Math Geosci (2019) 51:485–526
https://doi.org/10.1007/s11004-018-9769-3

TEACHING AID

Geostatistics for Compositional Data: An Overview

Raimon Tolosana-Delgado[1] · **Ute Mueller**[2] · **K. Gerald van den Boogaart**[1,3]

Received: 26 December 2017 / Accepted: 12 October 2018 / Published online: 19 November 2018

Abstract This paper presents an overview of results for the geostatistical analysis of collocated multivariate data sets, whose variables form a composition, where the components represent the relative importance of the parts forming a whole. Such data sets occur most often in mining, hydrogeochemistry and soil science, but the results gathered here are relevant for any regionalised compositional data set. The paper covers the basic definitions, the analysis of the spatial codependence between components, mapping methods of cokriging and cosimulation honoring compositional constraints, the role of pre- and post-transformations such as log-ratios or multivariate normal score transforms, and block-support upscaling. The main result is that multivariate geostatistical techniques can and should be performed on log-ratio scores, in which case the system data-variograms-cokriging/cosimulation is intrinsically consistent, delivering the same results regardless of which log-ratio transformation was used to represent them. Proofs of all statements are included in an appendix.

Keywords Simplex · Variogram · Cokriging · Co-simulation

✉ Raimon Tolosana-Delgado
r.tolosana@hzdr.de

Ute Mueller
u.mueller@ecu.edu.au

K. Gerald van den Boogaart
boogaart@hzdr.de

1 Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology, Chemnitzer Str. 40, 09599 Freiberg, Germany

2 School of Science, Edith Cowan University, Joondalup, WA, Australia

3 Institut für Stochastik, Technische Universität "Bergakademie" Freiberg, Freiberg, Germany

1 Introduction

Geostatistics is the name commonly given to a set of tools and concepts for modeling data exhibiting codependence related to the spatial proximity of their sampling locations. Since the founding works of Matheron (1963, 1965, 1971), geostatistics has become a standard method in many fields of geosciences (Cressie 1991; Chilès and Delfiner 1999; Journel and Huijbregts 1978; Wackernagel 2003). Two-point geostatistics was based on the concept of a covariance/correlation function. A covariance (correlation) function describes the evolution of the covariance (correlation) exhibited by two variables geographically a lag distance h apart, as a function of multiples of this lag. Under reasonable assumptions (ergodicity, spatial stationarity, some sort of anisotropy and Gaussianity, eventually after a point-wise transformation), such a function can be estimated and modelled from geo-referenced data sets and further used to obtain maps (interpolations or simulations) of the studied variables.

A multivariate observation is compositional if it is formed by several variables that jointly describe the relative weight, importance or influence with respect to a whole. It is well-known in the geosciences (Chayes 1960) that compositional data are affected by the spurious correlation problem, namely that the apparent linear correlation between two components of the system depends on what other components are considered in the system. As a consequence, the common interpretation of correlation is no longer valid, that is, correlation does not represent a valid measure of linear association between two components. This carries over to geo-referenced compositions, as their spatial auto- and cross-correlation functions are as spurious as correlation coefficients are for non-regionalized compositions (Pawlowsky 1984, 1989; Pawlowsky-Glahn et al. 1995).

Attending to certain consistency principles and basic mathematical tractability, Aitchison (1986) proposed to deal with compositional data after a log-ratio transformation. His methodology was first applied to regionalized compositions by Pawlowsky (1986). Several case studies were presented afterwards (Lark and Bishop 2007; Morales Boezio 2010; Morales Boezio et al. 2012; Pawlowsky 1989; Pawlowsky-Glahn and Burger 1992; Pawlowsky-Glahn et al. 1995; Sun et al. 2014; van den Boogaart and Tolosana-Delgado 2013; Ward and Mueller 2012, 2013); however, the geostatistical community has not yet accepted log-ratio methods, even after the appearance of a monograph (Pawlowsky-Glahn and Olea 2004), further theoretical developments (Pawlowsky-Glahn 2003; Tolosana-Delgado 2006; Tolosana-Delgado and van den Boogaart 2013) and some practical guides (Pawlowsky-Glahn and Egozcue 2016; Tolosana-Delgado et al. 2011). In addition to practical reasons (lack of implementation of the proposed methods in commercial software), some theoretical concerns might also explain that reluctance: the possible effect of the particular log-ratio transformation chosen, the alleged bias of estimates obtained, the supposed lack of upscaling models and block kriging methods, and the widespread tradition of applying a Gaussian anamorphosis (or normal scores transform) believed to magically solve all problems of non-normal data. This contribution presents a comprehensive discussion of the fundamentals of log-ratio methods for geostatistical applications, attending to each one of these issues. An illustration case study is provided throughout the text to illustrate the concepts, methods and tools together with their theoretical discussion.

Section 2 introduces this illustration case study. Section 3 reviews the grounding ideas, principles, concepts and operations of compositional data analysis relevant for geostatistical applications. Section 4 summarizes the common two-point geostatistical practice and existing alternative methods for the spatial modeling of compositional data, particularly focusing on their problems and caveats. Section 5 introduces several ways to specify the spatial dependence between the components of a regionalized composition and the links between them. Section 6 presents some considerations on cokriging and cross-validation, while Sect. 7 discusses the usage of the conditional distribution for cosimulation and estimation of non-linear quantities, such as block cokriging for compositional data. Proofs of the relevant statements are referenced to previous works or included in the first appendix. A second appendix includes a concise, practical workflow.

2 Illustration: The K-Pit Data Set

The K-Pit is a high-grade iron ore deposit (Angerer and Hagemann 2010) of banded iron formation (BIF) type, in the Archean Koolyanobbing Greenstone Belt, Western Australia. The deposit is located 360 km east of Perth in the Southern Cross Province of the Yilgarn Craton. The greenstone belt strikes northwest and is approximately 35 km long and 8 km wide. It is composed of folded sequences of amphibolites, meta-komatiites and intercalated metamorphosed BIF (Griffin 1981). This particular deposit occurs where the main BIF horizon is offset by younger striking faults (Angerer and Hagemann 2010). It consists of several mineralogical-textural types, including hard high-grade magnetite, hematite and goethite ores and medium-grade fault-controlled hematite-quartz breccia, as well as hematite-magnetite BIF.

Three domains, 202 (west main domain), 203 (east main domain) and 300 (hematite hanging wall), were selected as they can be considered reasonably homogeneous from a mineralogical point of view: the iron-rich facies is dominated by hematite in all of them, with minor contributions of magnetite or goethite/limonite.

This study focuses on the geostatistical modelling of the percentages of Al_2O_3, Fe, Mn, P and SiO_2 in 2495 samples assayed in drill core section of 2 m length from 226 drill holes within the three zones. Their classical statistics are shown in Table 1.

Table 1 Classical statistics of the K-Pit data set

	Minimum	Maximum	Mean	Variance	Skew
Al_2O_3	0.023	25.000	0.836	2.672	7.347
Fe	18.040	69.521	61.979	20.521	− 3.124
Mn	0.001	1.694	0.054	0.008	8.655
P	0.003	0.663	0.121	0.004	1.219
SiO_2	0.170	46.580	3.582	23.441	3.240
Sum	45.490	82.015	66.572	7.841	0.253

3 Compositional Data Analysis

3.1 The Classical Concept of Composition

A data set is compositional whenever its variables inform of the relative contribution of a set of components forming a whole. "*Relative importances*" cannot be negative numbers. The relative contribution of a component can be captured by the requirement that their total sum is bounded by 1 (or more loosely by a constant κ, such as 100%, 1 or 10^6 ppm), although this is not a necessary condition, as will be reported in the next subsection. A vector of D components $\mathbf{z} = [z_1, z_2, \ldots, z_D]$ is a composition if

$$z_i > 0 \quad \text{and} \quad \sum_{i=1}^{D} z_i = 1. \tag{1}$$

The set $\mathcal{S}^D$ of all points of D-dimensional real space $\mathbb{R}^D$ satisfying these conditions is often called the D-part *simplex*. The sum equality is fulfilled if all components of the whole are being considered, as the sum of contributions of an exhaustive partition must account for "everything".

It is common that only d, $d < D$, components have been analysed or are deemed important in modelling. In this case one speaks of a *subcomposition*, and there are two methods for constructing a composition from the given components. If the sum of the components is considered informative and is always below 1, it is possible to complement the given components with a further component, the fill-up variable

$$z_{d+1} = 1 - \sum_{i=1}^{d} z_i. \tag{2}$$

This option is particularly interesting if the fill-up variable has a physical interpretation, as happens with the K-pit case study, where it represents the totality of elements not assayed. If the total sum of the subcomposition is deemed uninformative, their information relative to the total can be recast by applying the closure operation, denoted as $\mathcal{C}[\cdot]$ and defined as

$$\mathcal{C}[\mathbf{z}] = \frac{1}{\sum_{i=1}^{d} z_i}\mathbf{z}, \tag{3}$$

effectively removing the uninformative variability from the total sum. Building/closing a subcomposition is routinely done in geology when representing three components of a composition in a ternary diagram. In either case, the resulting compositional vectors will be deemed to be elements of a simplex of the appropriate number of parts (D, d or $d + 1$).

In order to work with compositional data, Aitchison (1986) proposed two principles, subcompositional coherence and scale invariance, which should be satisfied by

any meaningful method of analysis to avoid the risks of spurious correlation. Subcompositional coherence implies that results obtained analysing a subcomposition cannot contradict those obtained analysing the whole composition. Scale invariance proposes that results should be the same for a composition $\mathbf{z}$ and any scaled version $p\mathbf{z}$, such as the composition resulting from a change of units from % to ppm. Both conditions imply working with ratios. In Eq. (1) the value zero was explicitly excluded in the definition to conform with the traditional definition by Aitchison (1986). However, zeroes occur quite often in compositional data, sometimes to designate missing values or values below detection limit. There is extensive literature (Aitchison 1986; Pawlowsky-Glahn et al. 2015; van den Boogaart and Tolosana-Delgado 2013) discussing their treatment. The common strategy is to replace these irregular values followed by the application of the spatial techniques described in this paper. The only exception known to the authors is the work by Tjelmeland and Lund (2003) who put forward an integrated approach dealing with zeroes and spatial dependence.

3.2 The Modern Concept of Composition: Equivalence Classes

The principle of scale invariance indeed leads to a more general definition of composition: if two vectors $\mathbf{x}$ and $\mathbf{y}$ of positive components are such that $\mathcal{C}[\mathbf{x}] = \mathcal{C}[\mathbf{y}]$, they are said to be compositionally equivalent, and the class of all vectors $\mathbf{y}$ compositionally equivalent to $\mathbf{x}$ is called a composition (Barceló-Vidal 2003; Barceló-Vidal and Martín-Fernández 2016). In fact, analyzing the original vector $\mathbf{x}$ or its closure $\mathcal{C}[\mathbf{x}]$ should give the same results, being compositionally equivalent. This concept of a composition may become relevant, for instance, in the study of geochemical surveys with partial leachates or of vegetal matrices, because the total sum may then be related to sample preparation issues or of the age of the plants.

Sometimes, compositional data are expressed in non-dimensionless units (mg/L, mol/Kg, meq/L), which do not appear to satisfy either of the two definitions given, like for instance in groundwater reservoir quality studies. In those cases it should be still possible to conceive a meaningful change to dimensionless units (mostly kg or mol of solute per kg or mol of solvent) in which one could decide which of the cases in Sect. 3.1 apply. In many of these cases (van den Boogaart and Tolosana-Delgado 2013), the theory of compositional analysis ensures that final results with dimensionless units or with the original units are mutually consistent.

3.3 Geometry

The simplex $\mathcal{S}^D$, the sample space of compositions, can be given an inner product space structure (Pawlowsky-Glahn and Egozcue 2001). The operations addition and subtraction, as well as scalar multiplication in the underlying vector space structure are given by the operations of perturbation ($\oplus$, $\ominus$; Aitchison 1982) and powering ($\odot$; Aitchison 1986), defined respectively as

$$\begin{aligned}\mathbf{a} \oplus \mathbf{b} &= \mathcal{C}[a_1 b_1, a_2 b_2, \ldots, a_D b_D],\\ \mathbf{a} \ominus \mathbf{b} &= \mathcal{C}[a_1/b_1, a_2/b_2, \ldots, a_D/b_D],\\ \lambda \odot \mathbf{a} &= \mathcal{C}[a_1^{\lambda}, a_2^{\lambda}, \ldots, a_D^{\lambda}].\end{aligned} \tag{4}$$

An inner product (Pawlowsky-Glahn and Egozcue 2001) on this vector space is given by

$$\langle \mathbf{a}, \mathbf{b} \rangle_A = \frac{1}{D} \sum_{i>j}^{D} \ln \frac{a_i}{a_j} \ln \frac{b_i}{b_j}. \tag{5}$$

The inner product induces a metric, called the Aitchison distance (Aitchison et al. 2000)

$$d_A(\mathbf{a}, \mathbf{b}) = \sqrt{\langle \mathbf{a} \ominus \mathbf{b}, \mathbf{a} \ominus \mathbf{b} \rangle_A} = \sqrt{\frac{1}{D} \sum_{i>j}^{D} \left(\ln \frac{a_i}{a_j} - \ln \frac{b_i}{b_j} \right)^2}. \tag{6}$$

In his seminal works about compositional data analysis, Aitchison (1982, 1986) defined the additive log-ratio transformation (alr), the centered log-ratio transformation (clr), and their inverses, respectively, as

$$\mathbf{y}^a = \operatorname{alr}(\mathbf{z}) = \ln \frac{\mathbf{z}_{-D}}{z_D} = \mathbf{J} \cdot \ln \mathbf{z}, \;\; \operatorname{alr}^{-1}(\mathbf{y}^a) = \mathcal{C}\left[\exp[\mathbf{y}^a; 0]\right], \tag{7}$$

$$\mathbf{y}^c = \operatorname{clr}(\mathbf{z}) = \ln \frac{\mathbf{z}}{\sqrt[D]{z_1 z_2 \cdots z_D}} = \mathbf{H} \cdot \ln \mathbf{z}, \;\; \operatorname{clr}^{-1}(\mathbf{y}^c) = \mathcal{C}\left[\exp(\mathbf{y}^c)\right]. \tag{8}$$

Here $\mathbf{z}_{-D}$ is the original composition without the last component, $[\mathbf{y}^a; 0]$ is a vector expanded with a zero component, logarithms and exponentials are applied componentwise, and

$$\begin{aligned}\mathbf{J} &= [\mathbf{I}_{(D-1)\times(D-1)}; -\mathbf{1}_{D-1}],\\ \mathbf{H} &= \mathbf{I}_{D\times D} - \frac{1}{D}\mathbf{1}_{D\times D}.\end{aligned} \tag{9}$$

In these and the following expressions, subscripts of a matrix or vector indicate their size, $\mathbf{I}$ is the identity matrix and $\mathbf{1}$ denotes a vector all of whose entries are 1. Note that the row and column sums of $\mathbf{H}$ are equal to 0, which makes the clr coefficients sum to 0 too.

With the introduction of the inner product space structure, it was recognized that alr-scores are nothing other but coordinates of compositions with respect to a non-orthogonal basis, and clr-scores are coefficients in a generating system of $\mathcal{S}^D$. In general terms, a composition and its coordinates in any basis are related by a transformation $\psi(\cdot)$

$$\mathbf{y} = \psi(\mathbf{z}) = \boldsymbol{\Psi} \cdot \ln \mathbf{z}, \quad \mathbf{z} = \psi^{-1}(\mathbf{y}) = \mathcal{C}\left[\exp(\boldsymbol{\Psi}^{-} \cdot \mathbf{y})\right], \tag{10}$$

where $\boldsymbol{\Psi}$ is a full-rank $(D-1) \times D$-matrix whose rows sum to 0

$$\boldsymbol{\Psi} \cdot \mathbf{1}_D = \mathbf{0}_{D-1}, \tag{11}$$

and $\boldsymbol{\Psi}^-$ is its Moore–Penrose generalized inverse. Note that the matrix $\mathbf{J}$ satisfies these conditions.

Based on this observation, Egozcue et al. (2003) proposed using orthonormal coordinate systems, to avoid having to work with generalized inverses. If the coordinate system is orthonormal, then the matrix $\boldsymbol{\Psi} \equiv \boldsymbol{\Phi}$ satisfies

$$\boldsymbol{\Phi} \cdot \boldsymbol{\Phi}^t = \mathbf{I}_{(D-1)\times(D-1)}, \quad \boldsymbol{\Phi}^t \cdot \boldsymbol{\Phi} = \mathbf{H}. \tag{12}$$

It follows that $\boldsymbol{\Phi}^- = \boldsymbol{\Phi}^t$, and the transformation is called the isometric log-ratio transformation (Egozcue et al. 2003). Readers interested in details about these operations and transformations are referred to Tolosana-Delgado et al. (2005) where some simple cases are presented.

Seen as a mapping between spaces, each log-ratio transformation has a geometric interpretation. The component-wise logarithm $\ln(\cdot)$ is a transformation from $\mathbb{R}^D_+$ onto $\mathbb{R}^D$. The $\mathrm{clr}(\cdot)$ maps a composition onto the hyperplane $\mathbb{H} \subset \mathbb{R}^D$ orthogonal to the vector $\mathbf{1}_D$. The matrix $\mathbf{H}$ is thus the orthogonal projection onto this hyperplane. Finally, each full-rank log-ratio transformation is a function from $\mathcal{S}^D \subset \mathbb{R}^D_+$ onto $\mathbb{R}^{D-1}$. The properties of Eq. (12) indicate that, in the case of an ilr, the matrix $\boldsymbol{\Phi}$ is an isometry on $\mathbb{H}$ (van den Boogaart and Tolosana-Delgado 2013).

3.4 Compositional Moments

In what follows, let $\mathbf{Z} = [Z_1, Z_2, \ldots, Z_D]$ denote a random composition. The appendix contains expressions for the empirical counterparts of these results, as well as the proofs of all statements of this section.

Classical first- and second-order moments can be defined through the clr or any other log-ratio transformation. If superscripts indicate the transformation used, mean and variance are given by

$$\begin{aligned}
\boldsymbol{\mu}^c &= \mathrm{E}\left[\mathbf{H} \cdot \ln(\mathbf{Z})\right], \quad \boldsymbol{\Sigma}^c = \mathrm{Var}\left[\mathbf{H} \cdot \ln(\mathbf{Z})\right], \\
\boldsymbol{\mu}^a &= \mathrm{E}\left[\mathbf{J} \cdot \ln(\mathbf{Z})\right], \quad \boldsymbol{\Sigma}^a = \mathrm{Var}\left[\mathbf{J} \cdot \ln(\mathbf{Z})\right], \\
\boldsymbol{\mu}^\psi &= \mathrm{E}\left[\boldsymbol{\Psi} \cdot \ln(\mathbf{Z})\right], \quad \boldsymbol{\Sigma}^\psi = \mathrm{Var}\left[\boldsymbol{\Psi} \cdot \ln(\mathbf{Z})\right].
\end{aligned}$$

Here, the superscript c stands for clr, a for alr and ψ for an arbitrary log-ratio transformation. Since the matrix $\mathbf{H}$ is a projection (i.e $\mathbf{H}^2 = \mathbf{H}$), the clr-variance matrix $\boldsymbol{\Sigma}^c$ is singular.

