

# Means and covariance functions for geostatistical compositional data: an axiomatic approach

Denis Allard<sup>a</sup>, Thierry Marchant<sup>b</sup>

<sup>a</sup>Biostatistics and Spatial Processes, BioSP, INRA, 84914, Avignon, France.

<sup>b</sup>Ghent University, Ghent, Belgium.

Corresponding author:

D. Allard, BioSP, 84914 Avignon, FRANCE

Tel: +33 432722171; Fax: +33 432722182

`denis.allard@inra.fr`

## Abstract

This work focuses on the characterization of the central tendency of a sample of compositional data. It provides new results about theoretical properties of means and covariance functions for compositional data, with an axiomatic perspective. Original results that shed new light on the geostatistical modeling of compositional data are presented. As a first result, it is shown that the weighted arithmetic mean is the only central tendency characteristic satisfying a small set of axioms, namely continuity, reflexivity and marginal stability. Moreover, this set of axioms also implies that the weights must be identical for all parts of the composition. This result has deep consequences on the spatial multivariate covariance modeling of compositional data. In a geostatistical setting, it is shown as a second result that the proportional model of covariance functions (i.e., the product of a covariance matrix and a single correlation function) is the only model that provides identical kriging weights for all components of the compositional data. As a consequence of these two results, the proportional model of covariance function is the only covariance model compatible with reflexivity and marginal stability.

**Keywords** Aitchison geometry; central tendency; functional equation; geostatistics; multivariate covariance function model

---

Authors are listed alphabetically. They contributed equally.

# 1 Introduction

This work focuses on the characterization of the central tendency of a sample of compositional data and on consequences regarding its geostatistical modeling. Compositional vectors “describe quantitatively the parts of some whole” (Egozcue, 2011). They convey information about relative values of components, which are usually expressed in proportions or percentages. Their modeling and their analysis is therefore different from those of unconstrained multivariate vectors. In order to facilitate their analysis, it is natural to select a representative of the equivalent class (Egozcue, 2011). Therefore, compositional data with  $p$  variables,  $\mathbf{x} = (x^1, \dots, x^p)$ , has positive components that add up to a constant, say  $\kappa$ . Without loss of generality, one can set  $\kappa = 1$  for the rest of this paper. Compositional data thus belongs to the positive simplex of dimension  $p - 1$

$$\mathbb{S}^{p-1} = \{(x^1, \dots, x^p) : x^k \geq 0 \text{ for } k = 1, \dots, p, \text{ and } x^1 + \dots + x^p = 1\}. \quad (1)$$

The central tendency of a sample of compositional data, which is also a compositional data, is denoted by  $\mathbf{M}(\mathbf{x}_1, \dots, \mathbf{x}_n)$ . Aitchison (1989) states that the arithmetic mean is “clearly useless as a measure of location because it falls outside of the array of compositions and is indeed very atypical of the data set” and that the normalized geometric mean (i.e., the back-transform of the arithmetic mean of the log-ratios) “serves equally well for curved data sets and for more linear and elliptical data sets”. On ternary sub-compositions of hongite Sharp (2006) noted that in many instances the geometric mean also falls outside of the array of compositions, a fact already pointed out in Shurtz (2000). Probably the most striking example is the Hongite 2 artificial data sets where all samples are projected on a line parallel to one side of the triangle. The arithmetic average is shown to belong to the same line, while the geometric average does not (Sharp, 2006, Fig. 2).

As an alternative to arithmetic or geometric means, Sharp (2006) proposed the graph median. It is built from the minimum spanning tree, which is the graph connecting all points of the data set whose total length (sum of the length of the edges of the graph) is minimal. The distance used in the simplex is the half-taxi metric (Miller, 2002)  $d(\mathbf{x}, \mathbf{x}') = 0.5 \sum_{k=1}^p |x^k - x'^k|$ . The graph median is then obtained by iteratively pruning the outermost branches of the tree until only one point or a pair of points remain. This point or the mid-point between the pair of points is then the graph median. The minimum spanning tree is not always unique, in which case a tie breaking rule is necessary. Very often, the graph median is one of the sample points of the data. Otherwise, it is the mid-point between two close data samples. By construction it is located in the innermost region of the data-set, thus defining a central tendency (Sharp, 2006). However, it is complicated to compute, and it cannot be easily related to the estimation of a total quantity or indeed to most statistical or geostatistical

analysis. Moreover, it is not continuous with respect to the data values. This alternative will not be considered in this work.

The statistical analysis of compositional data has received great attention in the last three decades. Compositional data are often transformed into a new  $p$ -dimensional vector of log-ratios of the components (Aitchison, 1982, 1986; Pawlowsky-Glahn and Olea, 2004; Egozcue, 2003). These transformations provide one-to-one mappings onto a real space, thereby allowing usual multivariate statistics methods to be applied to the transformed data. Any statement made in the transformed space is easily translated back into the compositional space. Billheimer (2001) and Pawlowsky-Glahn and Egozcue (2002) showed that the simplex  $\mathbb{S}^{p-1}$  equipped with the scalar product

$$\langle \mathbf{x}, \mathbf{y} \rangle_A = \frac{1}{p} \sum_{i=1}^p \sum_{j=i+1}^p \ln \frac{x_i}{x_j} \ln \frac{y_i}{y_j} = \sum_{i=1}^p \ln \frac{x_i}{g(\mathbf{x})} \ln \frac{y_i}{g(\mathbf{y})}, \quad \mathbf{x}, \mathbf{y} \in \mathbb{S}^{p-1}, \quad (2)$$

where  $g(\mathbf{x}) = [x^1 x^2 \dots x^p]^{1/p}$  is the geometric mean of the components of  $\mathbf{x}$ , induces a distance and thus a geometry on the simplex, called the Aitchison geometry. An orthonormal basis on the simplex was proposed in (Egozcue, 2003; Mateu-Figueras, 2011) where the data  $(x^1, \dots, x^p)$  are transformed by means of the isometric log-ratio (ilr) transformations

$$u^i = [i(i+1)]^{-1/2} \ln \frac{x^1 x^2 \dots x^i}{(x^{i+1})^i}, \quad i = 1, \dots, p-1, \quad (3)$$

with  $\mathbf{u} = (u^1, \dots, u^{p-1}) \in \mathbb{R}^{p-1}$ . Some properties of compositional data and their parameters can be worked out directly within the Aitchison geometry on the simplex. However, when back-transforming these to the standard Euclidean space of the raw data, unexpected behavior may arise. Following (Mateu-Figueras, 2011, pp. 35–36), let us consider the following example, with  $\mathbf{x} = (0.6, 0.3, 0.1)$  and  $\mathbf{x}' = (0.3, 0.3, 0.4)$  being two compositions in  $\mathbb{S}^2$ . Their ilr transforms  $\mathbf{u} = (0.490, 1.180)$  and  $\mathbf{u}' = (0.000, -0.235)$  correspond to orthonormal coordinates with respect to the ilr transformation. We can then apply standard operations on these coordinates. For example, their arithmetic mean is  $\bar{\mathbf{u}} = (0.245, 0.4725)$ . Once back-transformed by the inverse operations of Eq. (3), we obtain

$$\text{ilr}^{-1}(\bar{\mathbf{u}}) = \tilde{\mathbf{x}} = (0.459, 0.325, 0.216),$$

which is nothing but the normalized geometric mean of  $\mathbf{x}$  and  $\mathbf{x}'$ . An unexpected and intriguing fact is that even though the second coordinates are equal to 0.3 for both data, the second coordinate of  $\tilde{\mathbf{x}}$  is increased by 8.3%.

