (though this is sub-optimal). In fact, this procedure should only be used in this way, ignoring cross-covariances, because the full co-kriging system of disjunctive indicators is singular.

Finally, a way to correctly model indicator covariances is also found. This is important because raw (disjunctive) indicator covariances must sum up to zero both by rows and columns, and for each lag. To avoid dealing with these constraints, one

might model the covariance structure of the coordinates. The fitted covariance model could be eventually back-transformed to a covariance model for the classical indicators, automatically fulfilling all its needed constraints.

Acknowledgements This research was partially developed in two stages with Prof. Dr. K. Gerald van den Boogaart (Greifswald University, Germany), to whom we are indebted for the fruitful discussions on the earlier versions of this method. These stages were funded by a 2005 Student Grant of the IAMG, and grants from the German Academic Exchange Service (Ref: A/04/33586) and the AGAUR of the Catalan Government (Ref: 2004-BE-00147). This research was developed within the scope of a PhD-grant (Ref: BR01/03) of the University of Girona (Spain), and research projects BPM2003-05640 and MTM2006-03040, funded by the Spanish Ministry of Education and Science. The constructive and in-depth reviews of two anonymous referees greatly improved this contribution.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix: Proofs

For an orthonormal basis, the coordinates of a vector \mathbf{p} may be straightforwardly computed as $\boldsymbol{\pi} = \boldsymbol{\Psi} \cdot \ln \mathbf{p}$, where $\boldsymbol{\Psi}$ is an orthonormal, rank-deficient matrix (Egozcue et al. 2003). For ease of reference, the next property states these characteristics without proof. (Note that here \mathbf{I} is not the cumulative indicator function of Journel 1983.)

Property 1 (Orthonormality of $\boldsymbol{\Psi}$) A matrix $\boldsymbol{\Psi}$ of computation of coordinates of a compositional vector in an orthonormal basis of the simplex \mathbb{S}^D fulfills

$$\begin{aligned}\boldsymbol{\Psi} \cdot \boldsymbol{\Psi}^t &= \mathbf{I}_{(D-1)}, \\ \boldsymbol{\Psi}^t \cdot \boldsymbol{\Psi} &= \mathbf{I}_D - \frac{1}{D} \cdot \mathbf{1}_D,\end{aligned}\tag{13}$$

with \mathbf{I} the identity matrix and $\mathbf{1}$ a matrix with all elements equal to 1. Subindexes show their number of rows and columns.

Proposition 1 (Coordinates of the generalized indicator function) Let \mathbf{J} be a disjunctive indicator transform (1), and let \mathbf{G} be a generalized indicator transform (3) defined with the same set of cutoffs. Then the vector of coordinates of \mathbf{G} , represented by $\boldsymbol{\pi}$, fulfills

$$\begin{aligned}\boldsymbol{\pi} &= \beta \cdot \boldsymbol{\Psi} \cdot \mathbf{J}, \\ \mathbf{J} &= \frac{1}{\beta} \cdot \boldsymbol{\Psi}^t \cdot \boldsymbol{\pi} + \frac{1}{D} \cdot \mathbf{1},\end{aligned}$$

where $\mathbf{1}$ is a vector of D ones, $\beta = \ln((1-b)(D-1)/b)$, and $\boldsymbol{\Psi}$ is the matrix representing the basis used.

Proof To prove the first equality, assume that category k was observed. The i -th coordinate is

$$\begin{aligned} \pi_i^* &= \sum_{j=1}^D \Psi_{ij} \ln G_j = \ln \left((1-b)^{\Psi_{ik}} \left(\frac{b}{D-1} \right)^{\sum_{j \neq k} \Psi_{ij}} \right) \\ &= \ln \left((1-b)^{\Psi_{ik}} \left(\frac{b}{D-1} \right)^{-\Psi_{ik}} \right) = \Psi_{ik} \ln \frac{1-b}{\frac{b}{D-1}} = \Psi_{ik} \cdot \beta, \end{aligned}$$

with $\beta = \ln((1-b)(D-1)/b)$. The key point of this equivalence comes from the fact that the rows of Ψ sum up to zero, thus $\sum_{j \neq k} \Psi_{ij} = -\Psi_{ik}$. The second equality of the proposition is obtained multiplying both sides of the first one by Ψ^t and taking into account Property 1, yielding

