HAL
open science

A new class of α-transformations for the spatial analysis of Compositional Data

Lucia Clarotto, Denis Allard, Alessandra Menafoglio

To cite this version:

Lucia Clarotto, Denis Allard, Alessandra Menafoglio. A new class of α-transformations for the spatial analysis of Compositional Data. Spatial Statistics, 2022, 47, pp.100570. 10.1016/j.spasta.2021.100570 . hal-03379935v2

HAL Id: hal-03379935

https://hal.science/hal-03379935v2

Submitted on 20 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A new class of α-transformations for the spatial analysis of Compositional Data

Lucia Clarotto[a,*], Denis Allard[b], Alessandra Menafoglio[c]

[a]MINES ParisTech, PSL University, Centre de Geosciences, 77300 Fontainebleau, France

[b]Biostatistics and Spatial Processes (BioSP), INRAE, 84914 Avignon, France

[c]MOX, Department of Mathematics, Politecnico Milano, Milano, Italy

[*]Corresponding author: lucia.clarotto@mines-paristech.fr

Abstract

Georeferenced compositional data are prominent in many scientific fields and in spatial statistics. This work addresses the problem of proposing models and methods to analyze and predict, through kriging, this type of data. To this purpose, a novel class of α-transformations, named the Isometric α-transformation (α-IT), is proposed, which encompasses the traditional Isometric Log-Ratio (ILR) transformation. Similarly to other α-transformations existing in the literature, it is shown that the ILR is the limit case of the α-IT as α tends to 0 and that $\alpha = 1$ corresponds to a linear transformation of the data. Unlike the ILR, the proposed transformation accepts 0s in the compositions when $\alpha > 0$. Maximum likelihood estimation of the parameter α is established. Prediction using kriging on α-IT transformed data is validated on synthetic spatial compositional data, using prediction scores computed either in the geometry induced by the α-IT, or in the simplex. Application to land cover data shows that the relative superiority of the various approaches w.r.t. a prediction objective depends on whether the compositions contained any zero component. When all components are positive, the limit cases (ILR or linear transformations) are optimal for none of the considered metrics. An intermediate geometry, corresponding to the α-IT with maximum likelihood estimate, better describes the dataset in a geostatistical setting. When the amount of compositions with 0s is not negligible, some side-effects of the transformation gets amplified as α decreases, entailing poor kriging performances both within the α-IT geometry and for metrics in the simplex.

Keywords: Geostatistics, Kriging, Isometric Log-Ratio (ILR) transformation, Compositions with 0s.

1 Introduction

In statistics, compositional data are quantitative descriptions of the parts of some whole, conveying relative information. Mathematically, compositional data are represented by points in a simplex, i.e. vectors with non-negative coordinates whose sum is constant. In this paper, we are interested in geostatistics or spatial statistics for compositional data, which is the area of statistics developing methods to analyze and predict, through kriging, compositional data associated with spatial or spatio-temporal phenomena.

Georeferenced compositional datasets arise in varied fields of research, from geology to economics to chemistry to sociology, although most studies were historically concerned with topics in the geosciences. In this setting, it took a long time to find a solution to the problem of how to perform a proper statistical analysis of *closed* data – namely data with a constant-sum constraint – by taking into account the consequence of compositional constraints on correlations. Because some standard statistical techniques lose their applicability and classical interpretation when applied to compositional data, new techniques were needed. The first theoretically sound solution was proposed in the 1980's, when John Aitchison (Aitchison, 1986) built a consistent theory based on log-ratio transformations of compositional data. Later developments have shown that the mathematical foundation of a proper statistical analysis for this type of data is based on the definition of a specific geometry on the simplex, referred to as the *Aitchison geometry*. Based on it and on the Principles of Compositional Data Analysis (Egozcue et al., 2003), a relatively large body of literature has established a complete mathematical framework for statistical analysis which is nowadays widely accepted by the statistical community. Spatial statistics has also been widely treated in the compositional community (Pawlowsky-Glahn and Olea, 2004; Tolosana-Delgado et al., 2009; Tolosana-Delgado, 2006; Tolosana-Delgado et al., 2008, 2010, 2011; Tolosana-Delgado and van den Boogaart, 2013), by following the methods developed in the Aitchison geometry, i.e., by applying the log-ratio transformations that allow one to respect the Principles of Compositional Data Analysis. We refer to Pawlowsky-Glahn and Egozcue (2016) for a complete review of the historical evolution of spatial analysis of compositional data through the Aitchison geometry.