In the simplex, whenever one component is increased (resp. decreased), some or all other components must decrease (resp. increase) in order to satisfy the sum constraint, a fact

that has consequences on the conditions one wishes to impose on  $\mathbf{M}$ . Consider a dataset  $\mathbf{x}_1, \dots, \mathbf{x}_n$  of compositional data with  $q$  variables (i.e.,  $\mathbf{x}_i = (x_i^1, \dots, x_i^q)$ ) to be grouped into  $p < q$  variables in two different ways, thereby defining two new datasets  $\mathbf{y}_1, \dots, \mathbf{y}_n$  and  $\mathbf{z}_1, \dots, \mathbf{z}_n$ . We further assume that the first variable is identical in the two groupings, i.e.,  $y_i^1 = z_i^1$ ,  $i = 1, \dots, n$ . For instance, with  $q = 4$  and  $p = 3$ , one grouping is  $1, \{2, 3\}, 4$  and the other one is  $1, 2, \{3, 4\}$ . One does not expect the mean of the first variable to depend on the grouping of the other variables. This is the case for the arithmetic mean. Indeed, for the first grouping, the  $k$ th component of the arithmetic mean is  $M_a^k(\mathbf{y}_1, \dots, \mathbf{y}_n) = n^{-1} \sum_{i=1}^n y_i^k$ , and a similar expression holds for  $M_a^k(\mathbf{z}_1, \dots, \mathbf{z}_n)$ . Since the first variable is common to both datasets, it is clear that  $M_a^1(\mathbf{y}_1, \dots, \mathbf{y}_n) = M_a^1(\mathbf{z}_1, \dots, \mathbf{z}_n)$ . The arithmetic mean is said to satisfy marginal stability.

Let us now compute the normalized geometric mean independently for each variable. Then, for the first grouping

$$M_g^k(\mathbf{y}_1, \dots, \mathbf{y}_n) = \frac{(\prod_{i=1}^n y_i^k)^{1/n}}{\sum_{l=1}^p (\prod_{i=1}^n y_i^l)^{1/n}},$$

for  $k = 1, \dots, p$ , with a similar expression for the second grouping. Since all variables are involved in the computation of each component of the normalized geometric mean, equality of the mean of the first variable for the two groupings will be different, i.e.,  $M_g^1(\mathbf{y}_1, \dots, \mathbf{y}_n) \neq M_g^1(\mathbf{z}_1, \dots, \mathbf{z}_n)$ . In other words, the normalized geometric mean does not satisfy marginal stability. A similar result is obtained for any log-ratio transform. Marginal stability is an important property that one may choose to impose or not when analyzing compositional data. It plays a critical role in our results.

For applications, such as oil and mining industry or soil remediation, where compositions correspond to fractions of soil type or rock type, the above examples are puzzling. Based on samples of compositional samples, practitioners working in these fields need to compute reliable estimates of the total amount of material for a given volume, or equivalently, estimates of the mean of the fraction. Unanswered questions arise: Under which conditions is marginal stability always satisfied on central tendencies? In a geostatistical setting, what are the conditions for having unbiased estimates of means and what are the multivariate covariance functions compatible with marginal stability?

Through an axiomatic perspective, we provide answers to these questions. New mathematical results about theoretical properties of central tendencies and compatible covariance functions for compositional data are presented. These results shed new light on its geostatistical modeling. We start from a set of axioms that correspond to quite natural conditions: continuity, reflexivity (which will be defined later) and marginal stability. Other sets of axioms

would lead to different mathematical properties. For example, subcompositional dominance is violated by the Euclidean distance between compositional data (Egozcue, 2011). It is not the purpose of the present work to discuss the relevance of sets of axioms for the analysis of data. We merely derive “if and only if” relationships between one given set of axioms and some mathematical properties, which we believe are useful in some situations.

As a first result, it will be shown that the weighted arithmetic mean is the only central tendency characteristic satisfying reflexivity and marginal stability. Moreover, the weights must be identical for all components of the compositional vector. This first result holds in the standard Euclidean space of the raw data. It has deep consequences on the multivariate covariance modeling. It is well known (and easy to verify) that, if the multivariate covariance function belongs to the family of proportional models (i.e., it is the product of a covariance matrix and a single correlation function), the kriging weights are identical for all variables (Helterbrand and Cressie, 1994). It will be shown that the converse also holds for all multivariate random fields, compositional or not. As a second result on compositional data, it is then established that the proportional model of covariance functions is, in the Euclidean space of raw data, the only covariance model satisfying reflexivity and marginal stability. This second result does not necessarily hold within the Aitchison geometry on the simplex.

The paper is structured as follows. Section 2 provides a quick presentation of the axiomatic characterization of the possible definitions of means. Section 3 presents our first main result on the mean of compositional data. In Sect. 4, the consequences of this theorem on the covariance functions are shown. These results are then discussed in Sect. 5.

## 2 A primer to axiomatic definitions of means

### 2.1 Means for univariate data

Many ways exist of summarizing a sample  $x_1, \dots, x_n$  of a variable defined on an interval  $E \subseteq \mathbb{R}$  into a single value, usually called the mean, and sometimes also called central tendency. The most widely used are the arithmetic, geometric or harmonic means, the mode, the median, and more generally any desired quantile. They all satisfy different properties. For example, the arithmetic mean  $\bar{x} = n^{-1} \sum_{i=1}^n x_i$  is the real number  $m$  minimizing  $\sum_{i=1}^n (m - x_i)^2$ . It is unique, easy to compute, but sensitive to large values or outliers. The median is the real number  $m$  minimizing  $\sum_{i=1}^n |m - x_i|$ . It is sometimes not unique, but robust to large values and outliers. An axiomatic characterization of the arithmetic mean was obtained as early as Kolmogorov (1930). Let us denote by  $M(x_1, \dots, x_n)$  a mapping from  $E^n \rightarrow E$ , which will be the central tendency, or the mean, of the sample, and let us impose the following quite natural conditions, called axioms. Kolmogorov’s characterization follows.

K1 *Continuity and strict monotonicity*:  $M(x_1, \dots, x_n)$  is continuous and strictly monotonic in each of its arguments. Strict monotonicity states that if one of the values  $x_i$  increases, the central tendency must also increase. Continuity imposes that small variations in one of the values  $x_i$  leads to small variations of  $M$ .

K2 *Symmetry*:  $M(x_1, \dots, x_n)$  is symmetric, i.e., it is identical for any permutation of the sample.

K3 *Reflexivity*: The central tendency of identical values is equal to their common value, i.e.,  $M(x, \dots, x) = x$ .

K4 *Associativity*: A subset of the sample can be replaced by its central tendency with no effect on the total central tendency:

$$M(x_1, \dots, x_m, x_{m+1}, \dots, x_n) = M(x_*, \dots, x_*, x_{m+1}, \dots, x_n)$$

where  $x_* = M(x_1, \dots, x_m)$ .

**Theorem 1 (Kolmogorov, 1930)** *Conditions K1 to K4 hold if and only if the central tendency  $M$  has the form*

$$M(x_1, \dots, x_n) = \phi^{-1} \left( \frac{\phi(x_1) + \dots + \phi(x_n)}{n} \right), \quad (4)$$

where  $\phi$  is a continuous strictly increasing function on  $E$ , called the generating function.