$$\begin{aligned} \Psi^t \cdot \pi &= \Psi^t \cdot (\beta \cdot \Psi \cdot \mathbf{J}) = \beta \cdot \Psi^t \cdot \Psi \cdot \mathbf{J} = \beta \cdot \left(\mathbf{I}_D - \frac{1}{D} \cdot \mathbf{1}_{D \times D} \right) \cdot \mathbf{J} \\ &= \beta \cdot \left(\mathbf{I}_D \cdot \mathbf{J} - \frac{1}{D} \cdot \mathbf{1}_{D \times D} \cdot \mathbf{J} \right) = \beta \cdot \left(\mathbf{J} - \frac{1}{D} \cdot \mathbf{1}_D \right). \end{aligned}$$

Here, \mathbf{I}_D is the identity $D \times D$ matrix, and the subindex placed in the vector/matrix $\mathbf{1}$ shows the dimension of this object. The last step uses the fact that $\sum J_i = 1$. \square

Proposition 2 (Moments of the coordinate random function) Denote the mean of the disjunctive indicators by $E[\mathbf{J}(\vec{x})] = \bar{\mathbf{J}}$. This vector forms a one-way contingency table, or the marginal distribution of the categorical random function. Denote the mean of coordinates as $E[\boldsymbol{\pi}(\vec{x})] = \bar{\boldsymbol{\pi}}$. Let $\text{Cov}[\mathbf{J}(\vec{x}), \mathbf{J}(\vec{x} + \vec{h})] = \mathbf{C}^J(\vec{h})$ be the covariance of disjunctive indicators and $\text{Cov}[\boldsymbol{\pi}(\vec{x}), \boldsymbol{\pi}(\vec{x} + \vec{h})] = \mathbf{C}^\pi(\vec{h})$ the covariance of the coordinates. Then, these moments fulfill

$$\begin{aligned} \bar{\boldsymbol{\pi}} &= \beta \cdot \Psi \cdot \bar{\mathbf{J}}, \\ \mathbf{C}^\pi(\vec{h}) &= \beta^2 \cdot \Psi \cdot \mathbf{C}^J(\vec{h}) \cdot \Psi^t, \\ \bar{\mathbf{J}} &= \frac{1}{\beta} \cdot \Psi^t \cdot \bar{\boldsymbol{\pi}} + \frac{1}{D} \cdot \mathbf{1}, \\ \mathbf{C}^J(\vec{h}) &= \frac{1}{\beta^2} \cdot \Psi^t \cdot \mathbf{C}^\pi(\vec{h}) \cdot \Psi. \end{aligned}$$

Sketch of a proof If \mathbf{Z} is a random vector and $\mathbf{Y} = \mathbf{T} \cdot \mathbf{Z} + \mathbf{b}$ is an affine linear transformation, it is well known that

$$\begin{aligned} E[\mathbf{Y}] &= \mathbf{T} \cdot E[\mathbf{Z}] + \mathbf{b}, \\ \text{Cov}[\mathbf{Y}] &= \mathbf{T} \cdot \text{Cov}[\mathbf{Z}] \cdot \mathbf{T}^t. \end{aligned}$$

But, according to Proposition 1, the relationship between \mathbf{J} and $\boldsymbol{\pi}$ is linear. Taking $\mathbf{T} = \beta \cdot \Psi$ and $\mathbf{b} = \mathbf{0}$, the first two equations are proven immediately; the same is attained for the last two equations with $\mathbf{T} = \beta^{-1} \cdot \Psi^t$ and $\mathbf{b} = \frac{1}{D} \cdot \mathbf{1}$. \square

Proposition 3 (Kriging of a full-rank linear transformation) For $\mathbf{Z}(\vec{x})$ a p -dimensional vector-valued random function, (simple) co-kriging it at \vec{x}_0 and applying a full-ranked linear transformation \mathbf{T} to the result is equivalent to applying (simple) co-kriging to the transformed values $\mathbf{Y}(\vec{x}) = \mathbf{T} \cdot \mathbf{Z}(\vec{x})$.