Even though the Aitchison's approach is nowadays mainstream in the analysis of compositional data, a number of authors have pushed forward alternative viewpoints, arguing that the choice of the appropriate method for the statistical data analysis should not be determined *a priori* from a set of mathematical principles, but that this choice should rather depend, at least in part, on the data (Scealy and Welsh, 2014). In this vein, Tsagris et al. (2011) proposed a new family of transformations, called α-transformations, parameterized by a constant α. This parameter allows one to control the *degree of transformation* applied to the data, ranging from a linear transformation ($\alpha = 1$) to a log-ratio transformation ($\alpha = 0$). In this setting, the parameter α is chosen in a data-driven manner, thus allowing one to render the approach application-specific. Note that the use of α-transformations also enables one to deal with the presence of 0s in the compositions, unlike the log-ratio approach which is only suitable for strictly positive compositions. Beside these aspects, the approach based on α-transformations proved effective in real studies, both in classification (Tsagris et al., 2016) and in regression (Tsagris, 2015). In the same vein, but in the somewhat different domain of correspondence analysis, Greenacre (2009b,a, 2011) proposed a similar power

transformation and proved that it tends to the Aitchison log-ratio transformation as the power parameter approaches 0.

This paper follows the line pioneered by Tsagris et al. (2011) and Greenacre (2009b) and aims to establish a methodological framework for the statistical analysis of spatial compositional data, while finding a balance between the data-driven approach of Tsagris et al. (2011) and that of the compositional community. In this vein, a novel set of transformations is considered, based on the concept of α-contrasts – which generalizes that of log-contrasts upon which the Aitchison geometry is based. We show that, similarly to Tsagris et al. (2011), our approach coincides with that of Egozcue et al. (2003) for specific choices of the parameter α, while attaining in general a more flexible framework than that of the Aitchison simplex. Besides, we shall also establish an explicit link between the covariance structure induced by the proposed class of transformations and that defined under the Aitchison geometry, opening broad perspectives of potential application on a wide range of covariance-based methods for exploratory and inferential data analyses. In fact, the approach we propose is very general and not limited to spatial datasets. However, for the sake of brevity, this work will mainly focus on the problem of spatial analysis of compositional data with the lens of spatial prediction (kriging), which shall drive the formulation of the transformation, and particularly the choice of the best" parameterization.

The remaining of this work is organized as follows: Section 2 introduces the theoretical concepts of compositional data, focusing on the log-ratio transformations, the spatial statistics methodologies applied in this field and the α-transformations proposed by Tsagris et al. (2011). Section 3 explores the new class of transformations considered in this paper, i.e., the Centered and Isometric α-transformations (α-CT and α-IT), by discussing their properties, especially in a geostatistical setting. A maximum likelihood estimation method is proposed which maximizes the Gaussianity of the transformed data and, as a consequence, the kriging performances. Section 4 is interested in the application of the α-IT to a simulated spatial dataset, in order to evaluate the improvement of this transformation over the classical Aitchison transformations. Section 5 conducts a geostatistical analysis of land cover data, following the approach analyzed throughout the previous sections. Special attention is given to the analysis of data in presence of 0-parts in the compositions. Results show that when all parts are positive, the limit cases (ILR or linear transformations) are optimal for none of the considered metrics. An intermediate geometry, corresponding to the α-IT with maximum likelihood estimate better describes the dataset in a geostatistical setting. When the amount of compositions with 0s is not negligible, some side-effects of the transformation gets amplified as α decreases, entailing poor kriging performances both within the α-IT geometry and for metrics in the simplex. Finally, Section 6 reviews the main points of the paper.

2 State-of-the-art analysis of spatial compositions

2.1 Compositional Data analysis in the Aitchison simplex

In this section we provide a brief overview of the key concepts underlying the analysis of compositional data through the Aitchison geometry. We refer the reader to, e.g., Pawlowsky-Glahn et al. (2015) for a deeper account on the subject.