As an alternative to a specific log-ratio transformation for descriptive statistics Aitchison (1986) proposed the use of the compositional centre

$$\boldsymbol{\mu}^g = \mathcal{C}(\exp(\mathrm{E}\left[\ln(\mathbf{Z})\right])), \tag{13}$$

and the variation matrix $\mathbf{T} = [t_{ij}]$ with elements

$$t_{ij} = \mathrm{Var}\left[\ln \frac{Z_i}{Z_j}\right] = \mathrm{Var}\left[\ln(Z_i) - \ln(Z_j)\right] = \mathrm{Var}\left[\mathrm{clr}_i(\mathbf{Z}) - \mathrm{clr}_j(\mathbf{Z})\right]. \quad (14)$$

One of the typical concerns of analysts when starting with the log-ratio approach to compositional data is to ensure that statistical results do not depend on the transformation used. In fact, the four means introduced before are compatible, that is (Proposition 3 in the appendix)

$$\mathrm{clr}^{-1}(\boldsymbol{\mu}^c) = \mathrm{alr}^{-1}(\boldsymbol{\mu}^a) = \psi^{-1}(\boldsymbol{\mu}^\psi) = \boldsymbol{\mu}^g \in \mathcal{S}^D. \quad (15)$$

This equivalence holds for any full rank log-ratio transformation $\psi(\cdot)$. Similarly, all spread measures convey exactly the same information, and are related though the fundamental equations

$$\boldsymbol{\Sigma}^\psi = \boldsymbol{\Psi} \cdot \boldsymbol{\Sigma}^c \cdot \boldsymbol{\Psi}^t, \ \boldsymbol{\Sigma}^c = \boldsymbol{\Psi}^- \cdot \boldsymbol{\Sigma}^\psi \cdot \boldsymbol{\Psi}^{-t}, \quad (16)$$

$$\boldsymbol{\Sigma}^\psi = -\frac{1}{2}\boldsymbol{\Psi} \cdot \mathbf{T} \cdot \boldsymbol{\Psi}^t, \ \boldsymbol{\Sigma}^c = -\frac{1}{2}\mathbf{H} \cdot \mathbf{T} \cdot \mathbf{H}, \quad (17)$$

where $\boldsymbol{\Psi}^{-t}$ denotes the transposed Moore–Penrose generalized inverse of $\boldsymbol{\Psi}$. These expressions are valid for any basis, even a non-orthogonal one (e.g. an alr when taking $\boldsymbol{\Psi} = \mathbf{J}$), but at the price of having to deal with a more complex notation. Note also that this identity holds globally, for the entire composition considered, and not for individual parts. These statements are proved in Proposition 4 of the appendix.

3.5 Illustration: Compositional Moments

From a compositional point of view, the statistics of Table 1 are unsatisfactory, because they are not subcompositionally coherent. In the case of the K-pit data set, the total sum of the subcomposition is informative, and so a fill-up variable, called Rest, was introduced. The set of alternative basic compositional statistics and diagrams for the K-pit data set are summarized in Table 2 and Fig. 1, showing a very low variation entry for Fe-Rest, mirrored by a low-variance density plot of ln(Fe/Rest). This is an indication of a high proportionality between both parts, which suggests that the Rest is probably dominated by the oxygen and OH^--groups forming the various Fe-oxides present. In fact, the mass relation 63.2246:34.0875 on the geometric mean is almost equivalent to a 1:2 molar relation Fe:O, exactly of $(63.2246 \times 15.9994):(34.0875 \times 55.845) = 0.53$. The kernel density estimate of ln(Fe/Rest) shows that most of the samples have a Fe:O molar relation between 2:5 and 2:3. Figure 1 shows very strong closure effects in the raw data (e.g. upper triangle scatterplot of Fe vs. SiO_2), but also strong boundaries on the relation between several alr-transformed variables (alr.Al_2O_3-alr.SiO_2, for instance), where the alrs were calculated with respect to the Rest. The data are not normally distributed, as is obvious from Table 1 for the raw data and the final row of Fig. 1 for the alr scores.

Table 2 Compositional statistics of the K-Pit data set: centre (closed geometric mean) and variation matrix

Centre (%)	Al_2O_3	Fe	Mn	P	SiO_2	Rest
	0.4844	63.2246	0.0344	0.1068	2.0623	34.0875
Al_2O_3	0.0000	1.0032	1.1167	1.0516	1.5451	0.8975
Fe	1.0032	0.0000	0.8876	0.3459	1.1731	0.0123
Mn	1.1167	0.8876	0.0000	0.7287	1.8846	0.8002
P	1.0516	0.3459	0.7287	0.0000	1.5046	0.3043
SiO_2	1.5451	1.1731	1.8846	1.5046	0.0000	1.1529
Rest	0.8975	0.0123	0.8002	0.3043	1.1529	0.0000

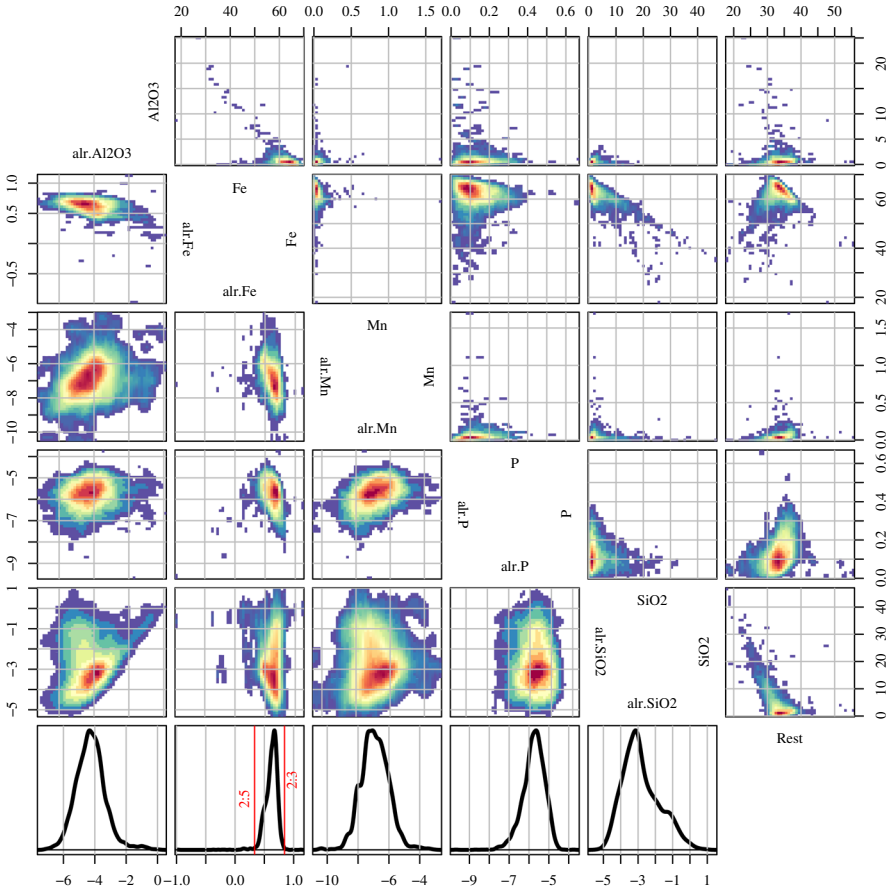

Fig. 1 Scatterplots of the K-Pit data set: raw data (upper triangle) and alr-transformed data with respect to the Rest (lower triangle), the colour scale represents sample density with red being highest. Kernel density estimates of the alr-transformed data are included in the bottom line. Note the strong effects of the closure and the positivity constraints

3.6 The Normal Distribution on the Simplex

As in the case of a real variable or a real vector, it is possible to define a normally distributed random composition characterized by a given compositional center $\mathbf{m}$ and a given covariance matrix $\boldsymbol{\Sigma}$. A random composition is said to follow a normal distribution on the simplex (Mateu-Figueras et al. 2013), denoted as $\mathcal{N}_{\mathcal{S}^D}$, if all of its log-ratio transformed scores follow a joint multivariate normal distribution. Taking a reference full-rank log-ratio transformation $\psi(\cdot)$, the normal distribution on the simplex can be parametrized by a mean vector $\boldsymbol{\mu}^{\psi}$ and a covariance matrix $\boldsymbol{\Sigma}^{\psi}$, so that the probability density is expressed as

$$
\begin{aligned}
f_{\mathbf{Z}}(\mathbf{z}) &= (2\pi)^{-(D-1)/2} |\boldsymbol{\Sigma}^{\psi}|^{-1/2} \exp\left[-\frac{1}{2} d^2_{AM}(\mathbf{z}, \boldsymbol{\mu}^g)\right], \\
d^2_{AM}(\mathbf{z}, \boldsymbol{\mu}^g) &= \psi(\mathbf{z} \ominus \boldsymbol{\mu}^g)^t \cdot [\boldsymbol{\Sigma}^{\psi}]^{-1} \cdot \psi(\mathbf{z} \ominus \boldsymbol{\mu}^g),
\end{aligned} \tag{18}
$$

where $d^2_{AM}(\mathbf{z}, \mathbf{z}')$ denotes the log-ratio Mahalanobis distance (here called *Aitchison–Mahalanobis distance*). Both the Aitchison–Mahalanobis distance and the normal distribution on the simplex are independent of the reference log-ratio transformation chosen, only their analytical expressions are more amenable with them. Propositions 5 and 6 in the appendix give formal proofs of these statements. The normal distribution on the simplex is a fundamental tool in geostatistics for compositional data, needed to define a Gaussian regionalized composition.

4 Conventional Multivariate Geostatistical Practice for Compositions

It is assumed that the reader is familiar with current cokriging (simple, ordinary, universal) and cosimulation techniques, their differences and similarities (Cressie 1991; Chilès and Delfiner 1999; Wackernagel 2003), as well as with the matrix notation introduced by Myers (1982).

The focus of classical geostatistics is to estimate the value of a vector of P space-dependent (real-valued) variables $\mathbf{Z}(x) = [Z_1(x), Z_2(x), \ldots, Z_P(x)]$, as a linear function of neighbouring observations $\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N\}$ with $\mathbf{z}_i = \mathbf{Z}(x_i)$ measured at locations $\{x_1, x_2, \ldots, x_N\}$ within a domain $\mathcal{D}$ of the geographical space. More generally, an estimate at location x_0 is given as

$$
\eta_i(z_i^*(x_0)) = y_i^*(x_0) = \sum_{n=1}^{N} \sum_{j=1}^{P} w_{ij}^n \eta_j(z_j(x_n)) = \sum_{n=1}^{N} \sum_{j=1}^{P} w_{ij}^n y_j(x_n), \tag{19}
$$

where $\eta_i(\cdot), i = 1, 2, \ldots, P$ can be (point-wise) linear or non-linear functions of the data, such as logarithms for lognormal kriging, or more typically a Gaussian anamorphosis or normal score transformation (Chilès and Delfiner 1999; Rossi and Deutsch 2014) forcing the $y_j(x_n)$ scores to exactly show an empirical normal marginal distribution. Gaussian anamorphosis is particularly critical for simulation, as most algorithms

Table 3 Problems and caveats of geostatistical approaches to regionalized compositional data sets (see text for details)

Method	Transformation	Ensures		Singular
		$y_i^* > 0$	$\sum_i y_i^* = \kappa$	Kriging matrix
Kriging	None	No	No	No
Kriging	Anamorphosis	Yes	No	No
Compositional kriging		Yes	Yes	No
Cokriging	None	No	Yes	Yes
Cokriging	Anamorphosis	Yes	No	No
Cokriging	Log-ratio	Yes	Yes	No

assume the unknown scores given the available data to have a joint conditional normal distribution.

When $\mathbf{Z}(x)$ is a regionalized composition, the (co)kriging systems should yield valid compositions everywhere in the domain, that is, the (co)kriging estimates satisfy the conditions stated in Eq. (1). However (as summarised in Table 3), kriging and cokriging are not convex operators (Chilès and Delfiner 1999), thus nothing forces their results to remain bounded by the data or any constraints on them. This implies that negative predictions are possible, as are total sums larger than 100%, as observed for example in Molayemat et al. (2018), Pawlowsky-Glahn and Olea (2004) or Pawlowsky-Glahn et al. (1995). Simulation algorithms tend to generate even larger violations of these constraints, as they often add some variability around the (co)kriging predictions. Gaussian anamorphosis might ensure positivity if cleverly chosen (i.e., similar to a logarithmic transformation), but it cannot ensure the constant sum constraint. This is ensured by cokriging (or cosimulation) of raw closed data only, if one manages to deal with the singularity of the cokriging system matrix.

Methods based on normal score transforms present a secondary problem: compositional data are multivariate by nature, and some form of multivariate Gaussian anamorphosis is needed. A naive marginal Gaussian anamorphosis of each individual variable (as implemented in most geostatistical software) does not satisfy this condition. In this context, the stepwise conditional transformation of Leuangthong and Deutsch (2003), the projection pursuit multivariate transform (PPMT, Barnett et al. 2014) or the multivariate flow normal transform of van den Boogaart et al. (2017) should be applied, as they are multivariate by construction.

Walwoort and de Gruijter (2001) took the term compositional kriging to describe a cokriging of all components where positivity and constant sum constraints are included in the cokriging as additional constraints. The singularity of the covariance of raw data is dealt with by ignoring cross-covariances, thus neglecting any natural link between variables conveyed by the cross-variograms. The only link preserved by this technique is the naive (and physically irrelevant) constant sum. Moreover, to the authors' knowledge, this method does not deliver a meaningful cokriging covariance matrix, which may explain why no cosimulation based on compositional kriging has been proposed yet.

Finally, a general comment on all approaches based on raw data, on simple logarithmic transformations or on individual Gaussian anamorphosis, even those taking into account the necessary constraints: These methods do not explicitly account for the relative scale of compositional data. Moreover, although in some cases they satisfy numerical constraints, they do not ensure that the problem of spurious spatial covariance is avoided, which affects regionalized compositions analogously to the way in which spurious correlation affects compositions in general (Pawlowsky 1984). To what extent these problems also affect a truly multivariate normal score transformed data set (e.g. one obtained using PPMT Barnett et al. 2014) is still unclear to the authors' knowledge.

5 Compositional Structural Analysis

Pawlowsky-Glahn and Olea (2004) propose several structural functions, which are essentially classical variograms and covariance functions defined either on clr-transformed data, or on alr-transformed data. In this section these definitions are summarized, and adapted to the case of data transformed via an arbitrary log-ratio transformation.

5.1 Structural Functions and Their Relations

Denote by $\mathbf{Y}(x)$ the scores obtained by transforming a regionalized composition $\mathbf{Z}(x)$ through a suitable full-rank log-ratio transformation $\psi(\cdot)$. For any choice of $\psi(\cdot)$, a matrix-valued variogram or covariance function can be defined, as shown in Proposition 9 in the appendix. The six specifications have the following properties (Proposition 7):

1. For a given log-ratio transformation, its variogram and covariance function satisfy

$$\boldsymbol{\Gamma}(h) = \mathbf{C}(0) - \frac{1}{2}(\mathbf{C}(h) + \mathbf{C}(-h));$$

   thus, for a symmetric covariance function, where $\mathbf{C}(h) = \mathbf{C}(-h)$, the log-ratio variogram and the log-ratio covariance function are equivalent, and they also fulfill $\mathbf{C}(h) = \mathbf{C}(0) - \boldsymbol{\Gamma}(h)$. These equivalences are honored whichever log-ratio transformation is used.
2. The covariance functions of ilr and clr transformed data are related by

$$\mathbf{C}^{\phi}(h) = \boldsymbol{\Phi} \cdot \mathbf{C}^{c}(h) \cdot \boldsymbol{\Phi}^{t}, \quad \mathbf{C}^{c}(h) = \boldsymbol{\Phi}^{t} \cdot \mathbf{C}^{\phi}(h) \cdot \boldsymbol{\Phi},$$

   where the superscripts c and ϕ relate to the clr and ilr transformed scores, respectively.
3. Analogously, the corresponding variograms are related by

$$\boldsymbol{\Gamma}^{\phi}(h) = \boldsymbol{\Phi} \cdot \boldsymbol{\Gamma}^{c}(h) \cdot \boldsymbol{\Phi}^{t}, \quad \boldsymbol{\Gamma}^{c}(h) = \boldsymbol{\Phi}^{t} \cdot \boldsymbol{\Gamma}^{\phi}(h) \cdot \boldsymbol{\Phi}. \tag{20}$$

4. Equivalent expressions link clr score variograms and covariance functions with those expressed in any arbitrary full rank log-ratio representation

$$\begin{aligned} \mathbf{C}^{\psi}(h) &= \boldsymbol{\Psi} \cdot \mathbf{C}^{c}(h) \cdot \boldsymbol{\Psi}^{t}, \quad \mathbf{C}^{c}(h) = \boldsymbol{\Psi}^{-} \cdot \mathbf{C}^{\psi}(h) \cdot \boldsymbol{\Psi}^{-t}, \\ \boldsymbol{\Gamma}^{\psi}(h) &= \boldsymbol{\Psi} \cdot \boldsymbol{\Gamma}^{c}(h) \cdot \boldsymbol{\Psi}^{t}, \quad \boldsymbol{\Gamma}^{c}(h) = \boldsymbol{\Psi}^{-} \cdot \boldsymbol{\Gamma}^{\psi}(h) \cdot \boldsymbol{\Psi}^{-t}. \end{aligned} \tag{21}$$

   In particular, these expressions are valid when working with alr scores, by taking $\boldsymbol{\Psi} = \mathbf{J}$. This also applies to the remainder of this section.

The preceding relationships are satisfied both by the theoretical and the empirical versions of these functions. Thus, when fitting models to empirical variogram systems defined in two different log-ratio specifications, it is necessary to force the fitted model to satisfy them too. In this case, the fitted models are said to be mutually compatible. In practice, ensuring compatibility during the modeling process may not be easy, as each variogram or covariance function derived from a particular log-ratio transformation may focus on some specific subcompositional features, while other might appear to mask the variability of that particular feature. It is thus possible that classical automatic fitting processes do not produce compatible models.