Proof Myers (1983) studied a similar expression, when \mathbf{T} is a (dual) vector. Let \mathbf{z}_0^* and \mathbf{y}_0^* represent the estimations obtained with simple co-kriging. Then, the goal is to show that $\mathbf{y}_0^* = \mathbf{T} \cdot \mathbf{z}_0^*$. Using matrix notation, it is common to write $\mathbf{z}_0^* = \mathbf{\Lambda}_z^t \cdot \mathbf{z}$, where $\mathbf{\Lambda}_z$ contains all the co-kriging weights arranged in a $pN \times p$ matrix, and \mathbf{z} is a vector containing the observations of the p components of $\mathbf{Z}(\vec{x})$ at N locations. These weights are computed with $\mathbf{\Lambda}_z = \mathbf{C}_z^{-1} \cdot \mathbf{C}_{z0}$, where \mathbf{C}_z contains the $pN \times pN$ covariances among all components at all observed locations, and \mathbf{C}_{z0} contains the $pN \times p$ covariances between the components at the observed locations and the components at the estimated location. The same can be written for $\mathbf{Y}(\vec{x})$, so that $\mathbf{y}_0^* = \mathbf{\Lambda}_y^t \cdot \mathbf{y}$, where $\mathbf{\Lambda}_y = \mathbf{C}_y^{-1} \cdot \mathbf{C}_{y0}$. If the covariance models built for \mathbf{Z} and \mathbf{Y} are compatible, then these matrices satisfy the same linear relationship as vectors, i.e. $\mathbf{C}_y = \mathbf{T}_N \cdot \mathbf{C}_z \cdot \mathbf{T}_N^t$ and $\mathbf{C}_{y0} = \mathbf{T}_N \cdot \mathbf{C}_{z0} \cdot \mathbf{T}^t$, with \mathbf{T}_N a block diagonal matrix with N reproductions of matrix \mathbf{T} on the diagonal. Then the co-kriging weights of these two systems fulfill:

$$\begin{aligned} \mathbf{\Lambda}_y &= \mathbf{C}_y^{-1} \cdot \mathbf{C}_{y0} = (\mathbf{T}_N \cdot \mathbf{C}_z \cdot \mathbf{T}_N^t)^{-1} \cdot \mathbf{T}_N \cdot \mathbf{C}_{z0} \cdot \mathbf{T}^t \\ &= \mathbf{T}_N^{-t} \cdot \mathbf{C}_z^{-1} \cdot \mathbf{T}_N^{-1} \cdot \mathbf{T}_N \cdot \mathbf{C}_{z0} \cdot \mathbf{T}^t = \mathbf{T}_N^{-t} \cdot \mathbf{C}_z^{-1} \cdot \mathbf{C}_{z0} \cdot \mathbf{T}^t = \mathbf{T}_N^{-t} \cdot \mathbf{\Lambda}_z \cdot \mathbf{T}^t. \end{aligned}$$

If this relationship can be inverted, $\mathbf{\Lambda}_z = \mathbf{T}_N^t \cdot \mathbf{\Lambda}_y \cdot \mathbf{T}^{-t}$. Now we can compute the predictor $\mathbf{z}_0^* = \mathbf{\Lambda}_z^t \cdot \mathbf{z} = (\mathbf{T}_N^t \cdot \mathbf{\Lambda}_y \cdot \mathbf{T}^{-t})^t \cdot \mathbf{z} = \mathbf{T}^{-1} \cdot \mathbf{\Lambda}_y^t \cdot \mathbf{T}_N \cdot \mathbf{z}$, or equivalently, $\mathbf{T} \cdot \mathbf{z}_0^* = \mathbf{\Lambda}_y^t \cdot \mathbf{T}_N \cdot \mathbf{z}$. But, given that $\mathbf{T}_N \cdot \mathbf{z} = \mathbf{y}$, this finally implies that $\mathbf{T} \cdot \mathbf{z}_0^* = \mathbf{\Lambda}_y^t \cdot \mathbf{y} = \mathbf{y}_0^*$. \square