A column vector, $\mathbf{x} = [x_1, \ldots, x_D]^\top$, is defined as a D-part composition when all its components are positive real numbers carrying only relative information. The sample space of compositional data is the simplex, defined as

$$\mathbb{S}^D = \{\mathbf{x} = [x_1, \ldots, x_D]^\top \mid x_i > 0,\ i = 1, \ldots, D;\ \sum_{i=1}^{D} x_i = \kappa\}, \tag{1}$$

where κ is an arbitrary constant which can be set to 1 without loss of generality. A D-part compositional vector $\mathbf{z} \in (0, \infty)^D$ can be represented as a point of the simplex $\mathbb{S}^D$ by use of the closure operator

$$C : C(\mathbf{z}) = \mathbf{x}, \quad x_i = \frac{z_i}{\sum_{i=1}^{D} z_i} \quad \forall i = 1, \ldots, D. \tag{2}$$

The perturbation of a composition $\mathbf{x} \in \mathbb{S}^D$ by a composition $\mathbf{y} \in \mathbb{S}^D$ is

$$\mathbf{x} \oplus \mathbf{y} = C\left([x_1 y_1, \ldots, x_D y_D]^\top\right).$$

The power transformation of a composition $\mathbf{x} \in \mathbb{S}^D$ by a constant $\alpha \in \mathbb{R}$ is

$$\alpha \odot \mathbf{x} = C\left([x_1^\alpha, \ldots, x_D^\alpha]^\top\right).$$

In particular, $(-1) \odot \mathbf{x} = C\left([1/x_1, \ldots, 1/x_D]^\top\right)$. The neg-perturbation is then defined in the following way: $\mathbf{x} \ominus \mathbf{y} = \mathbf{x} \oplus (-1 \odot \mathbf{y})$. The simplex $(\mathbb{S}^D, \oplus, \odot)$, equipped with the perturbation operation and the power transformation, is a vector space (Aitchison, 1986; Billheimer et al., 2001; Pawlowsky-Glahn and Egozcue, 2001). This implies that all properties of translation and scalar multiplication hold. Moreover, by adding an inner product, a norm and a distance on the vector space, the simplex becomes an Euclidean metric space (i.e., a finite-dimensional Hilbert space). To refer to the properties of $(\mathbb{S}^D, \oplus, \odot)$ as an Euclidean metric space, we call it *Aitchison geometry* on the simplex. We report here some definitions of the Aitchison geometry that will be useful for this work. The inner product of $\mathbf{x}, \mathbf{y} \in \mathbb{S}^D$ is

$$\langle \mathbf{x}, \mathbf{y} \rangle_a = \frac{1}{2D} \sum_{i=1}^{D} \sum_{j=1}^{D} \ln \frac{x_i}{x_j} \ln \frac{y_i}{y_j}, \tag{3}$$

and the norm of $\mathbf{x} \in \mathbb{S}^D$ is $\|\mathbf{x}\|_a = \langle \mathbf{x}, \mathbf{x} \rangle_a$. The Aitchison distance between $\mathbf{x}$ and $\mathbf{y} \in \mathbb{S}^D$ is

$$d_a(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} \ominus \mathbf{y}\|_a = \sqrt{\frac{1}{2D} \sum_{i=1}^{D} \sum_{j=1}^{D} \left(\ln \frac{x_i}{x_j} - \ln \frac{y_i}{y_j}\right)^2}. \tag{4}$$

Aitchison (1986) used the fact that for compositional data absolute quantities and units are irrelevant – as interest lies in relative proportions of the components measured – to introduce transformations based on log-ratios. We refer the reader to Aitchison (1986); Mateu-Figueras et al. (2011); Buccianti et al. (2006) and references therein for a detailed account. Here, we focus on the Centered Log-Ratio transform (CLR) and on the Isometric Log-Ratio transform (ILR) which will be necessary for the sequel.

The Centered Log-Ratio transformation (CLR) is defined as

$$\begin{aligned} \mathrm{clr} &: \mathbb{S}^D \to \mathbb{H} \subset \mathbb{R}^D \\ \mathrm{clr}(\mathbf{x}) &= \left[\ln \frac{x_1}{g(\mathbf{x})}, \cdots, \ln \frac{x_D}{g(\mathbf{x})}\right]^\top, \end{aligned} \tag{5}$$

where $g(\mathbf{x})$ is the geometric mean of $\mathbf{x}$, $g(\mathbf{x}) = \left(\prod_{i=1}^D x_i\right)^{1/D}$.