To avoid working with a specific transformation and having to check for mutual compatibility, a variogram function can be defined based on the variation matrix. The resulting matrix-valued function is called (pairwise) log-ratio variogram (Pawlowsky-Glahn and Olea 2004), or variation-variogram, and is denoted as the matrix $\mathbf{T}(h) = [t_{ij}(h)]_{i,j=1,\ldots,D}$ with elements

$$t_{ij}(h) = \mathrm{Var}\left[\ln \frac{Z_i(x+h)}{Z_j(x+h)} - \ln \frac{Z_i(x)}{Z_j(x)}\right]. \tag{22}$$

Variation-variograms contain the same information as classical clr or log-ratio variograms, so that the latter can be calculated from the former, with

$$\boldsymbol{\Gamma}^{\psi}(h) = -\frac{1}{2}\boldsymbol{\Psi} \cdot \mathbf{T}(h) \cdot \boldsymbol{\Psi}^{t}, \quad \boldsymbol{\Gamma}^{c}(h) = -\frac{1}{2}\mathbf{H} \cdot \mathbf{T}(h) \cdot \mathbf{H}, \tag{23}$$

and conversely from the clr-variograms

$$t_{ij}(h) = \gamma^{c}_{ii}(h) + \gamma^{c}_{jj}(h) - 2\gamma^{c}_{ij}(h). \tag{24}$$

There is no straightforward formula for conversion from log-ratio-variograms to variation-variograms. Therefore, it is convenient to use Eq. (20), resp. Eq. (21), first and then Eq. (24).

5.2 Variogram Estimation and Modelling

Empirical variograms satisfy Eqs. (20)–(24) for all lags, if they are computed with the same set of data. Fitted models should also satisfy them. In particular, the linear model

of coregionalization (LMC, Wackernagel 2003), being symmetric, can be expressed in any of these specifications

$$\mathbf{C}^{\psi}(h|\boldsymbol{\theta}) = \sum_{k=0}^{K} \mathbf{C}_k^{\psi} \cdot \rho_k(h|\theta_k),\ \mathbf{C}^{c}(h|\boldsymbol{\theta}) = \sum_{k=0}^{K} \mathbf{C}_k^{c} \cdot \rho_k(h|\theta_k),$$
$$\boldsymbol{\Gamma}^{\psi}(h|\boldsymbol{\theta}) = \sum_{k=0}^{K} \mathbf{C}_k^{\psi} \cdot (1-\rho_k(h|\theta_k)),\ \boldsymbol{\Gamma}^{c}(h|\boldsymbol{\theta}) = \sum_{k=0}^{K} \mathbf{C}_k^{c} \cdot (1-\rho_k(h|\theta_k)),$$
$$\mathbf{T}(h|\boldsymbol{\theta}) = \sum_{k=0}^{K} \mathbf{B}_k \cdot (1-\rho_k(h|\theta_k)),$$

where K denotes the number of structures, each with its correlogram $\rho_k(\cdot|\theta_k)$, and $\boldsymbol{\theta}$ denotes the vector of model parameters containing all individual ranges θ_k and sill matrices. These sills can be specified either as $\mathbf{B}_k$ or $\mathbf{C}_k$, as they are equivalent because of

$$\mathbf{C}_k^{\psi} = -\frac{1}{2}\boldsymbol{\Psi} \cdot \mathbf{B}_k \cdot \boldsymbol{\Psi}^t, \qquad \mathbf{C}_k^{c} = -\frac{1}{2}\mathbf{H} \cdot \mathbf{B}_k \cdot \mathbf{H}. \tag{25}$$

A model must be positive definite if specified in terms of $\mathbf{C}^{\psi}(\cdot)$ and positive semi-definite in terms of $\mathbf{C}^{c}$ because the clr-transformed covariances always have at least one zero eigenvalue. Equivalently, a model specification in $\boldsymbol{\Gamma}^{\psi}(\cdot)$ must be conditionally negative definite, and in terms of $\boldsymbol{\Gamma}^{c}(\cdot)$ conditionally negative semi-definite. Thus, a model specified in $\mathbf{T}(\cdot)$ must be conditionally positive semi-definite. For the LMC, these conditions are satisfied if each correlogram $\rho_k(h|\theta_k)$ is a positive definite function, and each matrix $\mathbf{C}_k$ is a valid covariance matrix (Wackernagel 2003) or $\mathbf{B}_k$ is a variation matrix (see Appendix "Spatial Results"), namely a conditionally negative definite matrix with zero diagonal elements. Conditional negative (semi)definiteness is a common term in the geostatistical literature, which means that the matrix must have one and only one positive singular value, the rest being negative or 0. It is well known that a variogram matrix must satisfy this condition.

Reduced-rank submodels of the LMC can also be used. Minimum/maximum autocorrelation factors (MAF) have been applied to compositions after an alr (Morales Boezio 2010; Morales Boezio et al. 2012; Ward and Mueller 2012) or a clr (Mueller and Grunsky 2016) transformation.

Individual variation-variograms can be estimated with any available procedure for estimation of direct variograms. Fitting must nevertheless be done jointly, to ensure that the validity conditions are satisfied. Fitting models for the variation-variograms has certain advantages. Most importantly, it allows using a goodness of fit that maximizes a logarithmic goodness-of-fit criterion (Tolosana-Delgado et al. 2011),

$$\text{gof}(\theta) = \sum_{i=1}^{D}\sum_{j=1}^{i-1}\sum_{n=1}^{N_{ij}} (\ln \hat{t}_{ij}(h_n) - \ln t_{ij}(h_n|\theta))^2, \tag{26}$$

where $\hat{t}_{ij}(h_n)$ is the empirical (i, j)-variation-variogram calculated for a lag distance class h_n, N_{ij} is the number of lags used for that calculation, and $t_{ij}(h_n|\theta)$ is the model evaluated at lag h_n. The choice of Eq. (26) as a fit criterion focuses the fitting procedure on the smaller values of t_{ij}, that is, around the origin and at short ranges. Two arguments support this choice. First, interpolation results are particularly sensitive to this part of the variogram, and they do not really depend on the sill: spending effort on fitting the sill will not translate in more reliable interpolations (Chilès and Delfiner 1999). Second, as sills are variances, their natural spread ranges in orders of magnitudes, or in other words, they should be compared in a relative scale, thus in logarithms: this view is consistent with the multiplicative confidence intervals given for the sill of a variogram (Cressie 1991).

As a second advantage, fitting variation-variograms enables working with data sets with many missing values, partially observed subcompositions and similar irregularities. As each component $t_{ij}(h_n)$ requires only that variables i and j are available, one uses pairwise elimination to calculate $\hat{t}_{ij}(h_n)$, thus using a maximum number of observations for each lag and each pair of variables. In contrast, estimating variograms with ilr- or clr-transformed data requires eliminating much more data: for instance, the clr vector is not available if one single component is missing (i.e., one would need complete row-wise elimination to deal with missing values); and though an ilr can be chosen to maximize the number of computable log-ratio scores, this only works if the missing values occur on a few components (Tolosana-Delgado et al. 2008). One could argue that the set of estimates $\{\hat{t}_{ij}(h)\}$ obtained in this way do not necessarily define a set of valid variation matrices $\{\hat{\mathbf{T}}(h)\}$, being computed from different subsets of the data. This is of no consequence in a spatial structural analysis, because the set $\{\hat{t}_{ij}(h)\}$ is only used to guide the fit of a parametric model $\mathbf{T}(h|\boldsymbol{\theta})$ to the data, which will in any case be obtained with an algorithm embedding the validity constraints on $\mathbf{T}(h|\boldsymbol{\theta})$ explicitly.

For these reasons, a structural analysis of regionalized compositional data in terms of variation-variograms should be desirable. One could use standard estimation algorithms to obtain each individual pairwise log-ratio variogram. Then, a procedure would be required for fitting an LMC using Eq. (26), subject to the condition that each matrix $\mathbf{B}_k$ is a valid variation matrix (zero diagonal, conditionally negative semidefinite). Finally, the fitted model could be compared to the empirical structural functions, either in terms of the variation-variograms or recasting variogram estimates and model to clr-/ilr-variograms or covariances, using Eq. (25). Unfortunately, no state-of-the-art geostatistical software incorporates any algorithm that can be adapted to work this way.

5.3 Illustration: Structural Analysis

The proposed compositional structural tool, the variation-variogram, is currently only available in the R package "compositions" (van den Boogaart et al. 2018). Instead, and in order to stay within the framework of user-friendly geostatistical software, the data were alr-transformed with respect to the Rest, and the resulting scores imported into ISATIS (Geovariances 2017). Direct- and cross-variograms were computed at lag

Table 4 Linear model of coregionalization of the alr-transformed composition

	alr.Al_2O_3	alr.Fe	alr.Mn	alr.P	alr.SiO_2
Nugget					
alr.Al_2O_3	0.1324	− 0.0098	0.1023	0.0415	0.0338
alr.Fe		0.0011	− 0.0059	0.0018	− 0.0029
alr.Mn			0.1267	0.0722	0.0177
alr.P				0.0907	− 0.0072
alr.SiO_2					0.3362
Sph(34.8, 34.8, 34.8)					
alr.Al_2O_3	0.199	− 0.0118	0.0864	0.0586	0.1185
alr.Fe		0.003	− 0.0109	− 0.0147	0.0073
alr.Mn			0.1998	0.042	− 0.0045
alr.P				0.0753	− 0.0409
alr.SiO_2					0.3486
Sph(39.05, 66.94, 11.16)					
alr.Al_2O_3	0.225	− 0.0026	− 0.02	− 0.0042	0.0253
alr.Fe		0.0034	− 0.0033	− 0.0033	− 0.0033
alr.Mn			0.1575	0.0138	− 0.0650
alr.P				0.0554	0.0145
alr.SiO_2					0.1330
Sph($3.3 \cdot 10^2$, $2.0 \cdot 10^2$, ∞)					
alr.Al_2O_3	0.3901	− 0.0147	0.0364	− 0.0602	0.1710
alr.Fe		0.0043	− 0.0059	0.0056	− 0.0134
alr.Mn			0.2806	0.0515	0.0685
alr.P				0.1159	− 0.0389
alr.SiO_2					0.3812

Sph = spherical model of the indicated (X,Y,Z) ranges, relative to reference direction N300°, dip 45° and pitch −14°

spacings of 2 m in the vertical direction and 25 m in the horizontal plane with direction of greatest continuity set to azimuth N300°, dip 45° and pitch −14°. A linear model of coregionalization (LMC) was fitted semi-automatically. The resulting parameters are summarized in Table 4 and the model fit is shown in Fig. 2.

Figures 3 and 4 show both the empirical variograms and the fitted LMC represented in variation-variograms and clr-variograms, respectively obtained with Eqs. (24) and (21). The same equations could be used to derive variograms and cross-variograms in any log-ratio representation, which were not included for space reasons. In all variogram figures, lower triangle plots (blue color) show vertical variograms and cross-variograms, while upper triangle diagrams show (red and black/grey colors) variograms/cross-variograms on the horizontal plane. All refer to the main directions relative to the direction of greatest continuityof anisotropy. The LMC satisfactorily fits the data in all these diagrams, a remarkable achievement given that these alternative log-ratio representations were not used at all during the fitting process.

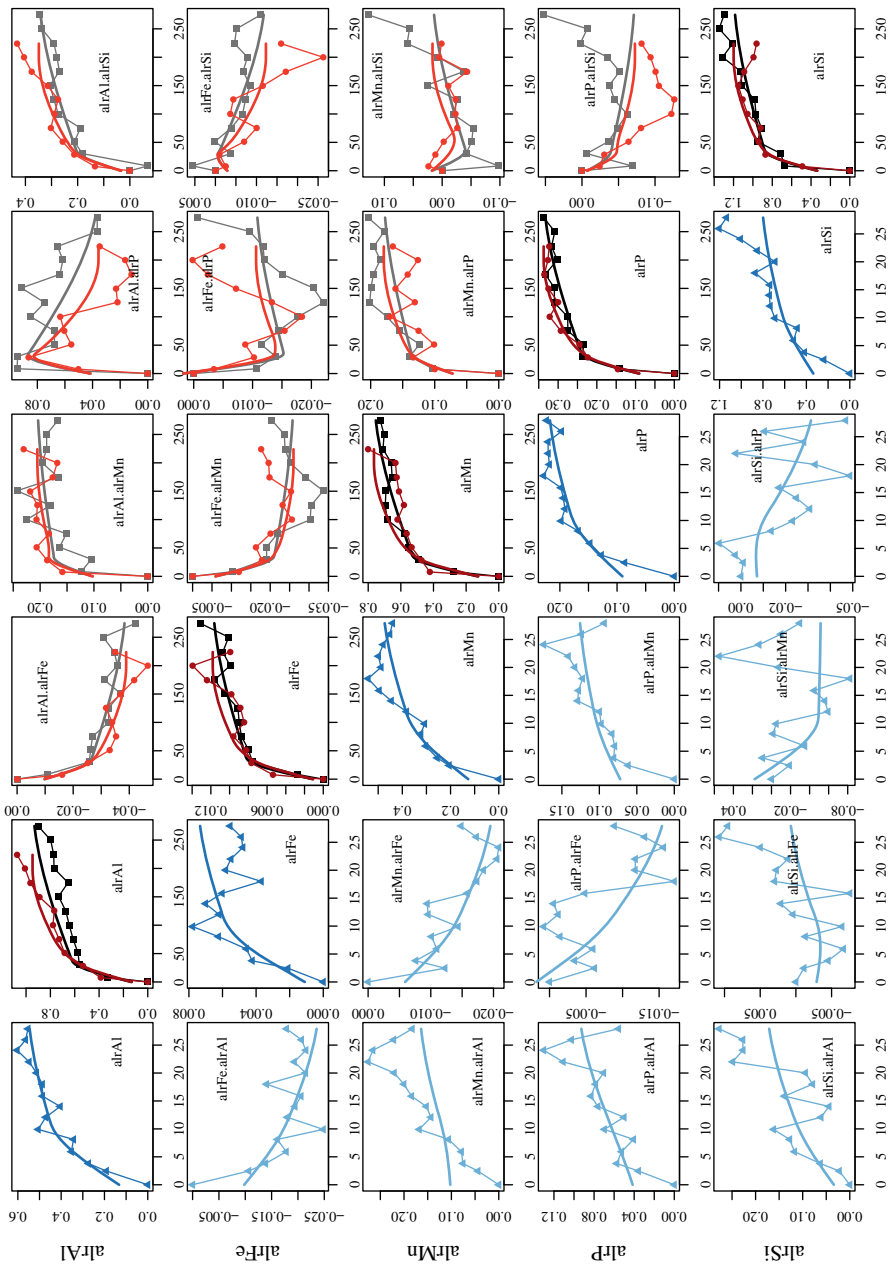

Fig. 2 Empirical alr-variograms and cross-variograms (dots) and fitted LMC (curves): diagonal plots (dark colors) are direct variograms, the lower triangle (blue) shows cross-variograms in the vertical direction, the upper triangle (red, black) shows cross-variograms in horizontal directions (after global rotation)

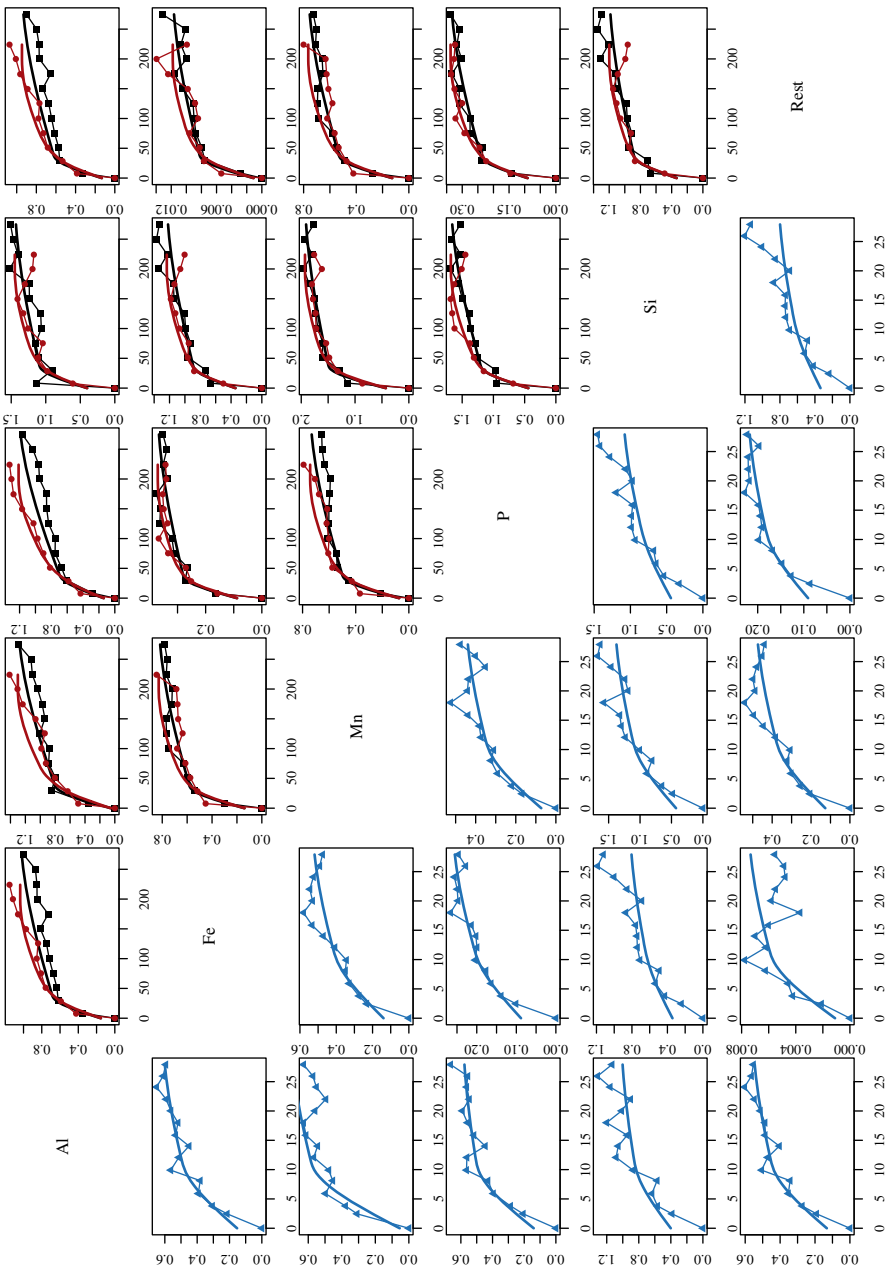

Fig. 3 Empirical variation-variograms (dots) and fitted LMC (curves): the lower triangle (blue) shows variograms for the vertical direction, the upper triangle (red, black) variograms in horizontal directions (after global rotation)