Functions  $M$  having the form of Eq. (4) are called quasi-arithmetic means in the functional equation literature (Aczél and Dhombres, 1989; Matkowski, 2010). It is worth noting that neither the mode nor the median belong to this family. The median is not continuous, while neither the mode nor the median satisfy K4. Quasi-arithmetic means cover a wide range of well known means: if  $\phi(x) = x$ ,  $M$  is the arithmetic mean. On  $E = (0, +\infty)$ ,  $\phi(x) = \ln x$  leads to the geometric mean, while  $\phi(x) = x^{-1}$  leads to the harmonic mean. More generally, when  $\phi(x) = x^\alpha$ ,  $\alpha \neq 0$  and  $x \in (0, +\infty)$  the associated mean is called the power mean.

## 2.2 Means for multivariate data

For multivariate data, each data is a vector of length  $p$ , i.e.,  $\mathbf{x}_i = (x_i^1, \dots, x_i^p)$ , with  $i = 1, \dots, n$ . In this case, the central tendency is a vector  $\mathbf{M} = (M^1, \dots, M^p)$ , where  $M^k$  is a mapping from  $E^n \rightarrow E$ . Possibilities are numerous. A tempting simplification is to impose that each component  $M^k$  depends only on the values of the corresponding variable. In this case one builds one central tendency per variable (e.g., the arithmetic mean or the median), independently of all other variables. When applying Kolmogorov's Theorem to

each variable independently, the associated central tendency  $M^k$  is a quasi-arithmetic mean if and only if conditions K1-K4 hold. When  $E \subseteq \mathbb{R}$ , the multivariate arithmetic mean is the point  $\mathbf{M}(\mathbf{x}_1, \dots, \mathbf{x}_n) = \mathbf{m} \in \mathbb{R}^p$  minimizing  $\sum_{i=1}^n \|\mathbf{m} - \mathbf{x}_i\|_p^2$ , where  $\|\cdot\|_p$  is the Euclidean distance in  $\mathbb{R}^p$ . In this case, each component of  $\mathbf{M}(\mathbf{x}_1, \dots, \mathbf{x}_n)$  is the arithmetic mean of the corresponding variable.

Although being natural and probably widely used, the componentwise simplification is by no means the only mathematical possibility. For instance, in analogy to the definition of the median for univariate data, the multivariate median is the  $p$ -dimensional vector  $\mathbf{m}$  minimizing  $\sum_{i=1}^n \|\mathbf{m} - \mathbf{x}_i\|_p$ . It is unique whenever the points are not colinear (Vardi and Zhang, 2000). Its  $k$ -th coordinate,  $m^k$ , depends on the values of all coordinates of the samples  $\mathbf{x}_1, \dots, \mathbf{x}_n$ . Many alternative multivariate medians can be defined, such as those based on the notion of statistical depth (Liu et al., 1999; Zuo and Sefling, 2000) or the graph median proposed in Sharp (2006).

### 3 An axiomatic characterization of the mean for compositional data

As seen in the Introduction, there exist many possible approaches to defining a mean for compositional data, and they all satisfy different properties. Inspired by the axiomatic approach briefly summarized in Sect. 2, a new characterization theorem for compositional data is presented. It relies on functional equation arguments and does not necessitate the use of inner products and distances. It is independent of any geometry on the simplex.

Consider a sample of fixed size  $n$  of compositional data belonging to the positive simplex  $\mathbb{S}^{p-1}$ , corresponding to a regionalized variable  $\mathbf{x}(\cdot)$  defined in  $\mathbb{R}^d$ , sampled at sites  $\mathbf{s}_1, \dots, \mathbf{s}_n$ . For the sake of simplicity, we will write  $\mathbf{x}_i = \mathbf{x}(\mathbf{s}_i)$ . We want to characterize the mapping  $\mathbf{M} = (M^1, \dots, M^p)$ , which associates to each data-set  $(\mathbf{x}_1, \dots, \mathbf{x}_n)$  a  $p$ -dimensional mean vector  $\mathbf{M}(\mathbf{x}_1, \dots, \mathbf{x}_n)$  with components  $M^k(\mathbf{x}_1, \dots, \mathbf{x}_n)$ , for  $k = 1, \dots, p$ . Notice that at this stage  $M^k(\mathbf{x}_1, \dots, \mathbf{x}_n)$  can depend on  $\mathbf{x}_1, \dots, \mathbf{x}_n$  and not just on  $x_1^k, \dots, x_n^k$ , so that the definition of  $\mathbf{M}$  is very general.

We will only consider mappings  $\mathbf{M}$  such that  $\sum_{k=1}^p M^k(\mathbf{x}_1, \dots, \mathbf{x}_n) = 1$ . In order to account for the fact that Kriging can result in negative weights and that this, in turn, can yield negative estimates for some specific data sets, we will not assume from the onset that  $0 \leq M^k(\mathbf{x}_1, \dots, \mathbf{x}_n) \leq 1$ . The set of all real vectors  $\mathbf{x}$  of size  $p$  satisfying  $\sum_{k=1}^p x^k = 1$  will be denoted by  $\mathcal{S}^{p-1}$ . It is a superset of the simplex  $\mathbb{S}^{p-1}$ .

Conditions imposed for the characterization of the central tendency of spatial compositional data differ from those imposed for Kolmogorov's characterization. Among Kol-

mogorov's conditions, we retain only two: reflexivity and continuity. As discussed in the previous section, marginal stability is also imposed. Our main theorem, characterizing the means for compositional data, follows.

C1 Reflexivity: for all  $\mathbf{x} \in \mathbb{S}^{p-1}$ ,  $\mathbf{M}(\underbrace{\mathbf{x}, \dots, \mathbf{x}}_n) = \mathbf{x}$ .

C2 Marginal stability: for any  $k = 1, \dots, p$ , any  $i = 1, \dots, n$  and any  $\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{x}'_i$  in  $\mathbb{S}^{p-1}$ , if  $x_i^k = x_i'^k$ , then

$$M^k(\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_n) = M^k(\mathbf{x}_1, \dots, \mathbf{x}'_i, \dots, \mathbf{x}_n).$$

C3 Continuity:  $\mathbf{M}$  is continuous in each argument.

In presence of spatial correlation, spatial symmetry is in general not desired, but can be an option. For the sake of completeness it is recalled.

C4 Symmetry:

$$M^k(\mathbf{x}_1, \dots, \mathbf{x}_n) = M^k(\mathbf{x}_{\sigma(1)}, \dots, \mathbf{x}_{\sigma(n)})$$

for any permutation  $\sigma$  of  $\{1, \dots, n\}$  and any  $\mathbf{x}_1, \dots, \mathbf{x}_n$  of  $\mathbb{S}^{p-1}$ .

We are now ready to state our first result. In part A, we do not impose that the components of  $\mathbf{M}(\mathbf{x}_1, \dots, \mathbf{x}_n)$  be nonnegative (i.e.  $\mathbf{M} : (\mathbb{S}^{p-1})^n \rightarrow \mathbb{S}^{p-1}$ ). In part B, this nonnegativity constraint is imposed (i.e.  $\mathbf{M} : (\mathbb{S}^{p-1})^n \rightarrow \mathbb{S}^{p-1}$ ).