The CLR transform is an injective transformation between $\mathbb{S}^D$ and $\mathbb{R}^D$ which preserves distances, in the sense that $d_a(\mathbf{x}, \mathbf{y}) = d_e(\mathrm{clr}(\mathbf{x}), \mathrm{clr}(\mathbf{y}))$, with d_e the Euclidean distance. However, the CLR is characterized by a constraint on the transformed sample, as the sum of the components of $\mathrm{clr}(\mathbf{x})$ is 0 by definition. The transformed sample thus lie on a hyper-plane, $\mathbb{H}$, which goes through the origin of $\mathbb{R}^D$ and is orthogonal to the vector of ones $[1, \ldots, 1]^\top$. The CLR transformation cannot be directly associated with an orthogonal coordinate system in the simplex. For this reason, Egozcue et al. (2003) introduced a new transformation, called Isometric Log-Ratio (ILR), defined as follows.

Any vector $\mathbf{x} \in \mathbb{S}^D$ can be written as

$$\mathbf{x} = \bigoplus_{i=1}^{D} \ln x_i \odot \mathbf{w}_i,$$

where $\{\mathbf{w}_1, \ldots, \mathbf{w}_D\}$ are the generator vectors $\mathbf{w}_i = C(\exp(\mathbf{d}_i))$, $i = 1, \ldots, D$, where $\mathbf{d}_i$ is the unit vector of $\mathbb{R}^D$ associated to the ith coordinate. Omitting one vector of the generating system, a basis is obtained, e.g. $\{\mathbf{w}_1, \ldots, \mathbf{w}_{D-1}\}$, which is not orthonormal. However, a new basis, orthonormal with respect to the inner product, can be obtained using the Gram-Schmidt procedure, so that it is possible to express a composition $\mathbf{x} \in \mathbb{S}^D$ as

$$\mathbf{x} = \bigoplus_{i=1}^{D-1} x_i^* \odot \mathbf{e}_i, \quad x_i^* = \langle \mathbf{x}, \mathbf{e}_i \rangle_a,$$

where $\{\mathbf{e}_1, \ldots, \mathbf{e}_{D-1}\}$ is a generic orthonormal basis of the simplex $\mathbb{S}^D$ and $\langle \cdot, \cdot \rangle_a$ is the inner product in the Aitchison geometry of Equation (3).

The Isometric Log-Ratio (ILR) of a composition $\mathbf{x}$ is defined as

$$\begin{aligned} \mathrm{ilr} &: \mathbb{S}^D \to \mathbb{R}^{D-1} \\ \mathrm{ilr}(\mathbf{x}) &= \mathbf{x}^* = [\langle \mathbf{x}, \mathbf{e}_1 \rangle_a, \ldots, \langle \mathbf{x}, \mathbf{e}_{D-1} \rangle_a]^\top. \end{aligned}$$

The ILR is an isomorphism between $(\mathbb{S}^D, d_a)$ and $(\mathbb{R}^{D-1}, d_e)$ which preserves distances (i.e., it is an isometry), and it can be retrieved from the CLR transform with the following equality:

$$\mathrm{ilr}(\mathbf{x}) = \mathbf{H}_D \, \mathrm{clr}(\mathbf{x}),$$

where $\mathbf{H}_D$ is the $(D-1, D)$ Helmert matrix whose rows are $\mathrm{clr}(\mathbf{e}_i)$. $\mathbf{H}_D$ satisfies $\mathbf{H}_D \, \mathbf{H}_D^\top = \mathbf{I}_{D-1}$ and $\mathbf{H}_D^\top \, \mathbf{H}_D = \mathbf{G}_D$, where $\mathbf{G}_D = (\mathbf{I}_D - D^{-1} \, \mathbf{J}_D)$ is the D-dimensional centering matrix and $\mathbf{J}_D$ is the (D, D) matrix of ones.