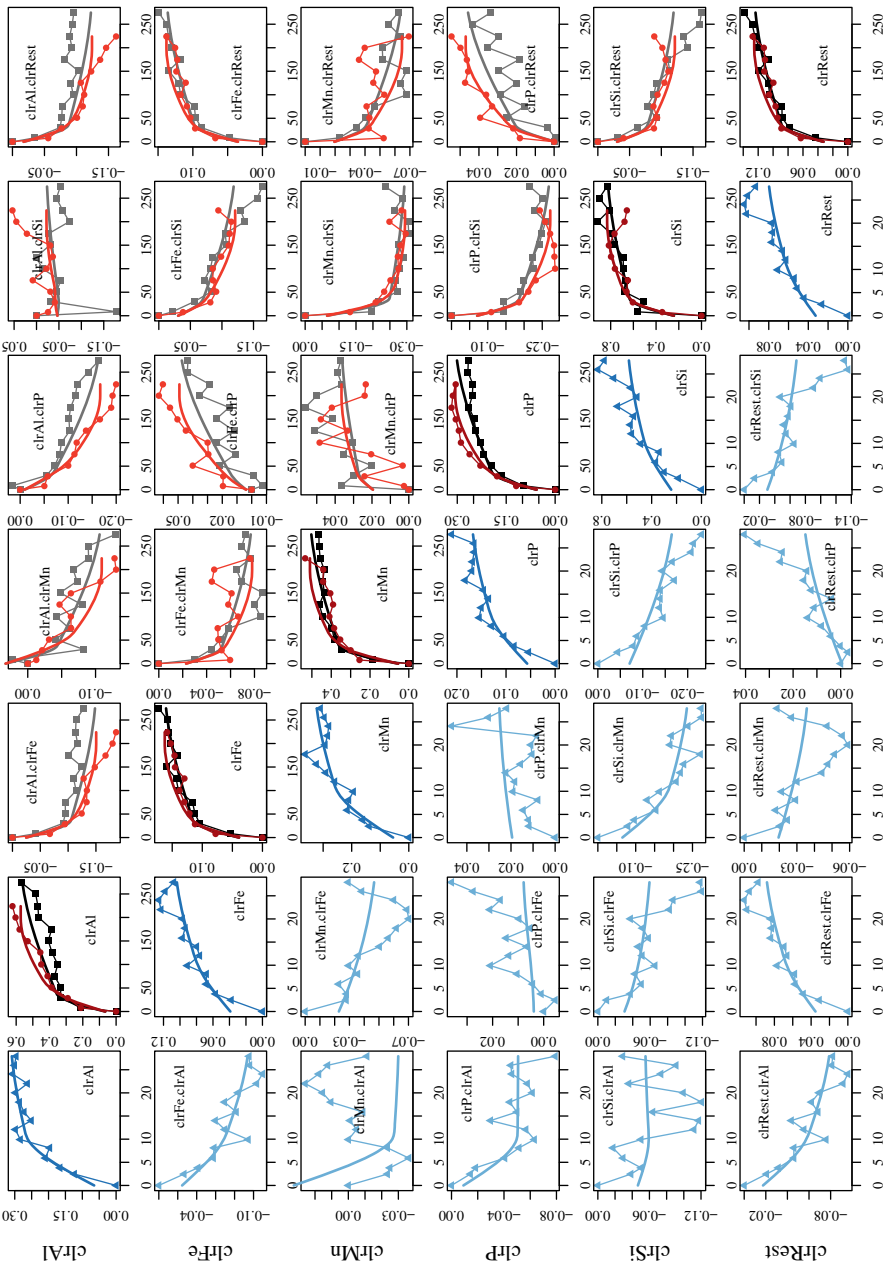

Fig. 4 Empirical clr-variograms and cross-variograms (dots) and fitted LMC (curves): diagonal plots (dark colors) are direct variograms, the lower triangle (blue) shows cross-variograms in the vertical direction, the upper triangle (red, black) show cross-variograms in horizontal directions (after global rotation)

6 Pointwise Estimation

6.1 Cokriging

Once a structural model is available, this can be used in interpolation and simulation procedures, most often ordinary cokriging. This is best done working in a one-to-one set of log-ratios: (a) choosing a particular full-rank log-ratio representation, (b) expressing data and variation-variogram model in this representation by Eqs. (10) and (25) respectively, (c) applying cokriging/cosimulation to the obtained log-ratio scores, and (d) back-transforming the interpolated/cosimulated scores to interpolated/cosimulated compositions through Eq. (10). This procedure ensures that the final outcomes are exactly the same, whichever log-ratio transformation was used (Proposition 9). It is nevertheless recommended to avoid the clr transformation, as clr scores sum to zero, and this collinearity makes the cokriging matrix singular. This could be resolved by Moore–Penrose generalized inversion, if this is available, but producing the same solution at a larger computational cost is seldom a good choice.

6.2 Cross-Validation and Error Assessment

When considering the application of log-ratio methods, practitioners often wonder whether this approach is "better" or "worse" than "classical" methods, and try to answer that by a cross-validation exercise, where a root mean square error (RMSE), a standardized residual sum of squares (STRESS) or a similar measure of difference between observations and independent predictions is drawn (Lark and Bishop 2007; Morales Boezio 2010; Sun et al. 2014; Rossi and Deutsch 2014; Ward and Mueller 2012). It should be noted that this procedure is fair only when comparing methods that minimize the same error measure, it is not appropriate to compare estimates obtained with Eq. (19) by using two different $\eta(\cdot)$ functions, that is: the identity (cokriging raw data) versus a log-ratio transformation (cokriging log-ratio data). If the goodness-of-fit measure is related to a distance $d(\mathbf{y}(x_i), \hat{\mathbf{y}}_i)$, then estimators minimizing its expected value will be favored. Pawlowsky-Glahn and Egozcue (2001) called such estimators metric expectations, and Pawlowsky-Glahn and Egozcue (2002) showed that these coincide with linear estimators in terms of any isometric transformation of the data (isometric with respect to the distance chosen). As a consequence, if the chosen measure of goodness of fit is an absolute difference between estimates and true values, one immediately favors cokriging of raw components. In contrast, by choosing a relative difference (like the Aitchison distance), one favors log-ratio methods. There is thus no metric producing a fair comparison between them.

A second consideration relates to the multivariate nature of compositional data. A compositional goodness-of-fit cross-validation criterion should be multivariate rather than univariate. For instance, one could compare some distance between the observed compositions $\mathbf{y}(x_i)$ and their cross-validation predictions $\hat{\mathbf{y}}_i$, preferentially a Mahalanobis distance (Eq. 18) taking into account the cokriging error covariance (Lark and Bishop 2007).

Connecting these two ideas, it seems reasonable to evaluate compositional methods by a Aitchison–Mahalanobis distance between predictions and values of the log-ratio transformed data (Eq. 18). This distance will deliver the same results, whichever log-ratio transformation is used, as long as it is one-to-one (Filzmoser and Hron 2008; or Proposition 5 in the appendix). This invariance property is not satisfied by the Euclidean distance of the scores, which requires working with isometric transformations (clr or ilr) in order to give rise to the Aitchison distance (Eq. 6).

Classical cross-validation (leave-one-out, n-fold, etc.) typically involves a visual assessment of the adequacy of a series of diagrams, judged by experience. These diagrams are histograms of residuals or standardized residuals, scatter plots of residuals against predictions, and scatterplots of true values against predictions (Bivand et al. 2013). For the purposes of validation with compositional data, all these scatter plots can be obtained for the log-ratio scores, and for the back-transformed components as well, if the input data and the cross-validated values are computed in the same (sub)composition. Histograms and qq-plots, on the other hand, are better obtained only for the log-ratio scores, because the original components will usually have strongly non-normal distributions. Histograms of Mahalanobis norms of residuals may also be useful.

6.3 Illustration: Cokriging and Cross-Validation

A fivefold cross-validation was conducted to assess the adequacy of the LMC of Sect. 5.3. The data set was randomly split into five subsets, and for each subset co-kriging estimates were computed based on the data from the remaining subsets. Cokriging covariance matrices were also produced. Computations were obtained with R. Figure 5 shows the estimates against true values and against residuals in log-ratio scale, while Fig. 6 shows them in raw scale, in the original units. In all cases, a red line shows a reference line, representing the identity line for plots of observations versus predictions, or the horizontal line $y = 0$ for diagrams of residuals versus predictions. All diagrams show a reasonable absence of trends or non-linearity structures, although some residuals in raw scale show a clear heteroskedasticity. The fit of the densities to the reference lines is also very good.

As a multivariate measure of goodness of fit, one can study the Aitchison–Mahalanobis cokriging residual (squared) norms: these are (squared) Aitchison–Mahalanobis distances (Eq. 18) between true values and cokriging estimates, with respect to the cokriging covariance matrix. Given that the dimension of the space is 5, these squared norms should follow a χ^2 distribution with five degrees of freedom. Figure 7 shows this comparison, in the form of a histogram and a qq-plot for a robust subset of the predictions. This selection was necessary because relative residuals are non-robust quantities: to account for this problem, it is common in conventional geostatistical practice to filter standardized residuals larger than 2.5 in absolute value. The outcome shows a reasonable fit, although with a slightly heavier tail.

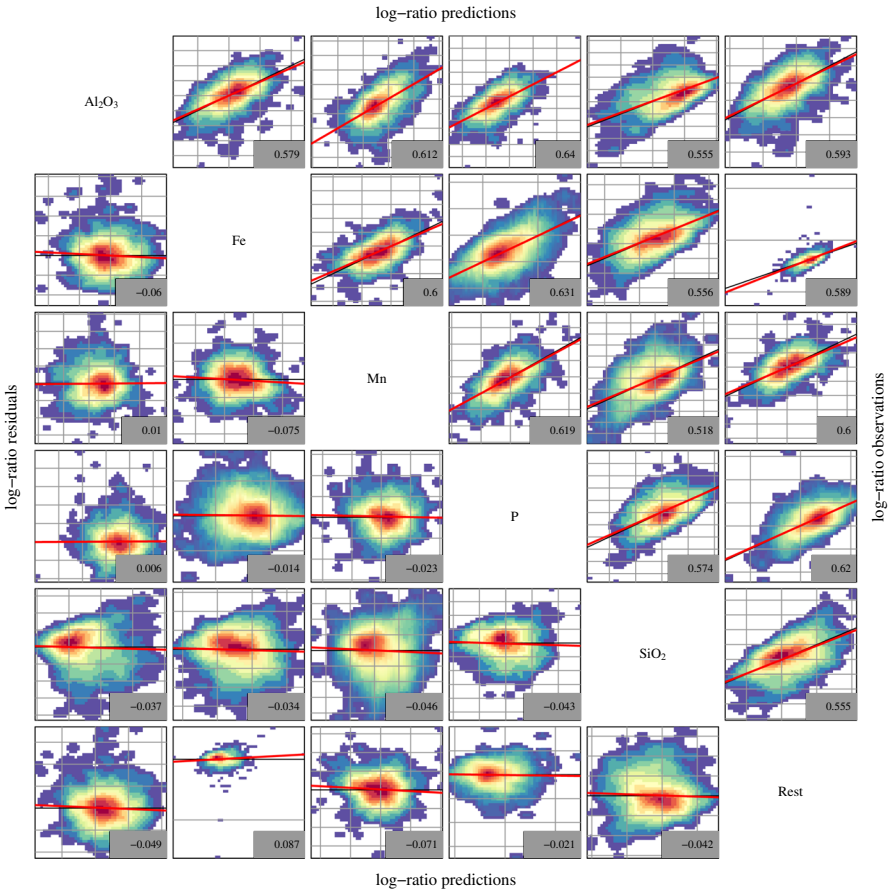

Fig. 5 Scatter density of residuals against predictions (lower triangle) and of true values against predictions (upper triangle) for a fivefold cross-validation. Warmer colors represent higher sample densities. Each diagram represents outcomes for the log-ratio of the row component divided by the column component. Regression lines (red) and correlation coefficients are reported. Reference lines are also shown in black: the $y = 0$ line for the lower triangle plots, the $y = x$ line for the upper triangle plots. For space reasons, no axes are included; a grid of one log-ratio unit spacing is given for reference

7 Simulation and Block-Wise Estimation

7.1 Conditional Joint Distribution

As happens with conventional (real-valued) variables, cokriging of regionalized compositional data sets delivers something more than just an "interpolation". Ordinary cokriging provides the expected value of the log-ratio transformed composition at the predicted location given the surrounding observed compositions that can, by virtue of Eq. (15), be back-transformed to an unbiased estimate of the corresponding conditional expected value of the composition. This unbiasedness is with respect to the Euclidean structure introduced in Sect. 3.3 (Pawlowsky-Glahn and Egozcue 2002). Moreover,

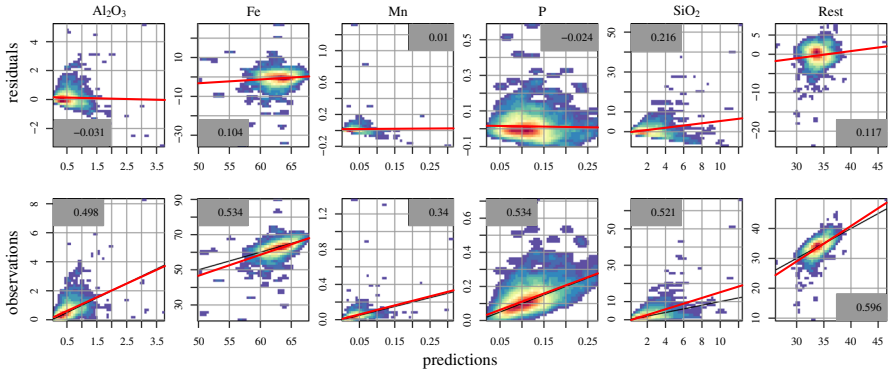

Fig. 6 Scatter density of residuals against predictions (upper row) and of true values against predictions (lower row) for a fivefold cross-validation. Warmer colors represent higher sample densities. Each diagram represents outcomes for one original component, in %. Regression lines (red) and correlation coefficients are reported. Reference lines are also shown in black: the $y = 0$ line for the upper row plots, the $y = x$ line for the lower row plots

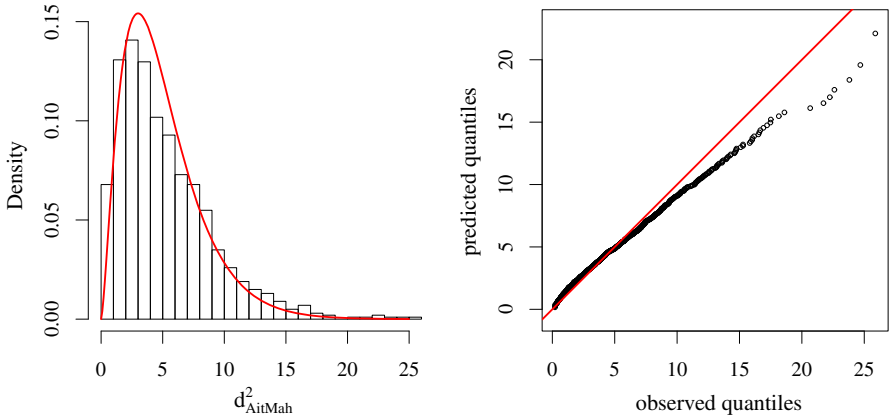

Fig. 7 Comparison of Aitchison–Mahalanobis cokriging residual norms with a χ^2-distribution of five degrees of freedom (in log scale): histogram (left) and qq-plot (right)

under the key assumption of (transformed) Gaussianity discussed in the introduction, simple cokriging delivers the full conditional distribution of the log-ratio transformed composition at the predicted location given the surrounding observed compositions: a multivariate normal distribution with mean equal to the simple cokriging predictor and covariance matrix equal to the cokriging error covariance. But as stated in Sect. 3.6, these two moments of any log-ratio representation completely specify the normal distribution on the simplex that describes the conditional uncertainty on the interpolated composition. Conditional distributions are primarily used for two purposes: cosimulation and estimation of non-linear quantities.

7.2 Cosimulation

Any cosimulation algorithm applied to the log-ratio transformed composition will provide valid simulated log-ratios, that can be back transformed to simulated compositions through Eq. (10). Mueller et al. (2014) show that this procedure adequately reproduces the mean value and the sample distribution of the variables (on both scales: the original proportions and the log-ratio transformed values), as long as the log-ratio transformed data do not depart notably from normality. Reproduction of the variation-variogram is, by construction, as good (or as bad) as the simulation method actually used. These caveats are well-known for simulation procedures of raw data, which often rely on Gaussian anamorphosis to enforce these global statistics reproduction. Recent concerns (Rossi and Deutsch 2014) on this issue for the log-ratio procedure are thus quite partial.

As a matter of good practice, one should avoid any procedure for which final simulated compositions would depend on the log-ratio transformation used. For this reason, separate simulation of each log-ratio variable (with or without a normal score transform) should be avoided. However, a multivariate Gaussian anamorphosis (Barnett et al. 2014) will be necessary when departures from normality on the simplex are notable. The recently introduced flow anamorphosis (van den Boogaart et al. 2017) provides a Gaussian anamorphosis that ensures independence of the results from the log-ratio transformation applied.

7.3 Estimation of Non-linear Quantities

With regard to the estimation of non-linear quantities (non-linear in terms of the log-ratio transformed composition), either Monte Carlo simulation or Gauss–Hermite quadratures have been proposed. If $\hat{\mathbf{y}}_0$ and $\hat{\mathbf{S}}_y$ denote respectively the simple cokriging estimate and covariance matrix of a log-ratio transformed composition $\mathbf{Y}(x_0)$, then

$$[\mathbf{Y}(x_0)|\mathbf{y}_1, \ldots, \mathbf{y}_N] \sim \mathcal{N}^{D-1}(\hat{\mathbf{y}}_0, \hat{\mathbf{S}}_y). \tag{27}$$

If interest lies in a non-linear function $g(\mathbf{Y}(x_0))$, a Monte Carlo sample of this quantity can be obtained from a set of K simulations $\{\tilde{\mathbf{y}}^{(1)}, \ldots, \tilde{\mathbf{y}}^{(K)}\}$. By transforming each one of them, the set $\{g(\tilde{\mathbf{y}}^{(1)}), \ldots, g(\tilde{\mathbf{y}}^{(K)})\}$ will describe the distribution of $g(\mathbf{Y}(x_0))$, and any relevant statistics (mean, variance, quantiles, etc.) can be derived from this transformed sample. This Monte Carlo strategy can also be used for spatially averaged quantities, such as block kriging estimates (Tolosana-Delgado et al. 2013). This methodology is affine equivariant.