**Theorem 2 (Characterization of the mean)** *Consider compositional data  $(\mathbf{x}_1, \dots, \mathbf{x}_n)$  with  $p \geq 3$  and  $n \geq 2$ .*

A) *The mapping  $\mathbf{M} : (\mathbb{S}^{p-1})^n \rightarrow \mathbb{S}^{p-1}$  satisfies conditions C1-C3 if and only if, for  $k \in \{1, \dots, p\}$ ,*

$$M^k(\mathbf{x}_1, \dots, \mathbf{x}_n) = \sum_{i=1}^n \lambda_i x_i^k, \quad (5)$$

*for some real numbers  $(\lambda_1, \dots, \lambda_n)$  satisfying  $\sum_{i=1}^n \lambda_i = 1$ .*

B) *The mapping  $\mathbf{M} : (\mathbb{S}^{p-1})^n \rightarrow \mathbb{S}^{p-1}$  satisfies conditions C1-C2 if and only if, for  $k \in \{1, \dots, p\}$ ,*

$$M^k(\mathbf{x}_1, \dots, \mathbf{x}_n) = \sum_{i=1}^n \lambda_i x_i^k, \quad (6)$$

*for some non-negative real numbers  $(\lambda_1, \dots, \lambda_n)$  satisfying  $\sum_{i=1}^n \lambda_i = 1$ .*

*Furthermore, if symmetry holds, then  $\lambda_i = 1/n$  for all  $i \in \{1, \dots, n\}$  and  $\mathbf{M}(\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathbb{S}^{p-1}$ .*

The key ingredients of the proof, presented in the Appendix, are marginal stability and sum to 1. Marginal stability imposes that, for each  $k = 1, \dots, p$ ,  $M^k$  depends only on  $x_1^k, \dots, x_n^k$ . Then, the sum to 1 condition implies that the generating function  $\phi$  must be the identity function, i.e., that the only means are linear combinations. Clearly, the same result holds if the data sum to another constant than 1. Notice that when the mapping  $\mathbf{M}$  is bounded, continuity of the mapping (axiom C3) is not necessary for the proof. However, since linear combinations are continuous, continuity is satisfied in both cases. As a consequence, the only means satisfying C1-C3 are linear combinations, whether or not  $M^k(\mathbf{x}_1, \dots, \mathbf{x}_n)$  is forced to belong to  $\mathbb{S}^{p-1}$ .

Theorem 2 holds whenever there are at least three components. With two components there is only one independent variable, and hence many more means satisfying axioms C1-C3 exist. As an example, it can easily be shown that the Kolmogorov means characterized in Eq. (4)

$$M^k(\mathbf{x}_1, \dots, \mathbf{x}_n) = \phi^{-1} \left( \sum_{i=1}^n \lambda_i \phi(x_i^k) \right)$$

also satisfy marginal stability and sum to 1 whenever  $\phi : [0, 1] \rightarrow \mathbb{R}$  is such that  $\phi(t) = 1 - \phi(1 - t)$ , with  $\phi(0) = 0$  and  $\phi(1) = 1$ . Indeed,

$$\begin{aligned} M^1(\mathbf{x}_1, \dots, \mathbf{x}_n) + M^2(\mathbf{x}_1, \dots, \mathbf{x}_n) &= \phi^{-1} \left( \sum_{i=1}^n \lambda_i \phi(x_i^1) \right) + \phi^{-1} \left( \sum_{i=1}^n \lambda_i \phi(1 - x_i^1) \right) \\ &= \phi^{-1} \left( \sum_{i=1}^n \lambda_i \phi(x_i^1) \right) + \phi^{-1} \left( \sum_{i=1}^n \lambda_i (1 - \phi(x_i^1)) \right) \\ &= \phi^{-1} \left( \sum_{i=1}^n \lambda_i \phi(x_i^1) \right) + \phi^{-1} \left( 1 - \sum_{i=1}^n \lambda_i \phi(x_i^1) \right) \end{aligned}$$

Since  $\phi(t) = 1 - \phi(1 - t)$  implies  $\phi^{-1}(u) = 1 - \phi^{-1}(1 - u)$  for  $0 \leq u \leq 1$ , the last equality implies

$$M^1(\mathbf{x}_1, \dots, \mathbf{x}_n) + M^2(\mathbf{x}_1, \dots, \mathbf{x}_n) = \phi^{-1} \left( \sum_{i=1}^n \lambda_i \phi(x_i^1) \right) + 1 - \phi^{-1} \left( \sum_{i=1}^n \lambda_i \phi(x_i^1) \right) = 1.$$

The case  $p = 2$  is not of central interest for the analysis of compositional data. In the rest of this work, it is thus assumed that  $p \geq 3$ .

## 4 Geostatistical compositional data

### 4.1 Characterization of compatible covariance models

Within a (geo-) statistical setting, the condition  $\sum_{i=1}^n \lambda_i = 1$  in Eq. (5) corresponds to an unbiasedness condition. Theorem 2 states that means of compositional data satisfying axioms C1-C3 correspond to linear unbiased estimators of the population mean. However, this theorem does not provide any criterion for choosing the weights in Eq. (5). In particular, it does not make explicit reference to the location of the data.

In geostatistics, it is well known that, provided the covariance function is known, the Best Linear Unbiased Estimator of the mean vector is the generalized-least-squares (gls) estimator (Cressie, 1993). It is also referred to as “kriging of the mean” in Wackernagel (2013) and Chilès and Delfiner (2012), a term that will be used in the rest of this work. In the following, “kriging weights” is a (slight) abuse of language to refer to the weight of the gls estimator of the mean. The fact that the kriging weights when estimating the mean must be identical for all variables under axioms C1-C3 has deep consequences on the covariance modeling of spatial compositional data, which are now examined in detail. The usual geostatistical setting is considered. The compositional data  $(\mathbf{x}_1, \dots, \mathbf{x}_n)$  is a sample of a second order stationary multivariate random field  $\mathcal{X}(\mathbf{s})$  with  $\mathbf{s} \in \mathbb{R}^d$ . For the simplicity of exposition, we consider  $p \geq 3$  and  $n \geq 2$ . Under the assumption of second order stationarity, the multivariate covariance model of  $\mathcal{X}(\mathbf{s})$  is defined by a matrix of functions

$$\begin{pmatrix} C_{11}(\mathbf{h}) & \cdots & C_{1p}(\mathbf{h}) \\ \vdots & \ddots & \vdots \\ C_{p1}(\mathbf{h}) & \cdots & C_{pp}(\mathbf{h}) \end{pmatrix}, \mathbf{h} \in \mathbb{R}^d, \quad (7)$$

which must be positive definite (Chilès and Delfiner, 2012). For a given set of  $n$  locations  $(\mathbf{s}_1, \dots, \mathbf{s}_n)$ , this model induces a  $np \times np$  multivariate covariance block-matrix of the form

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{11} & \cdots & \mathbf{C}_{1p} \\ \vdots & \ddots & \vdots \\ \mathbf{C}_{p1} & \cdots & \mathbf{C}_{pp} \end{pmatrix}, \quad (8)$$

where each matrix  $\mathbf{C}_{kl}$  is such that its elements are  $[\mathbf{C}_{kl}]_{ij} = C_{kl}(\mathbf{s}_j - \mathbf{s}_i)$ , with  $1 \leq k, l \leq p$  and  $1 \leq i, j \leq n$ .