Proposition 4 (Estimator of simplicial Indicator Kriging) If \mathbf{j}_0^* is the estimator obtained by the classical indicator co-kriging on $\mathbf{J}(\vec{x})$, then the estimator of the probability vector $\mathbf{p}(\vec{x}_0)$ from a simplicial point of view is

$$\mathbf{p}_{sIK}^*(\vec{x}_0) = \mathcal{C}(\exp(\beta \cdot \mathbf{j}_0^*)) = \beta \odot \exp(\mathbf{j}_0^*).$$

Proof To prove Proposition 4, one must replace (11) within (6), to get $\mathbf{p}^*(\vec{x}_0) = \mathcal{C}(\exp(\mathbf{\Psi}^t \cdot \boldsymbol{\pi}^*)) = \mathcal{C}(\exp(\mathbf{\Psi}^t \cdot \beta \cdot \mathbf{\Psi} \cdot \mathbf{j}_0^*))$. Then, using relation (13), one obtains $\mathbf{p}^*(\vec{x}_0) = \mathcal{C}(\exp(\beta \cdot (\mathbf{I}_D - \frac{1}{D} \cdot \mathbf{1}_D) \cdot \mathbf{j}_0^*))$, and some standard algebraic manipulations provide

$$\begin{aligned} \mathbf{p}^*(\vec{x}_0) &= \mathcal{C}\left(\exp\left(\beta \cdot \mathbf{j}_0^* - \frac{\beta}{D} \cdot \mathbf{1}_D \cdot \mathbf{j}_0^*\right)\right) \\ &= \mathcal{C}\left(\exp\left(\beta \cdot \mathbf{j}_0^*\right) \cdot \exp\left(-\frac{\beta}{D} \cdot \mathbf{1}_D \cdot \mathbf{j}_0^*\right)\right) \\ &= \mathcal{C}\left(\exp\left(\beta \cdot \mathbf{j}_0^*\right)\right) \oplus \mathcal{C}\left(\exp\left(-\frac{\beta}{D} \cdot \mathbf{1}_D \cdot \mathbf{j}_0^*\right)\right). \end{aligned}$$

Recalling that $\mathbf{1}_D$ is a matrix with $D \times D$ ones, $\mathbf{1}_D \cdot \mathbf{j}_0^*$ becomes a constant vector, thus $\mathcal{C}(\exp(-\frac{\beta}{D} \cdot \mathbf{1}_D \cdot \mathbf{j}_0^*)) = \mathbf{n}$, the neutral element of perturbation. But perturbing a vector by the neutral element leaves it unchanged, thus

$$\mathbf{p}^*(\vec{x}_0) = \mathcal{C}(\exp(\beta \cdot \mathbf{j}_0^*)) \oplus \mathbf{n} = \mathcal{C}([\exp(\mathbf{j}_0^*)]^\beta),$$

which yields the sought result taking into account the definition of powering. \square