An alternative is the use of Gauss–Hermite quadratures to estimate the expected value of $g(\mathbf{Y}(x_0))$. This requires a factorization of $\hat{\mathbf{S}}_y = \mathbf{R}\mathbf{R}^t$, for instance a Cholesky decomposition or a decomposition based on the singular value decomposition of $\hat{\mathbf{S}}_y$. Applying the change of variable $\mathbf{U} = \mathbf{R}^{-1} \cdot (\mathbf{Y} - \hat{\mathbf{y}}_0)/\sqrt{2}$, the expectation sought can be written as

$$\int_{\mathbb{R}^{D-1}} g(\mathbf{Y}) f_{\mathcal{N}}(\mathbf{Y}|\hat{\mathbf{y}}_0, \hat{\mathbf{S}}_y) d\mathbf{Y} = \int_{\mathbb{R}^{D-1}} \pi^{-(D-1)/2} g(\mathbf{Y}(\mathbf{U})) \exp(-\mathbf{U}^t\mathbf{U}) d\mathbf{U} = I,$$

where $f_{\mathcal{N}}(\cdot|\mathbf{m}, \mathbf{S})$ denotes a normal distribution with specified mean $\mathbf{m}$ and covariance $\mathbf{S}$. An order k Gauss–Hermite approximation can be computed using weights $w_1, w_2, \ldots, w_k$ and quadrature points $u_1, u_2, \ldots, u_k$ as

$$I \approx \sum_{i_1=1}^{k} \sum_{i_2=1}^{k} \cdots \sum_{i_{D-1}=1}^{k} w_{i_1} w_{i_2} \cdots w_{i_{D-1}} g(\hat{\mathbf{y}}_0 + \sqrt{2} \cdot \mathbf{R} \cdot \mathbf{u}_{[i_1, i_2, \ldots, u_{D-1}]}), \quad (28)$$

involving k^{D-1} $(D-1)$-tuples of Hermite quadrature points $\mathbf{u}_{[i_1, i_2, \ldots, u_{D-1}]} = [u_{i_1}, u_{i_2}, \ldots, u_{i_{D-1}}]$. These expressions were used by Aitchison (1986; p. 314) to obtain estimates of the composition $\mathbf{Z}(x_0)$ unbiased in terms of the original units (ppm, %, etc) under the assumption that the geometry of real space induced in the simplex holds. This is readily available using the inverse log-ratio function $g(\mathbf{Y}(x_0)) = \psi^{-1}(\mathbf{Y}(x_0))$ (Pawlowsky-Glahn and Olea 2004; Lark and Bishop 2007; Ward and Mueller 2013).

Being a Gauss quadrature approximation of $\mathrm{E}[\mathbf{Z}(x_0)]$, the result of this numerical integration is an unbiased estimator of $\mathbf{Z}(x_0)$ in the classical, Euclidean difference based sense. Note that the final results of this approximation to I might depend on which log-ratio was used, as the quadrature is not affine equivariant (Appendix "Spatial Results") and should, therefore, be avoided.

7.4 Illustration: Cosimulation in Point and Block Support

Cosimulation requires joint normality of the log-ratio scores. Figure 1 (lower triangle plots) show that the alr data do not have uni- and bi-normal marginals (nor will any other log-ratio transform). Hence, a form of joint multivariate normal anamorphosis is required. The flow anamorphosis (van den Boogaart et al. 2017) was applied with parameters $\sigma_0 = 0.1$ and $\sigma_1 = 1.1$. Resulting scores were shown to be spatially decorrelated, something often observed in flow-anamorphosed scores. Thus, only direct variograms of the five normal score transformed variables were modelled, with structures reported in Table 5 and the fit shown in Fig. 8. This allowed applying a separate simulation procedure for each of them. In a general framework with non-zero cross-variograms, one should use cosimulation in order to ensure that results are independent of the log-ratio transformations used.

The turning bands algorithm in ISATIS was used to generate the realisations. Outcomes were backtransformed via the inverse flow anamorphosis, and the inverse alr-transformation to obtain point-support simulations of the composition. These were upscaled to block support, with blocks of $12 \times 12 \times 6\,\mathrm{m}^3$, that is, averaging simulations at eight points. Figure 9 shows three views of the results. The spatial averages were all computed in mass scale, that is for each component tons/block were added at the different locations.

The E-type estimates were computed in the relevant scale. For single variables expressed in percentages, the relevant scale is additive, because the goal is to generate

Table 5 Variogram models of the flow-anamorphosed scores; Sph = spherical model of the indicated (X, Y, Z) ranges, relative to reference direction N300°, dip 45° and pitch −14°

	Ana_1	Ana_2	Ana_3	Ana_4	Ana_5
Nugget	0.24	0.18	0.26	0.13	0.13
$Sill_1$	0.22	0.25	0.15	0.26	0.3
Sph_1	(20, 5, 24)	(20, 5, 24)	(40, 5, 20)	(14, 14, 5)	(14, 20, 10)
$Sill_2$	0.19	0.19	0.3	0.31	0.35
Sph_2	(60, 100, 14.5)	(30, 100, 14.5)	(50, 70, 28)	(60, 60, 30)	(200, 100, 30)
$Sill_3$	0.36	0.38	0.3	0.3	0.22
Sph_3	(600, 200, ∞)	(400, 200, ∞)	(200, 400, ∞)	(180, 120, ∞)	(200, 300, ∞)

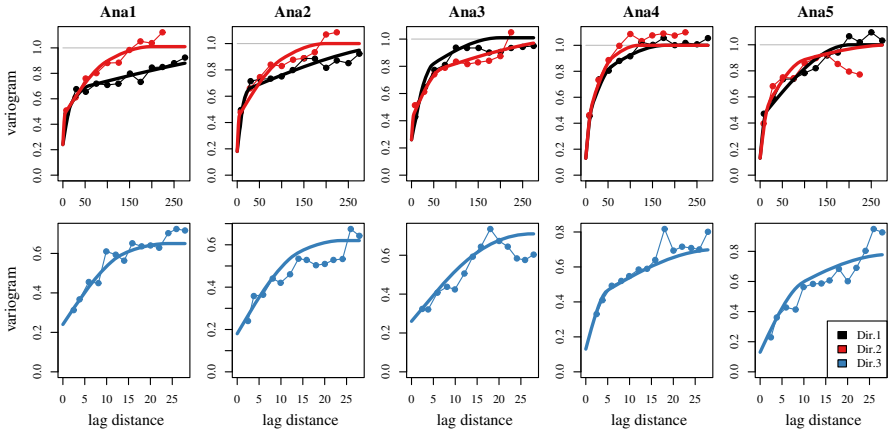

Fig. 8 Experimental variograms and fitted models for the flow-anamorphosed variables. The top row (red, black) shows direct variograms in horizontal directions (after global rotation), the bottom row (blue) those in the vertical direction

unbiased estimates of the mass of a certain element within each block, and masses are additive. In contrast, estimates of proportions should be unbiased regardless of what subcomposition is considered, consequently a compositional scale is relevant. Hence, out of $N_{sim} = 100$ simulations, E-types for %Fe at block V were obtained as

$$\text{Etype}[\%\text{Fe}(V)] = \frac{1}{N_{sim}} \sum_{k=1}^{N_{sim}} \sum_{x \in V} w^{(k)}(x) \text{Fe}^{(k)}(x), \tag{29}$$

while E-types for the (log)-proportion Fe:Mn were produced with

$$\text{Etype}\left[\ln \frac{\text{Fe}(V)}{\text{Mn}(V)}\right] = \frac{1}{N_{sim}} \sum_{k=1}^{N_{sim}} \ln \frac{\sum_{x \in V} w^{(k)}(x) \text{Fe}^{(k)}(x)}{\sum_{x \in V} w^{(k)}(x) \text{Mn}^{(k)}(x)}, \tag{30}$$

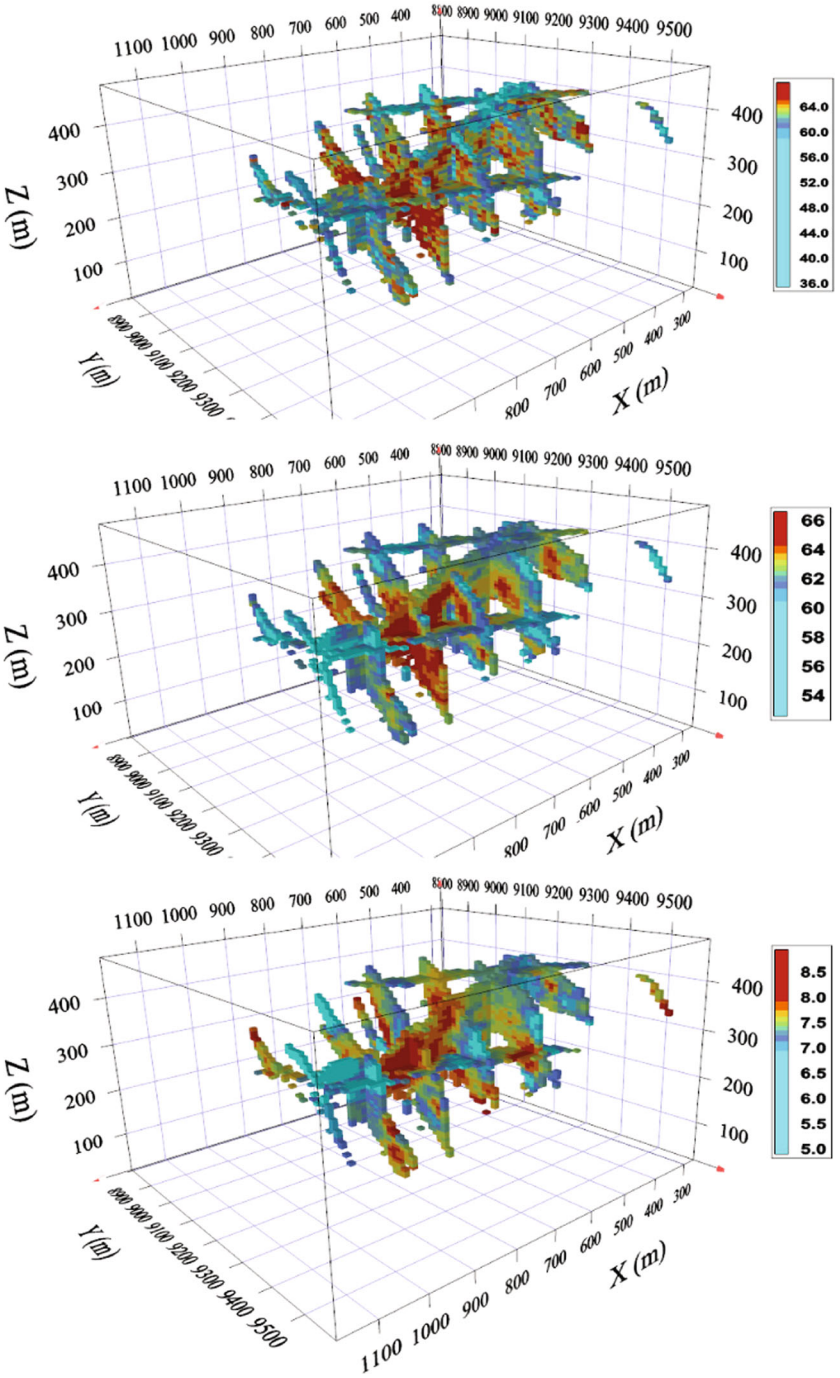

Fig. 9 North view of the deposit, one realization of the simulated block %Fe (top), E-type estimates of block %Fe (center) and of $\ln(Fe/Mn)$ (bottom), the last two computed from 100 simulations. Two horizontal slices, seven vertical transversal slices and one vertical longitudinal slice of the block model are shown

where $\mathrm{Fe}^{(k)}(x)$ or $\mathrm{Mn}^{(k)}(x)$ denote the values of iron or manganese at a location x for the k-th simulation, and $w^{(k)}(x)$ is the tonnage value allocated to point x at simulation k. This is the portion of mass allocated to a small block v of $6 \times 6 \times 3$ m^3 centered at x with respect to the mass of the block V of $12 \times 12 \times 6$ m^3. Estimating it properly requires cosimulating the rock density $\rho(x)$ as well (Tolosana-Delgado et al. 2015). In this illustration the density was assumed constant, so that $w^{(k)}(x) = 1/8$ for this configuration of simulation points within the block V. In Eq. (30) it is assumed that the processing of the block will occur after some form of spatial mixing, hence the spatial average is taken as an arithmetic mean: this will be a common situation in mining, due to the effect of milling. However, one can imagine other situations in which a spatial average is not relevant, in which case a quantity such as

$$\mathrm{Etype}\left[\ln\frac{\mathrm{Fe}}{\mathrm{Mn}}(V)\right] = \frac{1}{N_{sim}}\sum_{k=1}^{N_{sim}}\sum_{x\in V} w^{(k)}(x)\ln\frac{\mathrm{Fe}^{(k)}(x)}{\mathrm{Mn}^{(k)}(x)} \tag{31}$$

could be more appropriate.

The choice of the appropriate way to compute spatial averages or E-type estimates depends on the question at hand, that is: depends on the future uses of the summaries so obtained. Users should carefully consider which quantity are they going to use in the future (total mass of metal or pollutant in a block, ratio between deleterious and value elements/between two pollutants, cashflow produced by processing/remediating that block, etc), and choose an averaging strategy (Eqs. 29–31) that unbiasedly estimates it. No universal recipe is possible.

8 Conclusions

Compositional data should be treated geostatistically expressed in terms of log-ratios. The idea is to apply a log-ratio transformation and use conventional multivariate geostatistical tools on the transformed scores, then back-transform interpolations or simulations. This is analogous to the conventional Gaussian anamorphosis or normal score transform, but multivariate and concept-driven, instead of univariate and data-driven.

All log-ratio transformations available yield the same final compositional results, as long as the transformation is one-to-one and the variogram models used are compatible with each other. Indeed, given that the choice of one or another log-ratio transformation is somewhat an arbitrary one, this invariance (affine equivariance) is highly desirable. Moreover, it makes it unnecessary to discuss or consider that each alr or ilr could give different results. However, numerical approximations of non-linear transformations (for instance, the unbiased estimations of the composition in percentages within the original subcomposition) by Gauss–Hermite quadratures are not recommended because they are not affine equivariant and so results might differ for different log-ratio transformations.

Note that no univariate normal score transformations (Gaussian anamorphosis) is admissible, if this invariance is desired. Appropriate multivariate Gaussian

anamorphoses exist, such as the flow anamorphosis, which is by construction affine equivariant.

The variographic structure should be characterized by exploring several possible log-ratio sets, to check that the fit of the model to the empirical variograms is consistent across them. These should include at least variation-variograms, formed by direct variograms of all pairwise log-ratios. In addition to classical automatic fitting procedures, these variation-variograms can be fitted with a logarithmic goodness of fit criterion which focuses on improving the fit at short ranges.

Finally, bias and cross-validation error measures should be computed with caution. First, they should not be used at all to rank methods that minimize different kriging error variances. Second, given the multivariate nature of compositional data, multivariate measures of goodness-of-fit should be preferred over an individual component by component error assessment: a distance (preferentially an Aitchison–Mahalanobis distance) between each vector of observations and its predictions should be used.

With regard to representing compositions, a powerful option is to use matrices of pairwise log-ratios. This concept was first used in the variation matrix, as a matrix of variances, and has been extended to diagrams like the variation-variogram, or the visualization of cross-validation prediction versus true values.

Upscaling to block estimates as well as computing expectations of non-linear functions of the composition can be achieved by using simulations. This typically requires taking a multivariate Gaussian anamorphosis of the log-ratio transformed scores, and apply conventional variography and cosimulation procedures to the Gaussian scores obtained. Simulations can then be back-transformed to log-ratio scores and to compositions, and then upscaled or transformed as required. An average of the results will deliver a Monte Carlo estimation of the target quantities.

Acknowledgements This paper was compiled during research visits between Perth and Freiberg within the project "CoDaBlockCoEstimation", jointly funded by the German Academic Exchange Service (DAAD) and Universities Australia. The authors warmly thank Vera Pawlowsky-Glahn and Juan José Egozcue for their constructive comments on a previous version of this manuscript.

Appendix A: Formal Statements and Proofs

In this appendix, proofs for the results in the main text are provided. They are organised into a "non-spatial" part and a spatial part. A few of these results have been published before, albeit mostly for alr, clr or ilr transformations. They are included for reasons of self-containedness and generality, as the expressions in this contribution are valid for any full rank log-ratio transformation. These pre-existing results are cited as appropriate. In what follows the variables $\mathbf{z}$ or $\mathbf{Z}$ will denote a composition in the original scale and $\boldsymbol{\zeta}$ or $\boldsymbol{Z}$ will be used for the log-ratio scores.

A.1 Non-spatial Results

Definition 1 (*composition as a closed vector*) A vector $\mathbf{z} \in \mathbb{R}^D$ is called a composition if its k^{th} component z_k represents the relative importance of part k with respect to the remaining components.

Typically, $z_k \geq 0$ and $z_1 + z_i + \cdots + z_D = \kappa$, with $\kappa = 1$ (for proportions), $\kappa = 100$ (for percentages) and $\kappa = 10^6$ (for ppm). However, the variables under consideration might only represent a subset of all possible variables in which case the constant sum constraint is not necessarily satisfied. Subsequent treatment of the data then depends on whether or not the resulting non-constant sum is meaningful and less than κ. In this case a fill-up variable (Eq. 2) can be added to retain that information and fulfill the constraint. On the other hand, if the non-constant sum is meaningless, the data can be reclosed (Eq. 3) without losing any information. Mathematically, this last case gives rise to the definition of compositions as equivalence classes (Barceló-Vidal 2003), the modern, more general definition of composition.

Definition 2 (*log-ratio representation*) A function $\psi(\cdot)$ is a full-rank log-ratio representation of the composition $\mathbf{z}$ if its image $\boldsymbol{\zeta}$ satisfies

$$\boldsymbol{\zeta} = \psi(\mathbf{z}) = \boldsymbol{\Psi} \cdot \ln \mathbf{z},$$

where $\boldsymbol{\Psi}$ is a $(D-1) \times D$ matrix of rank $(D-1)$ with $\boldsymbol{\Psi} \cdot \mathbf{1}_D = \mathbf{0}_{D-1}$ (Barceló-Vidal and Martín-Fernández 2016).