Kriging weights depend on the multivariate covariance model and on sample locations. For some specific combinations of covariance models and sample locations, some of the kriging weights can be negative. Obviously, for all practical purposes, a basic requirement is that the mean  $\mathbf{M}(\mathbf{x}_1, \dots, \mathbf{x}_n)$  lies in the simplex  $\mathbb{S}^{p-1}$ . Since  $M^k(x_1^k, \dots, x_n^k)$  is the central tendency of

composition  $k$ , it is expected that it lies within the interval defined by the minimum and maximum of the observed values of  $\mathbf{x}^k$ , which is a sufficient condition for  $0 \leq M^k(x_1^k, \dots, x_n^k) \leq 1$ . However, in presence of negative weights it is mathematically not impossible that for some very specific datasets one gets  $M^k(x_1^k, \dots, x_n^k) \notin [0, 1]$ , which is not acceptable in our context. In Section 4.2, we will discuss how the condition  $\mathbf{M}(\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathbb{S}^{p-1}$  can be enforced. At this point, let us simply notice that imposing nonnegative weights is a sufficient condition.

It is well known (and easy to verify) that if the multivariate covariance function belongs to the family of proportional models (i.e., the product of a covariance matrix and a single correlation function), the kriging weights are identical for all variables (Helterbrand and Cressie, 1994). The following Theorem shows that the converse also holds. It is first stated independently of the compositional data setting considered so far. A formal link to the compositional data setting will be made later.

**Theorem 3** *Let us consider a second-order stationarity multivariate random field of dimension  $p \geq 2$ .*

A) *For all  $n \geq 2$  and all  $(\mathbf{s}_1, \dots, \mathbf{s}_n)$ , the co-kriging weights of the  $p$  means are equal for all variables if and only if the multivariate covariance model is a proportional model, i.e., for  $1 \leq k, l \leq p$*

$$C_{kl}(\mathbf{h}) = \sigma_{kl}\rho(\mathbf{h}), \quad \mathbf{h} \in \mathbb{R}^d,$$

*for some  $p \times p$  covariance matrix  $\mathbf{\Sigma} = [\sigma_{kl}]_1^p$  and some correlation function  $\rho(\mathbf{h})$ .*

B) *Moreover, the same result holds if nonnegativity of the kriging weights is imposed.*

The proof of this Theorem is deferred to the Appendix. Particular cases of this model are absence of correlation between variables, with  $\sigma_{kl} = 0$  for all  $k \neq l$ . This particular case must be ruled out when considering compositional data since negative correlation between some variables is implied by the fact that the parts sum to one.

Theorems 2 and 3 considered together imply that, when estimating the central tendency of geostatistical compositional data, the only minimum variance estimator satisfying reflexivity, marginal stability and continuity is the kriging of the mean with identical weights for all compositions, whether or not imposing nonnegative weights to enforce  $\mathbf{M}(\mathbf{x}_1, \dots, \mathbf{x}_n)$  to be in the simplex is needed. Moreover, the only multivariate covariance function model satisfying these requirements is the proportional model

$$\mathbf{C}(\mathbf{h}) = \mathbf{\Sigma} \otimes \rho(\mathbf{h})$$

where  $\rho(\mathbf{h})$  is a correlation function and  $\mathbf{\Sigma}$  a valid covariance matrix for compositional data. This is now formally stated in the following Theorem.

**Theorem 4** Let  $\mathcal{X}(\mathbf{s})$  be a multivariate compositional random field as described above and let  $\mathbf{M} = (M^1, \dots, M^p)$ , with  $\mathbf{M}(\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathbb{S}^{p-1}$ , be a compositional mean such that  $\text{tr}\{\text{Var}(\mathbf{M})\}$  is minimal over  $\mathbb{S}^{p-1}$ . Then, the following propositions are equivalent:

1.  $\mathbf{M}$  satisfies C1-C3.
2. For each variable  $k = 1, \dots, p$ ,  $M^k(\mathbf{x}_1, \dots, \mathbf{x}_n)$  is a linear combination of  $(x_1^k, \dots, x_n^k)$ , i.e.,  $M^k = \sum_{i=1}^n \lambda_i x_i^k$ , where the weights are the same for all variables. The  $n$ -dimensional vector of weights  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)^\top$  is the solution of a unique (possibly constrained) kriging system.
3. The multivariate covariance function model is a proportional model.

**Proof.** The proof consists in collecting the results established in Theorems 2 and 3.

- a) According to Theorem 2, “ $\mathbf{M}$  satisfies C1-C3” is equivalent to “ $M^k(\mathbf{x}_1, \dots, \mathbf{x}_n)$  is a linear combination of  $(x_1^k, \dots, x_n^k)$  only. Moreover, the weights are identical for all variables  $k = 1, \dots, p$ ”. Therefore, imposing  $\text{tr}\{\text{Var}(\mathbf{M})\}$  to be minimal on  $\mathbb{S}^{p-1}$  is equivalent to imposing  $\text{Var}(M^k)$  to be minimal for each  $k = 1, \dots, p$ . Hence, it is equivalent to impose  $M^k(\mathbf{x}_1, \dots, \mathbf{x}_n)$  to be equal to the (possibly constrained) kriging of the mean of  $(x_1^k, \dots, x_n^k)$ , for some unique kriging system. Hence  $1 \Leftrightarrow 2$ .
- b) In Theorem 3 it is proven that  $2 \Leftrightarrow 3$ .

In conclusion, the three statements are equivalent. □

## 4.2 Enforcing the mean to be in the simplex

As already pointed out earlier, in practice the mean  $\mathbf{M}(\mathbf{x}_1, \dots, \mathbf{x}_n)$  obtained by unrestricted kriging of the mean (i.e., without imposing nonnegative weights) will almost always lie in the simplex. Indeed, since the correlation model  $\rho(\mathbf{h})$  has been fitted on the dataset, it is expected that even though some of the weights are negative, these weights will be of small magnitude. Moreover, the combination of a high magnitude negative weight with extreme compositional value, leading to  $M^k(x_1^k, \dots, x_n^k) \notin [0, 1]$ , is very unlikely. In this case, part A) of Theorem 2 and Theorem 4 apply. In the unlikely event where one gets  $M^k(x_1^k, \dots, x_n^k) \notin [0, 1]$  for some composition and some specific dataset, several possibilities are now discussed.

1. In order to force  $\mathbf{M}(\mathbf{x}_1, \dots, \mathbf{x}_n)$  to belong to  $\mathbb{S}^{p-1}$ , Walvoort and de Gruijter (2001) proposed compositional kriging, for which the sum of the  $p$  prediction variances is minimized, subject to unbiasedness, nonnegativity and sum to 1 of the estimated parts.

Unfortunately, this approach leads to weights that i) depend on the observed values  $\mathbf{X}$  and ii) are different for each part  $k = 1, \dots, p$ . The compositional kriging proposed in Walvoort and de Gruijter (2001) is thus incompatible with axiom C2.