2.2 Geostatistics for Compositional Data

Throughout this section, we consider a D-part composition $\mathbf{X}(s)$ with $X_i(s) > 0$, for $i = 1, \ldots, D$, $\forall s \in \mathcal{D} \subset \mathbb{R}^d$. We further assume that it is second-order stationary, i.e. the expected values of the pairwise log-ratios at each point s exist and do not depend on s, and the cross-covariances between every pairwise log-ratio at two different points s_1 and s_2 exist and depend only on the directional vector $h = s_2 - s_1$. Such a compositional random field is said to be *second-order Log-Ratio (LR) stationary*. Let $\{s_j : j = 1, \ldots, n\}$ be a set of n spatial locations in a spatial domain $\mathcal{D}$ containing georeferenced compositional data $\{\mathbf{X}(s_j)\}_{j=1}^n = \{[X_1(s_j), \ldots, X_D(s_j)]^\top\}_{j=1}^n$ in $\mathbb{S}^D$. Now, let $\mathbf{X}(s_0)$ be the unobserved vector of compositional data at the prediction location s_0. The task of predicting $\mathbf{X}(s_0)$ using linear combinations of $\{\mathbf{X}(s_j)\}_{j=1}^n$ is known as cokriging in the geostatistics literature. We refer to Wackernagel (2003) for a general introduction to multivariate geostatistics and to cokriging.

Direct implementation of cokriging on $\mathbf{X}(s_0)$ is impractical for several reasons. For instance, the associated covariance matrix must satisfy the closure relations induced by the constant sum constraint, $\sum_{j=1}^D C_{ij}(h) = 0$ for all $h \in \mathbb{R}^d$ and all $i = 1, \ldots, D$, which result in singular cokriging matrices. One way around would be to use cokriging on $(D-1)$ parts and deduce the last part from the closure property, but this approach also shows drawbacks. Predicted and simulated compositional vectors are expected to be elements of the simplex $\mathbb{S}^D$. Generally speaking, direct kriging or cokriging of the proportions, if achievable, cannot guarantee this property in all generality since kriging does not impose non-linear constraints such as positivity (Walvoort and de Gruijter, 2001).

In an ordinary cokriging setting, closure to 1 can be enforced by imposing identical kriging weights for all parts, since in this case

$$\sum_{i=1}^D X_i^\star(s_0) = \sum_{i=1}^D \sum_{j=1}^n \lambda_j X_i(s_0) = \sum_{j=1}^n \lambda_j \sum_{i=1}^D X_i(s_0) = \sum_{j=1}^n \lambda_j = 1,$$

where $[\lambda_1, \ldots, \lambda_n]^\top$ is the unique vector of weights and $X_i^\star(s_0)$ is the cokriging of $X_i(s_0)$. Allard and Marchant (2018) have shown that identical weights are obtained for all s_0 and all $\{\mathbf{X}(s_j)\}_{j=1}^n$ if and only if the multivariate covariance model for $\mathbf{X}(s)$ is proportional, i.e. it is the product of a covariance matrix and a single spatial covariance function. On the other hand, imposing positivity is possible, but requires quadratic programming (Walvoort and de Gruijter, 2001).

Moreover, working directly on $\mathbf{X}(s_0)$ does not account for the relative nature of parts which, following the above developments, should be properly acknowledged. To sum up, even though direct cokriging is in theory possible, it is not guaranteed to be optimal. In Tolosana-Delgado (2006), an approach for geostatistics of compositional data based on the application of the principle of working in coordinates using Isometric Log-Ratio representations is presented in detail. For a recent review, see Pawlowsky-Glahn and Egozcue (2016). Following this approach, compositional data belonging to $\mathbb{S}^D$ are first transformed to a set of $(D-1)$ unbounded scores by means of the ILR. Then, the multivariate geostatistical techniques (e.g., covariance modeling, kriging, stochastic simulation) are applied to the scores. Finally, the resulting interpolated or simulated scores are back-transformed to obtain values in $\mathbb{S}^D$. Since this technique applies a transformation to the compositional dataset

before any geostatistical method, the spatial structure must be defined on the transformed dataset (Pawlowsky-Glahn and Olea, 2004), as we describe in the following.

Let $\mathbf{X}(s)$ be a second-order LR stationary compositional random field. The *variation matrix*, $\mathbf{T}(h) = [\tau_{ij}(h)]_{i,j=1}^{D}$ of $\mathbf{X}(s)$ is the (D, D) matrix whose elements are the autocovariances of the corresponding log-ratios

$$\tau_{ij}(h) = \mathrm{Cov}\left[\ln\left(\frac{X_i(s)}{X_j(s)}\right), \ln\left(\frac{X_i(s+h)}{X_j(s+h)}\right)\right].$$

The variation matrix is symmetric in the indices and in h for any two components of $\mathbf{X}(s)$.