Lemma 1 (inversion) *If $\psi(\cdot)$ is a full-rank log-ratio transformation, then the corresponding matrix $\boldsymbol{\Psi}$ satisfies $\boldsymbol{\Psi}^- \cdot \boldsymbol{\Psi} = \mathbf{H}$, where $\mathbf{H}$ is the projection matrix on the orthogonal complement of the vector $\mathbf{1}_D$ in $\mathbb{R}^D$ and $\boldsymbol{\Psi}^-$ is its generalized inverse.*

Proof The singular value decomposition of $\boldsymbol{\Psi}$ is given by $\boldsymbol{\Psi} = \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^t$, where $\mathbf{U}$ is an orthogonal $(D-1) \times (D-1)$ matrix, $\mathbf{V}$ is an orthogonal $D \times D$ matrix with $\mathbf{V}^t\mathbf{V} = \mathbf{I}_D$ and $\mathbf{S} = \left[\mathbf{D}_{(D-1)}\ \mathbf{0}_{(D-1)}\right]$ is a $(D-1) \times D$ matrix with $\mathbf{D}$ an invertible real diagonal matrix and $\mathbf{0}_{(D-1)}$ a column vector of zeros. The Moore-Penrose inverse is, therefore, $\boldsymbol{\Psi}^- = \mathbf{V} \cdot \mathbf{S}^+ \cdot \mathbf{U}^t$ where $\mathbf{S}^+ = [\mathbf{D}^{-1}\ \mathbf{0}_{(D-1)}]^t$. Then

$$\begin{aligned}\boldsymbol{\Psi}^- \cdot \boldsymbol{\Psi} &= (\mathbf{V} \cdot \mathbf{S}^+ \cdot \mathbf{U}^t) \cdot (\mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^t) \\ &= \mathbf{V} \cdot \begin{bmatrix} \mathbf{I}_{(D-1)} & \mathbf{0}_{(D-1)} \\ \mathbf{0}^t_{(D-1)} & 0 \end{bmatrix} \cdot \mathbf{V}^t.\end{aligned}$$

Since $\mathbf{S}^+ \cdot \mathbf{S}$ has rank $D-1$ and $\mathbf{V}$ has full rank, $\boldsymbol{\Psi}^- \cdot \boldsymbol{\Psi}$ has rank $D-1$ and its eigenvalues are 1 and 0. Since $(\boldsymbol{\Psi}^- \cdot \boldsymbol{\Psi})^2 = \boldsymbol{\Psi}^- \cdot \boldsymbol{\Psi}$, and $(\boldsymbol{\Psi}^- \cdot \boldsymbol{\Psi})^t = \boldsymbol{\Psi}^- \cdot \boldsymbol{\Psi}$, the matrix $\boldsymbol{\Psi}^- \cdot \boldsymbol{\Psi}$ is an orthogonal projection. Moreover from the definition of $\boldsymbol{\Psi}$ it follows that $\boldsymbol{\Psi}^- \cdot \boldsymbol{\Psi} \cdot \mathbf{1} = \boldsymbol{\Psi}^- \cdot \mathbf{0} = \mathbf{0}$. Therefore, if the columns of $\mathbf{V}$ are denoted by $\mathbf{v}_i, i = 1, \ldots, D$, the eigenvector for 0 is given by $\mathbf{v}_D = \frac{1}{\sqrt{D}}\mathbf{1}$, so that $(\boldsymbol{\Psi}^- \cdot \boldsymbol{\Psi}) = \mathbf{I}_D - \mathbf{v}_D\mathbf{v}_D^t = \mathbf{I}_D - \frac{1}{D}\mathbf{1}_{D\times D} = \mathbf{H}$. □

Proposition 1 (inverse log-ratio representation) *A full-rank log-ratio representation* $\psi(\cdot)$ *is one-to-one, and its inverse is*

$$\mathbf{z} = \mathcal{C}[\exp(\boldsymbol{\Psi}^{-} \cdot \boldsymbol{\zeta})].$$

Proof From the previous lemma if follows that

$$\begin{aligned}\mathcal{C}[\exp(\boldsymbol{\Psi}^{-} \cdot \boldsymbol{\zeta})] &= \mathcal{C}[\exp(\boldsymbol{\Psi}^{-} \cdot \boldsymbol{\Psi} \cdot \ln \mathbf{z})] \\ &= \mathcal{C}[\exp((\mathbf{H} \cdot \ln \mathbf{z})] \\ &= \mathcal{C}[\exp(\mathrm{clr}(\mathbf{z}))] \equiv \mathbf{z}.\end{aligned}$$

It remains to be shown that $\psi(\cdot)$ is one-to-one when restricted to the orthogonal complement of $\mathbf{1}_D$, but this is a direct consequence of the definition of $\psi(\cdot)$. □

Proposition 2 (change of log-ratio representation) *Let* $\mathbf{z}$ *be a composition, and* $\psi_1(\cdot)$ *and* $\psi_2(\cdot)$ *be two full-rank log-ratio transformations characterized by the matrices* $\boldsymbol{\Psi}_1$ *and* $\boldsymbol{\Psi}_2$ *respectively. Then, its two log-ratio representations* $\boldsymbol{\zeta}_1 = \psi_1(\mathbf{z})$ *and* $\boldsymbol{\zeta}_2 = \psi_2(\mathbf{z})$ *are related through the linear relationship*

$$\boldsymbol{\zeta}_2 = \mathbf{A}_{12} \cdot \boldsymbol{\zeta}_1, \tag{32}$$

where the matrix $\mathbf{A}_{12} = \boldsymbol{\Psi}_2 \cdot \boldsymbol{\Psi}_1^{-}$ *is square and invertible.*

Proof From the preceding two propositions it follows that $\boldsymbol{\zeta}_2 = \psi_2(\mathbf{z}) = \boldsymbol{\Psi}_2 \cdot \ln \mathbf{z}$ and $\mathbf{z} = \mathcal{C}[\exp(\boldsymbol{\Psi}_1^{-} \cdot \boldsymbol{\zeta}_1)]$. Substituting the second expression into the first, one has

$$\boldsymbol{\zeta}_2 = \boldsymbol{\Psi}_2 \cdot \ln\left(\mathcal{C}[\exp(\boldsymbol{\Psi}_1^{-} \cdot \boldsymbol{\zeta}_1)]\right) = \boldsymbol{\Psi}_2 \cdot \left[\boldsymbol{\Psi}_1^{-} \cdot \boldsymbol{\zeta}_1 - \alpha \mathbf{1}\right] = \boldsymbol{\Psi}_2 \cdot \boldsymbol{\Psi}_1^{-} \cdot \boldsymbol{\zeta}_1,$$

where $\alpha = \ln(\mathbf{1}^t \cdot \exp(\boldsymbol{\Psi}_1^{-} \cdot \boldsymbol{\zeta}_1))$. This last term satisfies $\kappa \boldsymbol{\Psi}_2 \cdot \mathbf{1} = \mathbf{0}$, which delivers the final expression as sought. □

Proposition 3 (log-ratio representation of the mean) *Let* $\mathbf{Z} = [z_{kn}]$, $k = 1, 2, \ldots, D$, $n = 1, 2, \ldots, N$, *be a compositional data set with N observations and of D parts, and* $\psi(\cdot)$ *be a full-rank log-ratio transformation. Then* $\hat{E}[\psi(\mathbf{Z})] = \psi(\hat{\boldsymbol{\mu}}^g)$ *the log-ratio representation of the closed geometric mean* (Eq. 13).

Proof The empirical closed geometric center is $\hat{\mathbf{m}} = \mathcal{C}[\exp(\ln(\mathbf{Z}) \cdot \mathbf{1}_N/N)]$. The log-ratio mean is given by $\hat{\mathrm{E}}[\psi(\mathbf{Z})] = (\boldsymbol{\Psi} \cdot \ln \mathbf{Z}) \cdot \mathbf{1}_N/N$. Substituting this expression into the definition of the inverse log-ratio representation results in

$$\begin{aligned}\psi^{-1}(\hat{\mathrm{E}}[\psi(\mathbf{Z})]) &= \mathcal{C}[\exp(\boldsymbol{\Psi}^{-} \cdot \hat{\mathrm{E}}[\psi(\mathbf{Z})])] = \mathcal{C}[\exp(\boldsymbol{\Psi}^{-} \cdot \boldsymbol{\Psi} \cdot \ln(\mathbf{Z}) \cdot \mathbf{1}_N/N)] \\ &= \mathcal{C}[\exp(\ln(\mathbf{Z}) \cdot \mathbf{1}_N/N)] = \hat{\boldsymbol{\mu}}^g.\end{aligned}$$

□

This proposition also proves Eq. (15): Because the calculation of $\hat{\boldsymbol{\mu}}^g$ does not involve any log-ratio representation, all log-ratio representations are equivalent. The idea of deriving statistics for compositional data from transformed scores is an application of the principle of working in coordinates (Mateu-Figueras et al. 2011).

Proposition 4 (log-ratio representations of the covariance) *Let* $\mathbf{Z} = [z_{kn}]$, $k = 1, 2, \ldots, D$, $n = 1, 2, \ldots, N$, *be a compositional data set with N observations and D parts, and* $\psi(\cdot)$ *be a full-rank log-ratio transformation. Then the covariance matrix of the log-ratio representation can be obtained from the empirical variation matrix* $\hat{\mathbf{T}}$ *as* $\hat{\boldsymbol{\Sigma}}^{\psi} = -\frac{1}{2}\boldsymbol{\Psi} \cdot \hat{\mathbf{T}} \cdot \boldsymbol{\Psi}^t$.

Proof From (Aitchison 1986) it is known that the clr covariance $\hat{\boldsymbol{\Sigma}}^c$ is related to the empirical variation matrix by $\hat{\boldsymbol{\Sigma}}^c = -\frac{1}{2}\mathbf{H} \cdot \hat{\mathbf{T}} \cdot \mathbf{H}$ and $\boldsymbol{\Psi} \cdot \mathbf{H} = \boldsymbol{\Psi}$, which is a consequence the definition of the matrix $\mathbf{H}$ (Eq. 9), where

$$\boldsymbol{\Psi} \cdot \mathbf{H} = \boldsymbol{\Psi} \cdot \left(\mathbf{I}_{D\times D} - \frac{1}{D}\mathbf{1}_{D\times D}\right) = \boldsymbol{\Psi} \cdot \mathbf{I}_{D\times D} - \frac{1}{D}\boldsymbol{\Psi}\mathbf{1}_{D\times D} = \boldsymbol{\Psi} - \frac{1}{D}\mathbf{0} = \boldsymbol{\Psi},$$

because the rows of $\boldsymbol{\Psi}$ sum to zero. Therefore, it remains to be shown $\hat{\boldsymbol{\Sigma}}^{\psi} = \boldsymbol{\Psi} \cdot \hat{\boldsymbol{\Sigma}}^c \cdot \boldsymbol{\Psi}^t$. The (maximum likelihood) estimators of these two covariance matrices are

$$\hat{\boldsymbol{\Sigma}}^{\psi} = \frac{1}{N}\left(\boldsymbol{\Psi} \cdot (\ln(\mathbf{Z}) - \ln(\hat{\boldsymbol{\mu}}^g) \cdot \mathbf{1}_N^t)\right) \cdot \left((\ln(\mathbf{Z}) - \ln(\hat{\boldsymbol{\mu}}^g) \cdot \mathbf{1}_N^t)^t \cdot \boldsymbol{\Psi}^t\right),$$
$$\hat{\boldsymbol{\Sigma}}^{c} = \frac{1}{N}\left(\mathbf{H} \cdot (\ln(\mathbf{Z}) - \ln(\hat{\boldsymbol{\mu}}^g) \cdot \mathbf{1}_N^t)\right) \cdot \left((\ln(\mathbf{Z}) - \ln(\hat{\boldsymbol{\mu}}^g) \cdot \mathbf{1}_N^t)^t \cdot \mathbf{H}\right).$$

Since $\mathbf{H} = \mathbf{H}^t$, so that $\mathbf{H} \cdot \boldsymbol{\Psi}^t = \boldsymbol{\Psi}^t$, it follows that

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}^{\psi} &= \frac{1}{N}\left(\boldsymbol{\Psi} \cdot \mathbf{H} \cdot (\ln(\mathbf{Z}) - \ln(\hat{\boldsymbol{\mu}}^g) \cdot \mathbf{1}_N^t)\right) \cdot \left((\ln(\mathbf{Z}) - \ln(\hat{\boldsymbol{\mu}}^g) \cdot \mathbf{1}_N^t)^t \cdot \mathbf{H} \cdot \boldsymbol{\Psi}^t\right) \\ &= \boldsymbol{\Psi} \cdot \hat{\boldsymbol{\Sigma}}^c \cdot \boldsymbol{\Psi}^t.\end{aligned}$$

Therefore $\hat{\boldsymbol{\Sigma}}^{\psi} = \boldsymbol{\Psi} \cdot \hat{\boldsymbol{\Sigma}}^c \cdot \boldsymbol{\Psi}^t = -\frac{1}{2}\boldsymbol{\Psi} \cdot \mathbf{H} \cdot \hat{\mathbf{T}} \cdot \mathbf{H} \cdot \boldsymbol{\Psi}^t = -\frac{1}{2}\boldsymbol{\Psi} \cdot \hat{\mathbf{T}} \cdot \boldsymbol{\Psi}^t$. □

It is straightforward to show that the same properties hold for unbiased estimators (with denominator $N - 1$).

The preceding two propositions show that the empirical log-ratio mean vector and covariance matrix can be obtained directly from the empirical closed geometric center and the variation matrix. Equivalent relationships exist also between the theoretical counterparts of these statistics.

Corollary 1 *If* $\psi(\cdot)$ *is a full rank log-ratio transformation, then* $\boldsymbol{\Psi}^- \cdot \hat{\boldsymbol{\Sigma}}^{\Psi} \cdot \boldsymbol{\Psi}^{-t} = \hat{\boldsymbol{\Sigma}}^c$.

Proof From Proposition 4 it follows that $\boldsymbol{\Psi}^- \cdot \hat{\boldsymbol{\Sigma}}^{\Psi} \cdot \boldsymbol{\Psi}^{-t} = \boldsymbol{\Psi}^- \cdot \boldsymbol{\Psi} \cdot \hat{\boldsymbol{\Sigma}}^c \cdot \boldsymbol{\Psi}^t \cdot \boldsymbol{\Psi}^{-t} = \mathbf{H} \cdot \hat{\boldsymbol{\Sigma}}^c \cdot \mathbf{H}^t = \hat{\boldsymbol{\Sigma}}^c$. □

Corollary 2 *If $\psi_1(\cdot)$ and $\psi_2(\cdot)$ are full rank log-ratio transformations, then $\hat{\boldsymbol{\Sigma}}^{\psi_2} = \mathbf{A}_{12} \cdot \hat{\boldsymbol{\Sigma}}^{\psi_1} \cdot \mathbf{A}_{12}^t$, where $\mathbf{A}_{12} = \boldsymbol{\Psi}_2 \cdot \boldsymbol{\Psi}_1^-$.*

Proof From $\hat{\boldsymbol{\Sigma}}^{\psi_2} = \boldsymbol{\Psi}_2 \cdot \hat{\boldsymbol{\Sigma}}^{c} \cdot \boldsymbol{\Psi}_2^t$ and Corollary 1 it follows that $\hat{\boldsymbol{\Sigma}}^{\psi_2} = \boldsymbol{\Psi}_2 \cdot \boldsymbol{\Psi}_1^- \cdot \hat{\boldsymbol{\Sigma}}^{\psi_1} \cdot \boldsymbol{\Psi}_1^{-t} \cdot \boldsymbol{\Psi}_2^t = \mathbf{A}_{12} \cdot \hat{\boldsymbol{\Sigma}}^{\psi_1} \cdot \mathbf{A}_{12}^t$. □

Corollary 3 *If $\psi(\cdot)$ is a full rank log-ratio transformation, then $(\hat{\boldsymbol{\Sigma}}^{c})^- = \boldsymbol{\Psi}^t \cdot (\hat{\boldsymbol{\Sigma}}^{\psi})^{-1} \cdot \boldsymbol{\Psi}$ is a generalised inverse of $\hat{\boldsymbol{\Sigma}}^{c}$.*

Proof Firstly, $\hat{\boldsymbol{\Sigma}}^{\Psi}$ has full rank and so is invertible, thus

$$\begin{aligned}
\hat{\boldsymbol{\Sigma}}^{c} \cdot (\hat{\boldsymbol{\Sigma}}^{c})^- &= \boldsymbol{\Psi}^- \cdot \hat{\boldsymbol{\Sigma}}^{\psi} \cdot \boldsymbol{\Psi}^{-t} \cdot \boldsymbol{\Psi}^t \cdot (\hat{\boldsymbol{\Sigma}}^{\psi})^{-1} \cdot \boldsymbol{\Psi} \\
&= \boldsymbol{\Psi}^- \cdot \hat{\boldsymbol{\Sigma}}^{\psi} \cdot (\boldsymbol{\Psi} \cdot \boldsymbol{\Psi}^-)^t \cdot (\hat{\boldsymbol{\Sigma}}^{\psi})^{-1} \cdot \boldsymbol{\Psi} \\
&= \boldsymbol{\Psi}^- \cdot \hat{\boldsymbol{\Sigma}}^{\psi} \cdot (\hat{\boldsymbol{\Sigma}}^{\psi})^{-1} \cdot \boldsymbol{\Psi} \\
&= \boldsymbol{\Psi}^- \boldsymbol{\Psi} = \mathbf{H}.
\end{aligned}$$

since $(\boldsymbol{\Psi} \cdot \boldsymbol{\Psi}^-) = \mathbf{I}_{(D-1)}$, so that $\hat{\boldsymbol{\Sigma}}^{c} \cdot (\hat{\boldsymbol{\Sigma}}^{c})^-$ is symmetric. Secondly, $\hat{\boldsymbol{\Sigma}}^{c} \cdot (\hat{\boldsymbol{\Sigma}}^{c})^- \cdot \hat{\boldsymbol{\Sigma}}^{c} = \mathbf{H} \cdot \hat{\boldsymbol{\Sigma}}^{c} = \hat{\boldsymbol{\Sigma}}^{c}$ and $(\hat{\boldsymbol{\Sigma}}^{c})^- \cdot \hat{\boldsymbol{\Sigma}}^{c} \cdot (\hat{\boldsymbol{\Sigma}}^{c})^- = (\hat{\boldsymbol{\Sigma}}^{c})^- \cdot \mathbf{H} = \boldsymbol{\Psi}^t \cdot (\hat{\boldsymbol{\Sigma}}^{\psi})^{-1} \cdot \boldsymbol{\Psi} \cdot \mathbf{H} = (\hat{\boldsymbol{\Sigma}}^{c})^-$ Similarly, $(\hat{\boldsymbol{\Sigma}}^{c})^- \cdot \hat{\boldsymbol{\Sigma}}^{c} = \mathbf{H}$. Therefore, $\boldsymbol{\Psi}^t \cdot (\hat{\boldsymbol{\Sigma}}^{\psi})^{-1} \cdot \boldsymbol{\Psi}$ satisfies all conditions of a generalised inverse. □

Proposition 5 (invariance of the Mahalanobis distance) *Let $\mathbf{Z}$ be a random composition, with variation matrix $\mathbf{T}$. The Aitchison–Mahalanobis distance between any two of its realisations $\mathbf{z}_1$ and $\mathbf{z}_2$*

$$d_M^2(\mathbf{z}_1, \mathbf{z}_2) = \psi(\mathbf{z}_1 \ominus \mathbf{z}_2)^t \cdot [\boldsymbol{\Sigma}^{\psi}]^{-1} \cdot \psi(\mathbf{z}_1 \ominus \mathbf{z}_2),$$

is invariant under the choice of full-rank log-ratio representation $\psi(\cdot)$.