2. A sufficient condition is to impose that the weights  $\lambda_1, \dots, \lambda_n$  are nonnegative. In this case, Part B) of Theorem 2 and Theorem 4 apply. From a geostatistical point of view, the kriging variance will be higher than that of the unrestricted kriging, but the condition  $\mathbf{M} \in \mathbb{S}^{p-1}$  will be satisfied.
3. However, in our opinion, obtaining  $M^k(x_1^k, \dots, x_n^k) \notin [0, 1]$  for some composition  $k$  can be interpreted as a lack of fit between the model and that specific dataset. Finding a more adequate model for the correlation function  $\rho(\mathbf{h})$  will fix the problem. There are several possibilities: decreasing the regularity of  $\rho(\mathbf{h})$  at the origin (e.g. from quadratic to linear behavior), decreasing its range, adding a small nugget effect. Of course, any combination of these is also possible.

## 5 Discussion

In this work several original results are established:

- Firstly, it is shown that for compositional data the only means that simultaneously satisfy reflexivity, continuity and marginal stability are weighted arithmetic means, to which kriging belongs. The simultaneous requirement of marginal stability and the sum to one is the main reason leading to this result.
- Secondly, a very general result in multivariate geostatistics is established. It is shown that the only multivariate covariance model for which the kriging weights are identical for all components is the proportional model. To the best of our knowledge, this result has not been shown earlier. It has consequences for the modeling of compositional data, but it also is of interest on its own. It is worth recalling here that Bogaert (2002) showed that, for a categorical random field, the only valid multivariate covariance model for indicator co-kriging is also the proportional model. Any other form of the linear model of corregionalization is not permissible.
- Thirdly, the combination of these two results is that, within a geostatistical setting in the Euclidean space of the raw data, the only covariance model leading to kriging that satisfies simultaneously reflexivity and marginal stability is the proportional model.

In summary, when performing the statistical analysis of compositional data in the standard Euclidean geometry of real space it is impossible, at the same time, to satisfy axioms C1-C3 and to consider complex modeling using log-ratio transformations, complex covariance

models, or both. It is a kind of impossibility result, in the spirit of Arrow’s impossibility theorem (Arrow, 1950). This result might be perceived as disappointing. Depending on the problem at hand, one either has to relax marginal stability or to restrict the modeling to the proportional covariance model and kriging of the mean on raw data. This is a restrictive model and chances are that the covariance structure inferred from the data does not follow this form. The scientist is in an “either/or” situation. On the one hand, imposing C2 when a more complex model is true corresponds to forcing an inefficient estimate of the mean. On the other hand, using a more complex model of covariance than the proportional model implies that marginal stability is lost. This might lead to counter-intuitive behaviors, similar to those shown in the Introduction section. The agreement or disagreement of conclusions derived from axioms with empirical results can provide a first preliminary assessment of the appropriateness of the axioms. In this respect, when not imposing marginal stability, the two examples shown in the Introduction can lead to results in contradiction with common sense and, perhaps, with the ultimate use of the estimates.

This work was intended to provide new mathematical results, and to re-open research directions. Our work is based on an axiomatic approach. Obviously, other approaches are possible. This work must be considered as complementary to the usual model-based approach, based on log-ratios. Instead of starting from a model and exploring its properties, we start from a set of properties (the axioms) and derive the class of models satisfying this set of axioms. Axioms are sets of properties we choose, and this choice can be discussed. In Egozcue and Pawlowski-Glahn (2011), a principle of coherence is chosen, from which the Aitchison geometry follows. In essence, this is also an axiomatic approach. Is the “principle of coherence” a better axiom than marginal stability? This is an interesting debate and an open question that would necessitate to first provide an operative definition of “better”.

Recent work (Scealy and Welsh, 2014, p.167) has shown that “subcompositional coherence [...] is essentially arbitrary and assertions on which it is based are much less natural than has been claimed. [...]”. Moreover, “the subcompositional coherence principle actually excludes all known methods of analysing compositional data, including the log-ratio methods it was intended to privilege.”

Log-ratio analysis remains, among others, a useful tool for the statistical analysis of compositional data for some kind of statistical problems. Usual geostatistical tools such as covariances, variograms and kriging can be redefined in Hilbert spaces defined in the so-called Aitchison simplicial geometry (Pawlowsky-Glahn and Egozcue, 2002; Tolosana-Delgado et al., 2011). An interesting development would be to reformulate an axiomatic approach in the simplex  $\mathbb{S}^{p-1}$  equipped with the isometric log-ratio transform (Egozcue, 2003) and check whether a result similar to Theorem 2 holds.

## Acknowledgements

We are truly indebted to one Advisory Editor and to the Editor-in-Chief for their very constructive comments, which helped to improve this paper.

## References

- Arrow KL (1950) A Difficulty in the Concept of Social Welfare. *J Polit Econ* 58(4):328–346.
- Aczél J (1966) Lectures on functional equations and their applications. Academic Press.
- Aczél J and Dhombres J (1989) Functional equations in several variables. Cambridge University Press
- Aitchison J (1982) The statistical analysis of compositional data. *J Royal Stat Soc Ser B (Stat Methodol)* 44(2):139–177
- Aitchison J. (1986) The statistical analysis of compositional data. Chapman and Hall
- Aitchison J (1989) Measures of location of compositional data sets. *Math Geol* 24(4):365–379
- Billheimer D, Guttorp P and Fagan WF (2001) Statistical Interpretation of Species Composition. *J Am Stat Assoc* 96:1205–1214
- Bogaert P (2002) Spatial prediction of categorical variables: the Bayesian maximum entropy approach. *Stoch Env Res Risk A* 16:425–448
- Cressie N (1993) Statistics for Spatial Data, Revised Edition. Wiley
- Chilès JP and Delfiner P (2012) Geostatistics: modeling spatial uncertainty; Second Edition. John Wiley & Sons
- Egozcue JJ, Pawłowsky-Glahn V, Figureas GM and Barceló-Vidal C (2003) Isometric Logratio Transformations for Compositional Data Analysis. *Math Geol* 35(3):279–300
- Egozcue JJ and Pawłowsky-Glahn V (2011) Basic Concepts and Procedures. In *Compositional Data Analysis: Theory and Applications*, Eds. Pawłowsky-Glahn V, Buccianti A. John Wiley & Sons
- Griva I, Nash SG, Sofer A (2008) Linear and Nonlinear Optimization, 2nd ed. SIAM.
- Helterbrand JD and Cressie N (1994) Universal cokriging under intrinsic coregionalization. *Math Geol* 26:205–226

- Kolmogorov A (1930) Sur la notion de la moyenne. *Atti R Accad Naz Lincei Mem Cl Sci Fis Mat Natur Sez* 12:323–343
- Liu RY, Parelius JM and Singh K (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference, (with discussion and a rejoinder by Liu and Singh). *Ann Stat* 27(3):783–858
- Mateu-Figueras G, Pawłowsky-Glahn V and Egozcue JJ (2011) The principle of working on coordinates. In *Compositional Data Analysis: Theory and Applications*, Eds. Pawłowsky-Glahn V, Buccianti A. John Wiley & Sons
- Matkowski J (2010) Generalized weighted quasi-arithmetic means. *Aequ Math* 79:203-212
- Miller WE (2002) Revisiting the Geometry of a Ternary Diagram With the Half-Taxi Metric. *Math Geol* 34(3):275–290
- Pawłowsky-Glahn V and Olea RA (2004) *Geostatistical Analysis of Compositional Data*. Oxford University Press
- Pawłowsky-Glahn V and Egozcue JJ (2002) BLU Estimators and Compositional Data. *Math Geol* 34(3):259–274
- (2014) Scealy JL, and Welsh AH (2014) Colours and cocktails: Compositional data analysis 2013 lancaster lecture. *Aust NZ J Stat*, 56(2):145–169
- Sharp WE (2006) The graph median – a stable alternative measure of central tendency for compositional data sets. *Math Geol*, 38:221–229
- Shurtz RF (2000) Comment on “Logratios and natural laws in compositional data analysis” by J Aitchison. *Math Geol* 32:645-647
- Tolosana-Delgado R, van den Boogaart K and Pawłowsky-Glahn V (2011) Geostatistics for compositions. In *Compositional Data Analysis: Theory and Applications*, Eds. Pawłowsky-Glahn V, Buccianti A. John Wiley & Sons
- Vardi Y, Zhang CH (2000) The multivariate  $L_1$ -median and associated data depth. *Proc Natl Acad of Sc* 97(4):1423–1426
- Wackernagel H (2013) *Multivariate geostatistics: an introduction with applications*. Springer Science & Business Media, 2013.
- Walvoort DJJ and de Gruijter JJ (2001) Compositional Kriging: A Spatial Interpolation Method for Compositional Data. *Math Geol* 33(8): 951–966