The covariance matrix of the Centered Log-Ratio (CLR) transformed random field $\mathbf{Z}_{\mathrm{clr}}(s) = \mathrm{clr}(\mathbf{X}(s))$, called the *CLR cross-covariance matrix*, is the (D, D) matrix $\mathbf{\Xi}(h) = [\xi_{ij}(h)]_{i,j=1}^{D}$ which is the covariance between the elements of the CLR transformed data and the same elements lagged by h

$$\xi_{ij}(h) = \mathrm{Cov}\left[\mathbf{Z}_{\mathrm{clr}}(s), \mathbf{Z}_{\mathrm{clr}}(s+h)\right]_{ij} = \mathrm{Cov}\left[\ln\left(\frac{X_i(s)}{g(\mathbf{X}(s))}\right), \ln\left(\frac{X_j(s+h)}{g(\mathbf{X}(s+h))}\right)\right].$$

The covariance matrix of the Isometric Log-Ratio (ILR) transformed data $\mathbf{Z}_{\mathrm{ilr}}(s) = \mathrm{ilr}(\mathbf{X}(s))$, called the *ILR cross-covariance matrix* (or *coordinate cross-covariance matrix* in Tolosana-Delgado (2006)), is the $(D-1, D-1)$ matrix $\mathbf{\Phi}(h) = [\phi_{ij}(h)]_{i,j=1}^{D-1}$ with

$$\phi_{ij}(h) = \mathrm{Cov}\left[\mathbf{Z}_{\mathrm{ilr}}(s), \mathbf{Z}_{\mathrm{ilr}}(s+h)\right]_{ij}.$$

These cross-covariance matrices are related to each other in the following way (Pawlowsky-Glahn and Olea, 2004):

$$\mathbf{\Xi}(h) + \mathbf{\Xi}^{\top}(h) = -\,\mathbf{G}_D\,\mathbf{T}(h)\,\mathbf{G}_D^{\top}, \tag{6}$$

where $\mathbf{G}_D$ is the centering matrix of dimension D defined in Section 2.1. Moreover, the following relation between $\mathbf{\Xi}(h)$ and $\mathbf{\Phi}(h)$ can be proven by linearity: the ILR cross-covariance matrix $\mathbf{\Phi}(h)$ and the CLR cross-covariance matrix $\mathbf{\Xi}(h)$ satisfy

$$\mathbf{\Phi}(h) = \mathbf{H}_D\,\mathbf{\Xi}(h)\,\mathbf{H}_D^{\top}.$$

2.3 The α-transformation

In the last few years, Tsagris et al. (2011) have proposed a new class of transformations, called *α-transformations*, that encompasses the Aitchison transformation in the sense that it allows to deal with 0-values in the compositions and that it retrieves the Aitchison geometry when α tends to 0. The α-transformation, for $\alpha \neq 0$, of any compositional vector $\mathbf{x} \in \mathbb{S}^D$ is the mapping

$$\begin{aligned} A_\alpha &: \mathbb{S}^D \to \mathbb{R}^{D-1} \\ \mathbf{z} = A_\alpha(\mathbf{x}) &= \alpha^{-1}\,\mathbf{H}_D\,(D\,\mathbf{u}_\alpha(\mathbf{x}) - \mathbf{1}_D), \qquad \mathbf{u}_\alpha(\mathbf{x}) = C(\mathbf{x}^\alpha), \end{aligned} \tag{7}$$

where $C(\cdot)$ is the closure operator defined in Equation (2) and the powering is applied component-wise.

Detailed analyses of this transformation are available in Tsagris et al. (2011), Tsagris (2015) and Tsagris et al. (2016). Two important benefits of the α-transformation are that it is well-defined for

any $\alpha > 0$ for compositions containing 0s and that it tends to the ILR transform as α approaches 0.