Proof To show this proposition, it suffices to observe that from Corollary 3 and the proof of Proposition 4 one obtains $\mathbf{H} \cdot (\boldsymbol{\Sigma}^{c})^- \cdot \mathbf{H} = \mathbf{H} \cdot \boldsymbol{\Psi}^t \cdot (\boldsymbol{\Sigma}^{\Psi})^{-1} \cdot \boldsymbol{\Psi} \cdot \mathbf{H} = \boldsymbol{\Psi}^t \cdot (\boldsymbol{\Sigma}^{\Psi})^{-1} \cdot \boldsymbol{\Psi}$ so that

$$\begin{aligned}
d_M^2(\mathbf{z}_1, \mathbf{z}_2) &= \ln(\mathbf{z}_1 \ominus \mathbf{z}_2)^t \cdot \boldsymbol{\Psi}^t \cdot (\boldsymbol{\Sigma}^{\Psi})^{-1} \cdot \boldsymbol{\Psi} \cdot \ln(\mathbf{z}_1 \ominus \mathbf{z}_2) \\
&= -2 \ln(\mathbf{z}_1 \ominus \mathbf{z}_2)^t \cdot \boldsymbol{\Psi}^t \cdot \boldsymbol{\Psi}^{-t} \cdot \mathbf{T}^- \cdot \boldsymbol{\Psi}^- \cdot \boldsymbol{\Psi} \cdot \ln(\mathbf{z}_1 \ominus \mathbf{z}_2),
\end{aligned}$$

an expression, which, given that $\boldsymbol{\Psi}^- \cdot \boldsymbol{\Psi} = \mathbf{H}$, does not depend on the log-ratio representation at all. □

Filzmoser and Hron (2008) proved a more restricted version of this proposition, valid for the set of clr, alr and ilr log-ratio transformations. Proposition 6 is a direct consequence of the invariance property of the Mahalanobis distance.

Proposition 6 (invariance of the normal distribution) *The probability density function of the normal distribution on the simplex with center* $\mathbf{m}$ *and variation matrix* $\mathbf{T}$,

$$f_{\mathbf{Z}}(\mathbf{z}) = (2\pi)^{-(D-1)/2} |\boldsymbol{\Sigma}^{\psi}|^{-1/2} \exp\left[-\frac{1}{2} d_M^2(\mathbf{z}, \mathbf{m})\right],$$

does not depend on the choice of full-rank log-ratio representation $\psi(\cdot)$.

Analogous results are available for the case when the log-ratio transformation is not full-rank. In that case the determinant $|\boldsymbol{\Sigma}^{\psi}|$ needs to be generalised to the product of its non-zero eigenvalues. This invariance (Mateu-Figueras et al. 2013) is a direct consequence of the preceding Proposition 5 and the fact that the determinant of a matrix is one of its invariants.

A.2 Spatial Results

Definition 3 (*compositional random function*) A vector-valued random function $\mathbf{Z} = [Z_1, Z_2, \ldots, Z_D]$ on a spatial domain $\mathcal{D} \subset \mathbb{R}^p$, is called compositional if for each $x \in \mathcal{D}$ the vector of random variables $\mathbf{Z}(x) = [Z_1(x), Z_2(x), \ldots, Z_D(x)]$ shows the relative importance of a set of parts forming a total of interest.

Definition 4 (*regionalized composition*) Given a set of locations $\{x_1, x_2, \ldots, x_N\}$, a regionalized data set $\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N\}$ with $\mathbf{z}_i = \mathbf{z}(x_i) = [z_1(x_i), \ldots z_D(x_i)] = [z_{1i}, \ldots, z_{Di}]$, $i = 1, 2, \ldots, N$ is called a regionalized composition, if z_{ki} represents the relative importance of part k with respect to the set of components considered at location x_i.

Proposition 7 (log-ratio representation of the spatial structure) *Let* $\mathbf{Z} = [z_{ki}] = [z_k(x_i)]$, $k = 1, 2, \ldots, D$, $i = 1, 2, \ldots, N$, *be a regionalized compositional data set with* N *locations* x_i *and* D *parts, and* $\psi(\cdot)$ *be a full-rank log-ratio transformation. Then, for each lag* h, *the variogram of the log-ratio representation can be obtained from the empirical variation-variogram* $\hat{\mathbf{T}}(h)$ *as* $\hat{\boldsymbol{\Sigma}}^{\psi}(h) = -\frac{1}{2}\boldsymbol{\Psi} \cdot \hat{\mathbf{T}}(h) \cdot \boldsymbol{\Psi}^t$, *or from the clr-variogram matrix as* $\hat{\boldsymbol{\Gamma}}^{\psi}(h) = \boldsymbol{\Psi} \cdot \hat{\boldsymbol{\Gamma}}^{c}(h) \cdot \boldsymbol{\Psi}^t$.

This is a direct consequence of Propositions 3 and 4.

Proposition 8 (equivalence of the spatial structure) *Let* $\mathbf{Z} = [z_{kn}] = [z_k(x_n)]$, $k = 1, 2, \ldots, D$, $n = 1, 2, \ldots, N$, *be a regionalized compositional data set with* N *locations* x_n *and* D *parts, and let* $\psi_1(\cdot)$ *and* $\psi_2(\cdot)$ *be two full-rank log-ratio transformations. Then, for each lag* h, *the empirical variograms* $\hat{\boldsymbol{\Gamma}}^{\psi_1}(h)$ *and* $\hat{\boldsymbol{\Gamma}}^{\psi_2}(h)$ *are related through the linear relationship*

$$\hat{\boldsymbol{\Gamma}}^{\psi_2}(h) = \mathbf{A}_{12} \cdot \hat{\boldsymbol{\Gamma}}^{\psi_1}(h) \cdot \mathbf{A}_{12}^t, \tag{33}$$

with matrix $\mathbf{A}_{12} = \boldsymbol{\Psi}_2 \cdot \boldsymbol{\Psi}_1^-$ *square and invertible.*

Proof From Proposition 7 it follows that $\hat{\boldsymbol{\Gamma}}^{\psi_2}(h) = \boldsymbol{\Psi}_2 \cdot \hat{\boldsymbol{\Gamma}}^{c}(h) \cdot \boldsymbol{\Psi}_2^t$; and because of Eq. (20), $\hat{\boldsymbol{\Gamma}}^{c}(h) = \boldsymbol{\Psi}_1^- \cdot \hat{\boldsymbol{\Gamma}}^{\psi_1}(h) \cdot \boldsymbol{\Psi}_1^{-t}$. Therefore

$$\hat{\boldsymbol{\Gamma}}^{\psi_2}(h) = \boldsymbol{\Psi}_2 \cdot \boldsymbol{\Psi}_1^- \cdot \hat{\boldsymbol{\Gamma}}^{\psi_1}(h) \cdot \boldsymbol{\Psi}_1^{-t} \cdot \boldsymbol{\Psi}_2^t,$$

which proves the desired equality because $\mathbf{A}_{12}^t = \boldsymbol{\Psi}_1^{-t} \cdot \boldsymbol{\Psi}_2^t$. □

Since Proposition 8 holds for all lags, it is normal to require that any fitted model satisfies the same relation. This is automatically satisfied if a linear model of coregionalization $\mathbf{T}(h|\boldsymbol{\theta})$ is fitted to the variation-variograms and then recast to each of the two log-ratio representations via Proposition 7.

Proposition 9 (invariance of the cokriging predictor and errors) *Let* $\mathbf{Z} = [z_{kn}] = [z_k(x_n)]$, $k = 1, 2, \ldots, D$, $n = 1, 2, \ldots, N$, *be a regionalized compositional data set with N locations* x_n *and D parts, and* $\psi_1(\cdot)$ *and* $\psi_2(\cdot)$ *be two full-rank log-ratio transformations. Then, the corresponding cokriging predictors* $\hat{\boldsymbol{\zeta}}_1(x_0)$ *and* $\hat{\boldsymbol{\zeta}}_2(x_0)$ *of the log-ratio transformed composition* $\boldsymbol{\zeta}_i(x_0) = \psi_i(\mathbf{Z}(x_0))$ *satisfy*

$$\hat{\boldsymbol{\zeta}}_2(x_0) = \mathbf{A}_{12} \cdot \hat{\boldsymbol{\zeta}}_1(x_0),$$

so that

$$\psi_1^{-1}(\hat{\boldsymbol{\zeta}}_1(x_0)) = \psi_2^{-1}(\hat{\boldsymbol{\zeta}}_2(x_0)) =: \hat{\mathbf{z}}(x_0),$$

gives a predicted composition independent of the log-ratio representation used in the computations. Moreover, the corresponding cokriging error covariance matrices $\mathbf{S}_1$ *and* $\mathbf{S}_2$ *are related by*

$$\mathbf{S}_2^K = \mathbf{A}_{12} \cdot \mathbf{S}_1^K \cdot \mathbf{A}_{12}^t,$$

with $\mathbf{A}_{12} = \boldsymbol{\Psi}_2 \cdot \boldsymbol{\Psi}_1^-$, *for all forms of cokriging (simple, ordinary, universal and cokriging with a trend) at all locations* x_0, *if both are derived from the same linear model of coregionalization* $\mathbf{T}(h|\boldsymbol{\theta})$.

Proof The case of simple cokriging (SK) under the assumption of second-order stationarity will be considered first. In both log-ratio representations, the SK predictor is of the form

$$\hat{\boldsymbol{\zeta}}(x_0) = \sum_{n=1}^{N} \boldsymbol{\lambda}_n^t \boldsymbol{\zeta}(x_n) = \boldsymbol{\Lambda}^t \boldsymbol{Z}, \tag{34}$$

where $\boldsymbol{Z} = [\boldsymbol{\zeta}(x_1); \boldsymbol{\zeta}(x_2); \ldots; \boldsymbol{\zeta}(x_N)]$ is the concatenated vector of all log-ratio transformed observations $\boldsymbol{\zeta}(x_n) = \boldsymbol{\Psi} \ln \mathbf{z}(x_n)$, and $\boldsymbol{\Lambda} = [\boldsymbol{\lambda}_1; \boldsymbol{\lambda}_2; \ldots; \boldsymbol{\lambda}_N]$ is the block matrix of all cokriging weight matrices, which are obtained as (Myers 1982)

$$\boldsymbol{\Lambda} = \underbrace{\begin{bmatrix} \boldsymbol{\Gamma}_{11} & \boldsymbol{\Gamma}_{12} & \cdots & \boldsymbol{\Gamma}_{1N} \\ \boldsymbol{\Gamma}_{21} & \boldsymbol{\Gamma}_{22} & \cdots & \boldsymbol{\Gamma}_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}_{N1} & \boldsymbol{\Gamma}_{N2} & \cdots & \boldsymbol{\Gamma}_{NN} \end{bmatrix}^{-1}}_{\mathbf{W}^{-1}} \underbrace{\begin{bmatrix} \boldsymbol{\Gamma}_{10} \\ \boldsymbol{\Gamma}_{20} \\ \vdots \\ \boldsymbol{\Gamma}_{N0} \end{bmatrix}}_{\mathbf{W}_0} = \mathbf{W}^{-1}\mathbf{W}_0,$$

where each block $\boldsymbol{\Gamma}_{nm} = \boldsymbol{\Gamma}(h|\boldsymbol{\theta}) = -\frac{1}{2}\boldsymbol{\Psi}\mathbf{T}(h|\boldsymbol{\theta})\boldsymbol{\Psi}^t$ using the fitted model $\mathbf{T}(h|\boldsymbol{\theta})$. With the same notation, the SK error covariance is given by

$$\mathbf{S}^K = \boldsymbol{\Gamma}_{00} - \boldsymbol{\Lambda}^t\mathbf{W}_0 = \boldsymbol{\Gamma}_{00} - \mathbf{W}_0^t\mathbf{W}^{-1}\mathbf{W}_0.$$

Considering these matrices obtained with the two distinct log-ratio representations, and taking Eq. (33) into account, then

$$\begin{aligned} \mathbf{W}^{(2)} &= \begin{bmatrix} \boldsymbol{\Gamma}_{11}^{(2)} & \boldsymbol{\Gamma}_{12}^{(2)} & \cdots & \boldsymbol{\Gamma}_{1N}^{(2)} \\ \boldsymbol{\Gamma}_{21}^{(2)} & \boldsymbol{\Gamma}_{22}^{(2)} & \cdots & \boldsymbol{\Gamma}_{2N}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}_{N1}^{(2)} & \boldsymbol{\Gamma}_{N2}^{(2)} & \cdots & \boldsymbol{\Gamma}_{NN}^{(2)} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A}_{12}\boldsymbol{\Gamma}_{11}^{(1)}\mathbf{A}_{12}^t & \mathbf{A}_{12}\boldsymbol{\Gamma}_{12}^{(1)}\mathbf{A}_{12}^t & \cdots & \mathbf{A}_{12}\boldsymbol{\Gamma}_{1N}^{(1)}\mathbf{A}_{12}^t \\ \mathbf{A}_{12}\boldsymbol{\Gamma}_{21}^{(1)}\mathbf{A}_{12}^t & \mathbf{A}_{12}\boldsymbol{\Gamma}_{22}^{(1)}\mathbf{A}_{12}^t & \cdots & \mathbf{A}_{12}\boldsymbol{\Gamma}_{2N}^{(1)}\mathbf{A}_{12}^t \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{12}\boldsymbol{\Gamma}_{N1}^{(1)}\mathbf{A}_{12}^t & \mathbf{A}_{12}\boldsymbol{\Gamma}_{N2}^{(1)}\mathbf{A}_{12}^t & \cdots & \mathbf{A}_{12}\boldsymbol{\Gamma}_{NN}^{(1)}\mathbf{A}_{12}^t \end{bmatrix} \\ &= \mathbf{A}\mathbf{W}^{(1)}\mathbf{A}^t, \end{aligned} \tag{35}$$

where $\mathbf{A} = \mathrm{diag}(\mathbf{A}_{12}, \mathbf{A}_{12}, \ldots, \mathbf{A}_{12})$ and similarly

$$\mathbf{W}_0^{(2)} = \mathbf{A}\mathbf{W}_0^{(1)}\mathbf{A}_{12}^t. \tag{36}$$

Now substituting Eqs. (35) and (36) into the expression for the weights

$$\begin{aligned} \boldsymbol{\Lambda}^{(2)} &= [\mathbf{W}^{(2)}]^{-1}\mathbf{W}_0^{(2)} = [\mathbf{A}\mathbf{W}^{(1)}\mathbf{A}^t]^{-1}\mathbf{A}\mathbf{W}_0^{(1)}\mathbf{A}_{12}^t \\ &= \mathbf{A}^{-t}[\mathbf{W}^{(1)}]^{-1}\mathbf{A}^{-1}\mathbf{A}\mathbf{W}_0^{(1)}\mathbf{A}_{12}^t = \mathbf{A}^{-t}[\mathbf{W}^{(1)}]^{-1}\mathbf{W}_0^{(1)}\mathbf{A}_{12}^t \\ &= \mathbf{A}^{-t}\boldsymbol{\Lambda}^{(1)}\mathbf{A}_{12}^t, \end{aligned} \tag{37}$$

which implies that the cokriging weight matrices of each datum satisfy

$$\boldsymbol{\lambda}_n^{(2)} = \mathbf{A}_{12}^{-t}\boldsymbol{\lambda}_n^{(1)}\mathbf{A}_{12}^t$$

due to the block-diagonal structure of $\mathbf{A}$. Finally, substituting these weights into the SK predictor of the second log-ratio representation, and taking into account Eq. (32) between the data, one obtains

$$\begin{aligned}\hat{\boldsymbol{\zeta}}_2(x_0) &= \sum_{n=1}^{N}[\boldsymbol{\lambda}_n^{(2)}]^t\boldsymbol{\zeta}_2(x_n) = \sum_{n=1}^{N}(\mathbf{A}_{12}^{-t}\boldsymbol{\lambda}_n^{(1)}\mathbf{A}_{12}^t)^t\mathbf{A}_{12}\boldsymbol{\zeta}_1(x_n)\\ &= \sum_{n=1}^{N}\mathbf{A}_{12}[\boldsymbol{\lambda}_n^{(1)}]^t\mathbf{A}_{12}^{-1}\mathbf{A}_{12}\boldsymbol{\zeta}_1(x_n) = \mathbf{A}_{12}\sum_{n=1}^{N}[\boldsymbol{\lambda}_n^{(1)}]^t\boldsymbol{\zeta}_1 = \mathbf{A}_{12}\hat{\boldsymbol{\zeta}}_1(x_0),\end{aligned}$$

thus establishing the identity between the cokriging predictors. To derive the relation for the cokriging error covariance, the same strategy can be used to express the error in terms of the second log-ratio representation as a function of that in terms of the first representation,

$$\begin{aligned}\mathbf{S}_{(2)}^K &= \boldsymbol{\Gamma}_{00}^{(2)} - [\boldsymbol{\Lambda}^{(2)}]^t\mathbf{W}_0^{(2)} = \mathbf{A}_{12}\boldsymbol{\Gamma}_{00}^{(1)}\mathbf{A}_{12}^t - [\mathbf{A}^{-t}\boldsymbol{\Lambda}^{(1)}\mathbf{A}_{12}^t]^t\mathbf{A}\mathbf{W}_0^{(1)}\mathbf{A}_{12}^t\\ &= \mathbf{A}_{12}\boldsymbol{\Gamma}_{00}^{(1)}\mathbf{A}_{12}^t - \mathbf{A}_{12}[\boldsymbol{\Lambda}^{(1)}]^t\mathbf{A}^{-1}\mathbf{A}\mathbf{W}_0^{(1)}\mathbf{A}_{12}^t\\ &= \mathbf{A}_{12}\boldsymbol{\Gamma}_{00}^{(1)}\mathbf{A}_{12}^t - \mathbf{A}_{12}[\boldsymbol{\Lambda}^{(1)}]^t\mathbf{W}_0^{(1)}\mathbf{A}_{12}^t = \mathbf{A}_{12}\left[\boldsymbol{\Gamma}_{00}^{(1)} - [\boldsymbol{\Lambda}^{(1)}]^t\mathbf{W}_0^{(1)}\right]\mathbf{A}_{12}^t\\ &= \mathbf{A}_{12}\mathbf{S}_{(1)}^K\mathbf{A}_{12}^t,\end{aligned}$$

which proves the desired equivalence.