References

- Aitchison J (1984) Reducing the dimensionality of compositional data sets. *Math Geol* 16(6):617–636
- Aitchison J (1986) The statistical analysis of compositional data, Monographs on statistics and applied probability. Chapman & Hall, London (Reprinted in 2003 with additional material by The Blackburn Press)
- Billheimer D, Guttorp P, Fagan W (2001) Statistical interpretation of species composition. *J Am Stat Assoc* 96(456):1205–1214
- Bogaert P (1999) On the optimal estimation of the cumulative distribution function in presence of spatial dependence. *Math Geol* 3(2):213–239
- Bogaert P (2002) Spatial prediction of categorical variables: The Bayesian maximum entropy approach. *Stoch Environ Res Risk Assess* 16:425–448
- Carle S, Fogg G (1996) Transition probability-based indicator geostatistics. *Math Geol* 28(4):453–476
- Carr J (1994) Order relation correction experiments for probability kriging. *Math Geol* 26(5):605–621
- Carr J, Mao N (1993) A general-form of probability kriging for estimation of the indicator and uniform transforms. *Math Geol* 25(4):425–438
- Christakos G (1990) A Bayesian/maximum entropy view to the spatial estimation problem. *Math Geol* 22(7):763–777
- Deutsch C, Journé A (1998) GSLIB—geostatistical software library and user’s guide, 2nd edn. Oxford University Press, New York
- Egozcue JJ, Pawłowsky-Glahn V (2005) Groups of parts and their balances in compositional data analysis. *Math Geol* 37(7):795–828
- Egozcue JJ, Pawłowsky-Glahn V (2006) Simplicial geometry for compositional data. In: Buccianti A, Mateu-Figueras G, Pawłowsky-Glahn V (eds) Compositional data analysis in the geosciences. Geological Society of London, London, pp 145–169
- Egozcue JJ, Pawłowsky-Glahn V, Mateu-Figueras G, Barceló-Vidal C (2003) Isometric logratio transformations for compositional data analysis. *Math Geol* 35(3):279–300
- Journé AG (1983) Nonparametric estimation of spatial distributions. *Math Geol* 15(3):445–468
- Juang K, Lee D, Hsiao C (1998) Kriging with cumulative distribution function of order statistics for delimitation of heavy-metal contaminated soils. *Soil Sci* 163(10):797–804
- Lindley DV (1964) The Bayesian analysis of contingency tables. *Ann Math Stat* 35(4):1622–1643
- Mateu-Figueras G, Pawłowsky-Glahn V, Barceló-Vidal C (2003) Distributions on the simplex. In: Thió-Henestrosa S, Martín-Fernández JA (eds) Compositional data analysis workshop—CoDaWork’03, proceedings. Universitat de Girona, Girona
- Matheron G (1976) A simple substitute for the conditional expectation: The disjunctive kriging. In: Guarascio M, David M, Huijbregts C (eds) Advanced geostatistics in the mining industry. NATO advances study institute series; series C: mathematical and physical sciences, vol 24. Reidel, Dordrecht, pp 221–236
- Matheron G (1993) Une conjecture sur la covariance d’un ensemble aleatoire. In: Cahiers de géostatistique, vol 3. Ecole de Mines de Paris, Fontainebleau, pp 107–113
- Myers DE (1983) Estimation of linear combinations and co-kriging. *Math Geol* 15(5):633–637
- Pardo-Igúzquiza E, Dowd P (2005) Multiple indicator cokriging with application to optimal sampling for environmental monitoring. *Comput Geosci* 31(1):1–13
- Pawłowsky V (1989) Cokriging of regionalized compositions. *Math Geol* 21(5):513–521
- Pawłowsky-Glahn V (2003) Statistical modelling on coordinates. In: Thió-Henestrosa S, Martín-Fernández JA (eds) Compositional data analysis workshop—CoDaWork’03, proceedings. Universitat de Girona, Girona

- Pawlowsky-Glahn V, Egozcue JJ (2001) Geometric approach to statistical analysis on the simplex. *Stoch Environ Res Risk Assess* 15(5):384–398
- Pawlowsky-Glahn V, Egozcue JJ (2002) BLU estimators and compositional data. *Math Geol* 34:259–274
- Pawlowsky-Glahn V, Olea RA (2004) Geostatistical analysis of compositional data. *Studies in mathematical geology*, vol 7. Oxford University Press, Oxford
- Sullivan J (1984) Conditional recovery estimation through probability kriging—theory and practice. In: *Geostatistics for natural resources characterization*, 2nd edn. NATO-ASI, Stanford
- Suro-Perez V, Journel A (1991) Indicator principal component kriging. *Math Geol* 23(5):759–788
- Tolosana-Delgado R (2006) Geostatistics for constrained variables: Positive data, compositions and probabilities. Applications to environmental hazard monitoring. PhD Thesis, Universitat de Girona, Girona, Spain, p 198
- Tolosana-Delgado R, Pawlowsky-Glahn V (2007) Kriging regionalized positive variables revisited sample space and scale considerations. *Math Geol* 39(6):529–558
- Tolosana-Delgado R, Pawlowsky-Glahn V, J Egozcue J, van den Boogaart KG (2005) A compositional approach to indicator kriging. In: Cheng Q, Bonham-Carter G (eds) 2005 annual conference of the international association for mathematical geology (IAMG), Toronto, Canada, pp 651–656
- Vargas-Guzman J, Dimitrakopoulos R (2003) Successive nonparametric estimation of conditional distributions. *Math Geol* 35(1):39–52