Zuo Y, Serfling R (2000). General notions of statistical depth function. *Ann stat* 28(2): 461–482

## Appendix A: Proof of Theorem 2

Our definition of the simplex is a closed set, i.e., some compositions are allowed to be equal to 0. While this might be a problem for the definition of log-ratios, it will not be a problem for us. Clearly the conditions are necessary: it is simple to check that (5) satisfies conditions C1-C3 for part A) and C1-C2 for part B).

We now prove sufficiency. By marginal stability, we have

$$M^k(\mathbf{x}_1, \dots, \mathbf{x}_n) = F_k(x_1^k, \dots, x_n^k)$$

for some function  $F_k : [0, 1]^n \rightarrow \mathbb{R}$ , for any  $k \in \{1, \dots, p\}$ . If we assume continuity, as in part A), then, by the Extreme Value Theorem,  $F_k$  is bounded since  $x_i^k \in [0, 1]$  for all  $i = 1, \dots, n$ . If we assume  $\mathbf{M}(\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathbb{S}^{p-1}$ , as in part B), then  $F_k$  is also bounded. Choose any  $l, l', l'' \in \{1, \dots, p\}$  and suppose  $x_i^k = 0$  for all  $k \neq l, l', l''$  and all  $i \in \{1, \dots, n\}$ . Since  $\sum_{k=1}^p M^k(\mathbf{x}_1, \dots, \mathbf{x}_n) = 1$  and  $x_i^{l''} = 1 - x_i^l - x_i^{l'}$  for all  $i \in \{1, \dots, n\}$ , it is the case that

$$F_l(x_1^l, \dots, x_n^l) + F_{l'}(x_1^{l'}, \dots, x_n^{l'}) + F_{l''}(1 - x_1^l - x_1^{l'}, \dots, 1 - x_n^l - x_n^{l'}) = 1. \quad (9)$$

Let us define the mapping  $G : [0, 1]^n \rightarrow \mathbb{R}$  by  $G(u_1, \dots, u_n) = 1 - F_{l''}(1 - u_1, \dots, 1 - u_n)$ . Equation (9) then becomes

$$F_l(x_1^l, \dots, x_n^l) + F_{l'}(x_1^{l'}, \dots, x_n^{l'}) = G(x_1^l + x_1^{l'}, \dots, x_n^l + x_n^{l'}), \quad (10)$$

for all  $x_1^l, x_1^{l'} \in [0, 1]$  such that  $x_1^l + x_1^{l'} \leq 1$ . In particular, it holds for all  $x_1^l, x_1^{l'} \in [0, 1/2]$ . Equation (10) is a generalized Pexider equation and, because  $F_l, F_{l'}$  and  $G$  are bounded, its unique solution is

$$\begin{aligned} F_l(u_1, \dots, u_n) &= \lambda_1 u_1 + \dots + \lambda_n u_n + \gamma_l, & \forall u_1, \dots, u_n \in [0, 1/2], \\ F_{l'}(u_1, \dots, u_n) &= \lambda_1 u_1 + \dots + \lambda_n u_n + \gamma_{l'}, & \forall u_1, \dots, u_n \in [0, 1/2], \\ G(u_1, \dots, u_n) &= \lambda_1 u_1 + \dots + \lambda_n u_n + \gamma_l + \gamma_{l'}, & \forall u_1, \dots, u_n \in [0, 1], \end{aligned}$$

for some real numbers  $\lambda_1, \dots, \lambda_n, \gamma_l$  and  $\gamma_{l'}$  (Aczél, 1966, p.302). The expression for  $G$  yields  $F_{l''}(u_1, \dots, u_n) = 1 - G(1 - u_1, \dots, 1 - u_n) = \lambda_1 u_1 + \dots + \lambda_n u_n + \beta_{l''}$  for all  $u_1, \dots, u_n \in [0, 1]$  and for some real  $\beta_{l''}$ .

Since our choice of components  $l, l'$  and  $l''$  in the above reasoning is arbitrary, we obtain  $M^k(\mathbf{x}_1, \dots, \mathbf{x}_n) = F_k(x_1^k, \dots, x_n^k) = \sum_{i=1}^n \lambda_i x_i^k + \beta_k$ , for all  $k = 1, \dots, p$  and all  $x_1^k, \dots, x_n^k$  in  $[0, 1]$ . By reflexivity,  $F_k(u, \dots, u) = \sum_{i=1}^n \lambda_i u + \beta_k = u$ , for all  $u \in [0, 1]$  and for all  $k = 1, \dots, p$ . This is possible only if  $\beta_k = 0$  for all  $k = 1, \dots, p$  and  $\sum_{i=1}^n \lambda_i = 1$ .  $\square$

## Appendix B: Proof of Theorem 3

In the following,  $\mathbf{1}_n$  denotes a vector of ones of length  $n$ ,  $\mathbf{I}_n$  denotes the identity matrix of dimension  $n$  and  $\mathbf{0}_{p,q}$  denotes a  $p \times q$  matrix of zeros. If  $\mathbf{A}$  is a  $m \times n$  matrix and  $\mathbf{B}$  is a  $p \times q$  matrix, the Kronecker product  $\mathbf{A} \otimes \mathbf{B}$  is the  $mp \times nq$  block matrix

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{pmatrix}.$$

In particular, the matrix  $\mathbf{J} = \mathbf{I}_p \otimes \mathbf{1}_n$  is the  $np \times p$  matrix

$$\begin{pmatrix} \mathbf{1}_n & \cdots & \mathbf{0}_{n,1} \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{n,1} & \cdots & \mathbf{1}_n \end{pmatrix}.$$

A) We first consider unconstrained kriging.