For the remaining cases of cokriging (which will be grouped under the name of universal cokriging, UK), the log-ratio mean is assumed to have the form

$$\boldsymbol{\mu}(x) = \sum_{l=1}^{L} g_l(x)\mathbf{b}_l,$$

with the typical cases $L = 1$ and $g_1(x) \equiv 1$ (for ordinary cokriging), $g_l(x) = x^{l-1}$ up to the desired order L (universal cokriging), or $L = 1$ and $g_1(x)$ an arbitrary function available everywhere in the estimation domain (for cokriging with a trend). In any case, the UK predictor has the same form (Eq. 34), where the weights are obtained from the solution of the system

$$\mathbf{W}\boldsymbol{\Lambda} = \mathbf{W}_0,$$

subject to the L unbiasedness conditions

$$\sum_{n=1}^{N} g_l(x_n)\boldsymbol{\lambda}_n^t = g_l(x_0)\mathbf{I}_{D-1}, \qquad l = 1, 2, \ldots, L.$$

where $\mathbf{I}_{D-1}$ is the identity matrix of size $(D-1)$, the dimension of the composition. It is known (Myers 1982; Tolosana-Delgado 2006) that this is equivalent to solving an extended system of equations

$$\mathbf{W}_e \boldsymbol{\Lambda}_e = \mathbf{W}_{e0}, \tag{38}$$

where

$$\mathbf{W}_e = \begin{bmatrix} \mathbf{W} & \mathbf{G} \\ \mathbf{G}^t & 0\mathbf{I}_{L(D-1)} \end{bmatrix}, \qquad \mathbf{W}_{e0} = \begin{bmatrix} \mathbf{W}_0 \\ \mathbf{G}_0^t \end{bmatrix}, \qquad \boldsymbol{\Lambda}_e = \begin{bmatrix} \boldsymbol{\Lambda} \\ \mathbf{N} \end{bmatrix},$$

with $\mathbf{N}^t = [\boldsymbol{\nu}_1; \boldsymbol{\nu}_2; \ldots; \boldsymbol{\nu}_L]$ the Lagrange multipliers for each unbiasedness condition, and $\mathbf{G}^t = [\mathbf{G}_1^t; \mathbf{G}_2^t; \ldots; \mathbf{G}_N^t]$ with

$$\mathbf{G}_i = [g_1(x_i)\mathbf{I}_{D-1}; g_2(x_i)\mathbf{I}_{D-1}; \ldots; g_L(x_i)\mathbf{I}_{D-1}], \qquad i = 0, 1, \ldots, N.$$

The UK error covariance matrix is then shown to be

$$\mathbf{S}^K = \boldsymbol{\Gamma}_{00} - \boldsymbol{\Lambda}_e^t \mathbf{W}_{e0} = \boldsymbol{\Gamma}_{00} - \mathbf{W}_{e0}^t \mathbf{W}_e^{-1} \mathbf{W}_{e0}.$$

Since the UK and SK system of equations, predictors and errors have analogous forms, the proposition for the case of UK can be proved by showing that, if the extended matrices satisfy Eqs. (35)–(37), then they satisfy the UK system of equations (Eq. 38) as well. That is, if $\mathbf{W}_e^{(2)} = \mathbf{A}\mathbf{W}_e^{(1)}\mathbf{A}^t$ (Eq. 35) and $\boldsymbol{\Lambda}_e^{(2)} = \mathbf{A}^{-t}\boldsymbol{\Lambda}_e^{(1)}\mathbf{A}_{12}^t$ (Eq. 37), then in Eq. (38), becomes

$$\mathbf{W}_e^{(2)}\boldsymbol{\Lambda}_e^{(2)} = [\mathbf{A}\mathbf{W}_e^{(1)}\mathbf{A}^t][\mathbf{A}^{-t}\boldsymbol{\Lambda}_e^{(1)}\mathbf{A}_{12}^t] = \mathbf{A}\mathbf{W}_e^{(1)}\boldsymbol{\Lambda}_e^{(1)}\mathbf{A}_{12}^t = \mathbf{A}\mathbf{W}_{e0}^{(1)}\mathbf{A}_{12}^t = \mathbf{W}_{e0}^{(2)},$$

which holds given Eq. (36). □

Lastly, the relationship is established between the quadratures for distinct log-ratio representations $\psi_1(\cdot)$ and $\psi_2(\cdot)$. The weights $w_1, w_2, \ldots, w_k$ and quadrature points $u_1, u_2, \ldots, u_k$ do not depend on the choice of log-ratio representation. If $\hat{\boldsymbol{\zeta}}_i$ is the predictor using the i-th log-ratio representation, then by Proposition 9 the representations $\hat{\boldsymbol{\zeta}}_1$ and $\hat{\boldsymbol{\zeta}}_2$ are related by $\mathbf{A}_{12} \cdot \hat{\boldsymbol{\zeta}}_1 = \hat{\boldsymbol{\zeta}}_2$.

The spectral decomposition of the cokriging error covariance matrix $\mathbf{S}_1^K$ is given by $\mathbf{S}_1^K = \mathbf{V}_1 \cdot \mathbf{D}_1 \cdot \mathbf{V}_1^t$, where $\mathbf{D}_1$ is a diagonal matrix and $\mathbf{V}_1$ is an orthogonal matrix of eigenvectors then $\mathbf{R}_1 = \mathbf{V}_1 \cdot \mathbf{D}_1^{1/2} \cdot \mathbf{V}_1^t$ is a square root of $\mathbf{S}_1^K$ and so from the congruence one has

$$\mathbf{S}_2^K = \mathbf{A}_{12} \cdot \mathbf{S}_1^K \cdot \mathbf{A}_{12}^t = \mathbf{A}_{12} \cdot (\mathbf{V}_1 \cdot \mathbf{D}_1 \cdot \mathbf{V}_1^t) \cdot \mathbf{A}_{12}^t.$$

This expression can be rewritten as

$$\begin{aligned} \mathbf{S}_2^K &= \mathbf{A}_{12} \cdot \mathbf{V}_1 \cdot \mathbf{D}_1^{1/2} \cdot \mathbf{V}_1^t \cdot (\mathbf{A}_{12}^{-1} \cdot \mathbf{A}_{12}) \cdot \mathbf{V}_1 \cdot \mathbf{D}_1^{1/2} \cdot \mathbf{V}_1^t \cdot \mathbf{A}_{12}^t \\ &= (\mathbf{A}_{12} \cdot \mathbf{R}_1 \cdot \mathbf{A}_{12}^{-1}) \cdot \mathbf{A}_{12} \cdot \mathbf{R}_1 \cdot \cdot \mathbf{A}_{12}^t, \end{aligned}$$

and so

$$\mathbf{R}_2 = \mathbf{A}_{12} \cdot \mathbf{R}_1 \cdot \mathbf{A}_{12}^t,$$

is a square root of $\mathbf{S}_2$ if and only if $\mathbf{A}_{12}^t = \mathbf{A}_{12}^{-1}$, that is, $\mathbf{A}_{12}$ is an orthogonal matrix. In that case the quadrature vectors $\boldsymbol{\zeta}(i_1, i_2, \ldots, i_{D-1})$ are related by

$$\boldsymbol{\zeta}_{(2)}(i_1, i_2, \ldots, i_{D-1}) = \boldsymbol{\zeta}_2 + \mathbf{R}_2 \cdot \mathbf{u}_{[i_1,i_2,\ldots,i_{D-1}]} \tag{39}$$

$$= \mathbf{A}_{12} \cdot \boldsymbol{\zeta}_1 + \mathbf{A}_{12} \cdot \mathbf{R}_1 \cdot \mathbf{A}_{12}^t \cdot \mathbf{u}_{[i_1,i_2,\ldots,i_{D-1}]} \tag{40}$$

$$= \mathbf{A}_{12} \cdot (\boldsymbol{\zeta}_1 + \mathbf{R}_1 \cdot \cdot \mathbf{v}_{[i_1,i_2,\ldots,i_{D-1}]}). \tag{41}$$

where $\mathbf{v}_{[i_1,i_2,\ldots,i_{D-1}]} = \mathbf{A}_{12} \cdot \mathbf{u}_{[i_1,i_2,\ldots,i_{D-1}]}$. Thus Gauss–Hermite quadratures are invariant under the choice of ilr transformation only, but they are not affine equivariant.

Appendix B: Compositional Geostatistics Workflow

B.1 Interpolation

1. Perform both classical and compositional exploratory analysis (Sect. 3.4)
2. Compute variation-variograms of the regionalized composition (Eq. 22)
3. Fit a valid model (Sect. 5.2); models such as the linear model of coregionalization or the minimum/maximum autocorrelation factors are useful
4. Recast both the experimental and the model variation-variograms via other log-ratio transformation with respectively Eqs. (23) and (25), in order to confirm that the model fits the data reasonably well in these other reference systems with respect to these other log-ration representations
5. Choose one of these alternative log-ratio transforms, and compute the scores of the data (Eq. 10)
6. Apply cokriging to the log-ratio scores with variogram model expressed in the same log-ratios on a suitably chosen grid; store cokriging covariance error matrices if cross-validation or Gauss–Hermite quadratures is desired
7. Backtransform the predicted values
8. If unbiased estimates of the mass of each component are required and an ilr is being used, estimate them through Gauss-Hermite quadratures (Eq. 28); otherwise, follow the procedure in the Sect. B.2
9. Further products (maps, cross-validation, block models, etc) can be derived from individual components of the composition or from relevant log-ratios; cross-validation studies should focus on multivariate quantities and pairwise log-ratio plots (Sect. 6.2).

Steps (2) and (3) can alternatively be applied to data transform via a particular log-ratio transformation. In this case, step (4) should also explore the fit of the model to the variation-variograms, and step (5) can be applied to the same log-ratio set as in step (2). This is the strategy followed in the paper, where all calculations were primarily done with the alr-transformed data.

B.2 Simulation

1. Apply a log-ratio transformation to the data, then transform the scores via multivariate Gaussian anamorphosis, such as the flow anamorphosis (Sect. 7.2)
2. Estimate direct and cross-variograms of the Gaussian scores
3. Fit a valid joint model to these variograms
4. Apply conditional simulation algorithms to produce simulations of the Gaussian scores
5. Transform the simulated Gaussian scores to log-ratio scores with the inverse Gaussian anamorphosis, then backtransform the log-ratio scores to compositions
6. Post-process simulations as desired, that is, produce point-wise estimates of non-linear quantities (Eq. 27), upscale them to block averages (Eqs. 29–31) or produce maps.

References

Aitchison J (1982) The statistical analysis of compositional data (with discussion). J R Stat Soc Ser B (Stat Methodol) 44:139–177

Aitchison J (1986) The statistical analysis of compositional data. Monographs on statistics and applied probability. Chapman & Hall Ltd., London (**Reprinted in 2003 with additional material by The Blackburn Press**)

Aitchison J, Barceló-Vidal C, Martín-Fernández JA, Pawlowsky-Glahn V (2000) Logratio analysis and compositional distance. Math Geol 32(3):217–275

Angerer T, Hagemann S (2010) The BIF-hosted high-grade iron ore deposits in the Archean Koolyanobbing Greenstone Belt, Western Australia: structural control on synorogenic- and weathering-related magnetite-, hematite- and goethite-rich iron ore. Econ Geol 105(3):917–945

Barceló-Vidal C (2003) When a data set can be considered compositional? In: Thió-Henestrosa S, Martin-Fernández JA (eds) Proceedings of CoDaWork'03, The 1st Compositional Data Analysis Workshop. Universitat de Girona

Barceló-Vidal C, Martín-Fernández JA (2016) The mathematics of compositional analysis. Austrian J Stat 45(4):57–71

Barnett RM, Manchuk JG, Deutsch CV (2014) Projection pursuit multivariate transform. Math Geosci 46(2):337–360

Bivand RS, Pebesma E, Gomez-Rubio V (2013) Applied spatial data analysis with R. Springer, New York

Chayes F (1960) On correlation between variables of constant sum. J Geophys Res 65(12):4185–4193

Chilès JP, Delfiner P (1999) Geostatistics. Wiley, New York

Cressie N (1991) Statistics for spatial data. Wiley, New York

Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, Barceló-Vidal C (2003) Isometric logratio transformations for compositional data analysis. Math Geosci 35(3):279–300

Filzmoser P, Hron K (2008) Outlier detection for compositional data using robust methods. Math Geosci 40(3):233–248

Geovariances (2017) Isatis geostatistical software. Avon, France

Griffin AC (1981) Structure and Iron Ore deposition in the Archaean Koolyanobbing Greenstone belt, Western Australia. Geol Soc Austra Spec Publ 7:429–438

Journel AG, Huijbregts CJ (1978) Mining geostatistics. Academic Press, London

Lark RM, Bishop TFA (2007) Cokriging particle size fractions of the soil. Eur J Soil Sci 58(3):763–774

Leuangthong O, Deutsch CV (2003) Stepwise conditional transformation for simulation of multiple variables. Math Geol 35(2):155–173

Mateu-Figueras G, Pawlowsky-Glahn V, Egozcue JJ (2011) The principle of working on coordinates. In: Pawlowsky-Glahn V, Buccianti A (eds) Compositional data analysis: theory and applications. Wiley, New York, pp 29–42

Mateu-Figueras G, Pawlowsky-Glahn V, Egozcue JJ (2013) The normal distribution in some constrained sample spaces. Stat Oper Res Trans 37(1):29–56

Matheron G (1963) Principles of geostatistics. Econ Geol 58:1246–1266
Matheron G (1965) Les variables régionalisées et leur estimation-une application de la théorie des fonctions aléatoires aux sciences de la nature. Masson et Cie, Paris
Matheron G (1971) The theory of regionalized variables and its applications. Technical Report C-5, École Nationale Supérieure des Mines de Paris, Centre de Geostatistique et de Morphologie Mathematique, Fontainebleau
Molayemat H, Torab FM, Pawlowsky-Glahn V, Hossein Morshedy A, Egozcue JJ (2018) The impact of the compositional nature of data on coal reserve evaluation, a case study in Parvadeh IV coal deposit, Central Iran. Int J Coal Geol 188:94–111. https://doi.org/10.1016/j.coal.2018.02.003
Morales Boezio MN (2010) Estudo das metodologias alternativas da geoestatística multivariada aplicadas a estimativa de teores de depósitos de ferro. Ph. D. thesis, Universidade Federal do Rio Grande do Sul
Morales Boezio MN, Costa JF, Koppe JC (2012) Cokrigagem de razões logarítmicas aditivas (alr) na estimativa de teores em depós de ferro (Cokriging of additive log-ratios (alr) for grade estimation in iron ore deposits. Rev Escol Minas 65:401–412
Mueller UA, Grunsky EC (2016) Multivariate spatial analysis of lake sediment geochemical data; Melville Peninsula, Nunavut, Canada. Appl Geochem 75:247–262. https://doi.org/10.1016/j.apgeochem.2016.02.007
Mueller U, Tolosana-Delgado R, van den Boogaart KG (2014) Simulation of compositional data: a nickel-laterite case study. In: Dimitrakopoulos R (ed) Advances in orebody modelling and strategic mine planning. AusIMM, Melbourne
Myers DE (1982) Matrix formulation of co-kriging. Math Geol 14(3):49–257
Pawlowsky V (1984) On spurious spatial covariance between variables of constant sum. Sci Terre Sér Inform 21:107–113
Pawlowsky V (1986) Räumliche Strukturanalyse und Schätzung ortsabhängiger Kompositionen mit Anwendungsbeispielen aus der Geologie. Ph.D. thesis, Freie Universität Berlin
Pawlowsky V (1989) Cokriging of regionalized compositions. Math Geol 21(5):513–521
Pawlowsky-Glahn V (2003) Statistical modelling on coordinates. In: Thió-Henestrosa S, Martin-Fernández JA (eds) Proceedings of CoDaWork'03, The 1st Compositional Data Analysis Workshop. Universitat de Girona
Pawlowsky-Glahn V, Burger H (1992) Spatial structure analysis of regionalized compositions. Math Geol 24(6):675–691
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(3):259–274
Pawlowsky-Glahn V, Egozcue JJ (2016) Spatial analysis of compositional data: a historical review. J Geochem Explor 164:28–32. https://doi.org/10.1016/j.gexplo.2015.12.010
Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R (2015) Modeling and analysis of compositional data. Wiley, Chichester
Pawlowsky-Glahn V, Olea RA (2004) Geostatistical analysis of compositional data. Studies in mathematical geology 7. Oxford University Press, Oxford
Pawlowsky-Glahn V, Olea RA, Davis JC (1995) Estimation of regionalized compositions: a comparison of three methods. Math Geol 27(1):105–127
Rossi ME, Deutsch CV (2014) Mineral resource estimation. Springer, New York
Sun XL, Wu YJ, Wang HL, Zhao YG, Zhang GL (2014) Mapping soil particle size fractions using compositional kriging, cokriging and additive log-ratio cokriging in two case studies. Math Geosci 46(4):429–443
Tjelmeland H, Lund KV (2003) Bayesian modelling of spatial compositional data. J Appl Stat 30(1):87–100
Tolosana-Delgado R (2006) Geostatistics for constrained variables: positive data, compositions and probabilities. Application to environmental hazard monitoring. Ph.D. thesis, Universitat de Girona
Tolosana-Delgado R, Egozcue JJ, Pawlowsky-Glahn V (2008) Cokriging of compositions: log ratios and unbiasedness. In: Ortiz JM, Emery X (eds) Geostatistics Chile 2008. Gecamin Ltd., Santiago, pp 299–308
Tolosana-Delgado R, Mueller U, van den Boogaart KG, Ward C (2013) Block cokriging of a whole composition. In: Costa JF, Koppe J, Peroni R (eds) Proceedings of the 36th APCOM international symposium on the applications of computers and operations research in the mineral industry. Fundacao Luiz Englert, Porto Alegre, pp 267–277

Tolosana-Delgado R, Mueller U, van den Boogaart KG, Ward C, Gutzmer J (2015) Improving processing by adaption to conditional geostatistical simulation of block compositions. J South Afr Inst Min Metall 115(1):13–26

Tolosana-Delgado R, Otero N, Pawlowsky-Glahn V (2005) Some basic concepts of compositional geometry. Math Geol 37(7):673–680

Tolosana-Delgado R, van den Boogaart KG (2013) Joint consistent mapping of high-dimensional geochemical surveys. Math Geosci 45(8):983–1004

Tolosana-Delgado R, van den Boogaart KG, Pawlowsky-Glahn V (2011) Geostatistics for compositions. In: Pawlowsky-Glahn V, Buccianti A (eds) Compositional data analysis: theory and applications. Wiley, New York, pp 73–86

van den Boogaart KG, Tolosana-Delgado R, Bren M (2018) Compositions: compositional data analysis package R package version 1.40-2

van den Boogaart KG, Tolosana-Delgado R (2013) Analysing compositional data with R. Springer, Heidelberg

van den Boogaart KG, Tolosana-Delgado R, Mueller U (2017) An affine equivariant multivariate normal score transform for compositional data. Math Geosci 49(2):231–252

Wackernagel H (2003) Multivariate geostatistics: an introduction with applications. Springer, Berlin

Walwoort DJ, de Gruijter JJ (2001) Compositional kriging: a spatial interpolation method for compositional data. Math Geol 33(8):951–966

Ward C, Mueller U (2012) Multivariate estimation using log ratios: a worked alternative. In: Abrahamsen P, Hauge R, Kolbjornsen O (eds) Geostatistics Oslo 2012. Springer, Berlin, pp 333–343

Ward C, Mueller U (2013) Compositions, log ratios and bias—from grade control to resource. In: Iron Ore 2013 shifting the paradigm. AUSIMM, Melbourne, pp 313–320