For each variable  $k = 1, \dots, p$ , the kriging of the mean,  $\hat{m}_k$ , is a linear combination of the data

$$\hat{m}_k = \sum_{l=1}^p \mathbf{X}_l^\top \boldsymbol{\lambda}_l^k,$$

where  $\mathbf{X}_l = (x_{1,l}, \dots, x_{n,l})^\top$  and  $\boldsymbol{\lambda}_l^k = (\lambda_{1,l}^k, \dots, \lambda_{n,l}^k)^\top$ . Unbiasedness conditions impose

$$\mathbf{1}_n^\top \boldsymbol{\lambda}_k^k = 1 \quad \text{and} \quad \mathbf{1}_n^\top \boldsymbol{\lambda}_l^k = 0, \quad \text{for } l \neq k.$$

Let  $\boldsymbol{\lambda}^k$  be the stacked  $np$ -vector  $\boldsymbol{\lambda}^k = ((\boldsymbol{\lambda}_1^k)^\top, \dots, (\boldsymbol{\lambda}_p^k)^\top)^\top$  and let  $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}^1, \dots, \boldsymbol{\lambda}^p)$  be the  $(np \times p)$  matrix of kriging weights. When solved simultaneously, the  $p$  kriging equations are, in matrix notation,

$$\begin{pmatrix} \mathbf{C} & \mathbf{J} \\ \mathbf{J}^\top & \mathbf{0}_{p,p} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Lambda} \\ -\boldsymbol{\mu} \end{pmatrix} = \begin{pmatrix} \mathbf{0}_{np,p} \\ \mathbf{I}_p \end{pmatrix}, \quad (11)$$

where  $\boldsymbol{\mu}$  is the  $p \times p$  matrix of Lagrange multipliers. The matrix  $\mathbf{C}$  arises from a valid model of covariance functions. If one excludes multiple values at the same location, it is invertible. The general solution for  $\boldsymbol{\Lambda}$  therefore satisfies

$$\mathbf{C}\boldsymbol{\Lambda} = \mathbf{J}\boldsymbol{\mu}, \quad (12)$$

where  $\boldsymbol{\mu} = (\mathbf{J}^\top \mathbf{C}^{-1} \mathbf{J})^{-1}$ . We will denote  $\mu_{kl}$  its elements, with  $1 \leq k, l \leq p$ .

For each  $k = 1, \dots, p$ , we wish to impose that  $\boldsymbol{\lambda}^k$  is a vector of zeros except at coordinates corresponding to the  $k$ -th variable where the weights are equal to a common vector  $\boldsymbol{\lambda}$ , i.e.,  $\boldsymbol{\lambda}^k = (\mathbf{0}_{1,n(k-1)}, \boldsymbol{\lambda}^\top, \dots, \mathbf{0}_{1,n(p-k)})^\top$ . Hence,  $\boldsymbol{\Lambda} = \mathbf{I}_p \otimes \boldsymbol{\lambda}$ . Thus, Eq. (12) becomes

$$\mathbf{C}(\mathbf{I}_p \otimes \boldsymbol{\lambda}) = \mathbf{J}\boldsymbol{\mu}, \quad (13)$$

Then, Eq. (13) is equivalent to

$$\mathbf{C}_{kl}\boldsymbol{\lambda} = \mu_{kl}\mathbf{1}_n, \quad \forall 1 \leq k, l \leq p. \quad (14)$$

Since  $\mathbf{C}$  is invertible,  $\mathbf{C}_{kk}$  is invertible and  $\mu_{kk} \neq 0$ , for all  $k = 1, \dots, p$ . Hence, plugging  $\mu_{kk}\mathbf{C}_{kk}^{-1}\mathbf{1}_n = \boldsymbol{\lambda}$  into Eq. (14) leads to

$$\frac{\mu_{kk}}{\mu_{ll}}\mathbf{C}_{kk}^{-1}\mathbf{C}_{ll} = \mathbf{I}_n, \quad \forall 1 \leq k, l \leq p.$$

This condition shows that  $\mathbf{C}_{kk} = a_{kk}\mathbf{R}$ , where  $\mathbf{R}$  is a correlation matrix with  $a_{kk} > 0$ . With a similar argument, one can show that  $\mathbf{C}_{kl} = a_{kl}\mathbf{R}$  when  $k \neq l$ . In conclusion, there is a single correlation matrix for describing all direct and cross covariance matrices,

$$\mathbf{C} = \boldsymbol{\Sigma} \otimes \mathbf{R}.$$

In other words, the model is proportional.

- B) We now impose non-negativity of the kriging weights. This constrained kriging is the solution of the quadratic system

$$\begin{aligned} \min_{\boldsymbol{\lambda}} \quad & \sum_{k=1}^p \boldsymbol{\lambda}^\top \mathbf{C}_{kk} \boldsymbol{\lambda} \\ \text{s.t.} \quad & \boldsymbol{\lambda}^\top \mathbf{1}_n = 1 \\ & \boldsymbol{\lambda} \geq \mathbf{0}_n, \end{aligned}$$

where the last inequality must be satisfied componentwise. The Kuhn-Tucker stationary conditions (Griva et al., 2008) corresponding to this system are

$$\begin{aligned} \mathbf{C}_{kk}\boldsymbol{\lambda} - \mu_k\mathbf{1}_n - \boldsymbol{\alpha} &= 0 \quad k = 1, \dots, p \\ \boldsymbol{\lambda}^\top \mathbf{1}_n &= 1 \\ \alpha_i &\geq 0 \quad i = 1, \dots, n \\ \alpha_i \lambda_i &= 0 \quad i = 1, \dots, n, \end{aligned}$$

where  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)^\top$  and  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)^\top$  are the  $n + p$  Lagrange multipliers. A non-negativity constraint is said to be active if  $\alpha_i > 0$  and non active when  $\alpha_i = 0$ .

The Lagrange multipliers  $\boldsymbol{\alpha}$ , and hence the set of active constraints, are identical for all  $k = 1, \dots, p$ . Let us re-order the dataset such that the first  $m$  elements, with  $1 < m \leq n$  correspond to inactive constraints, i.e.  $\alpha_i = 0$  and  $\lambda_i > 0$  for  $i = 1, \dots, m$ . Let us denote  $\boldsymbol{\lambda}_m$  the corresponding vector of non-null kriging weights and  $\mathbf{C}_{kk}^{m,m}$  the corresponding  $m \times m$  matrix with the  $m$  first rows and columns of  $\mathbf{C}_{kk}$ . Let us also denote  $\mathbf{C}_{kk}^{m,n-m}$  the  $m \times (n-m)$  matrix with elements from the  $m$  first rows and  $(n-m)$  last columns. Then,  $\boldsymbol{\lambda}_m$  is solution of the kriging system

$$\mathbf{C}_{kk}^{m,m} \boldsymbol{\lambda}_m - \mu_k \mathbf{1}_m = 0 \quad k = 1, \dots, p \quad (15)$$

$$\boldsymbol{\lambda}_m^\top \mathbf{1}_m = 1 \quad (16)$$

$$\begin{aligned} \boldsymbol{\lambda}_m \mathbf{C}_{kk}^{m,n-m} - \mu_k \mathbf{1}_{n-m} - \boldsymbol{\alpha}_{n-m} &= 0 \quad k = 1, \dots, p \\ \alpha_i &\geq 0 \quad i = m+1, \dots, n. \end{aligned}$$

Following arguments similar to those in part A), Eqs. (15) and (16) induce that all covariance matrices  $\mathbf{C}_{kk}^{m,m}$  must be proportional to each other, i.e.  $\mathbf{C}_{kk}^{m,m} = a_{kk} \mathbf{R}^{m,m}$ . This condition is satisfied for all  $m$  if and only if  $\mathbf{C}_{kk} = a_{kk} \mathbf{R}$ , i.e. if and only if  $\mathbf{C} = \boldsymbol{\Sigma} \otimes \mathbf{R}$ .

□