The α-transformation was shown to yield good results when used to analyze compositional datasets, especially for classification problems (Tsagris et al., 2016). However, when analyzing the peculiarities of this transformation, Tsagris and Stewart (2020) clarified that the images of the α-transformation lie in a codomain that has a simplex shape, except for $\alpha = 0$; in this case the codomain is the entire space $\mathbb{R}^{D-1}$. The simplex shape is due to the fact that a closure operator is introduced in the transformation, which allows for a closed form for the inverse transform. Beside this convenient property, the presence of the closure operation (2) does not seem to be mathematically necessary. Moreover, when embedded in a geostatistical setting, the spatial covariance structure is no longer explicitly connected to the variation matrix as in Equation (6). In order to overcome these limitations, we introduce a new family of α-transformations in the next Section, that we name Centered and Isometric α-transformations.

3 The Centered and Isometric α-transformations for spatial Compositional Data

3.1 The Centered and Isometric α-transformations

We here introduce a new class of transformations called *Centered* and *Isometric α-transformations* (α-CT and α-IT). Let us define $\mathbb{S}_0^D = \{\mathbf{x} = [x_1, \ldots, x_D]^\top \mid x_i \geq 0,\ i = 1, \ldots, D;\ \sum_{i=1}^D x_i = 1\}$ the simplex that admits one or more 0-values in its components.

Definition 1. Let $\alpha > 0$. The Centered α-transformation (α-CT) of a compositional vector $\mathbf{x} \in \mathbb{S}_0^D$ is the mapping $A_{\alpha-CT} : \mathbb{S}_0^D \to \mathbb{R}^D$

$$\mathbf{u}_{\alpha-CT} = A_{\alpha-CT}(\mathbf{x}) = \alpha^{-1}\, \mathbf{G}_D\, \mathbf{x}^\alpha, \tag{8}$$

where $\boldsymbol{G}_D$ is the (D, D) centering matrix defined above. The Isometric α-transformation (α-IT) of a compositional vector $\mathbf{x} \in \mathbb{S}_0^D$ is the mapping $A_{\alpha-IT} : \mathbb{S}_0^D \to \mathbb{R}^{D-1}$

$$\mathbf{z}_{\alpha-IT} = A_{\alpha-IT}(\mathbf{x}) = \alpha^{-1}\, \mathbf{H}_D\, \mathbf{G}_D\, \mathbf{x}^\alpha = \alpha^{-1}\, \mathbf{H}_D\, \mathbf{x}^\alpha, \tag{9}$$

where $\mathbf{H}_D$ is the $(D-1, D)$ Helmert matrix defined in Section 2.1.

To better understand the α-CT, it is useful to make the link between Equation (8) and the CLR transformation (5) more explicit. Indeed, the α-CT of a compositional vector $\mathbf{x}$ in $\mathbb{S}^D$ reads

$$\mathbf{u}_{\alpha-CT} = \left[\alpha^{-1}\left(x_1^\alpha - \frac{1}{D}\sum_{i=1}^D x_i^\alpha\right), \ldots, \alpha^{-1}\left(x_D^\alpha - \frac{1}{D}\sum_{i=1}^D x_i^\alpha\right)\right]^\top, \tag{10}$$

whereas the CLR transform of $\mathbf{x}$ in (5) is equivalent to

$$\mathrm{clr}(\mathbf{x}) = \left[\left(\ln(x_1) - \frac{1}{D}\sum_{i=1}^D \ln(x_i)\right), \ldots, \left(\ln(x_D) - \frac{1}{D}\sum_{i=1}^D \ln(x_i)\right)\right]^\top. \tag{11}$$

In fact, similarly as for the CLR transform, the α-CT of $\mathbf{x}$ operates a centering of a (power) transformation of $\mathbf{x}$ with respect to its average value, thus yielding a transformed vector $\mathbf{u}_{\alpha-CT}$ characterized by a zero-sum constraint. Hence, both $\mathbf{u}_{\alpha-CT}$ and $\mathrm{clr}(\mathbf{x})$ lie on the same hyper-plane $\mathbb{H} \subset \mathbb{R}^D$ defined in Section 2.1. Analogously as for the ILR transform, the α-IT then identifies a set of $(D-1)$ coordinates over an orthonormal basis of $\mathbb{H}$. The following Proposition sheds further light on the link between the proposed class of transformations and those in use under the Aitchison geometry.

Proposition 1. *The CLR transform is retrieved from the Centered α-transformation when $\alpha \to 0$ and the ILR transform is retrieved from the Isometric α-transformation when $\alpha \to 0$, provided that $\mathbf{x} \in \mathbb{S}^D$.*

Proof. Since both the ILR transform and the α-IT are defined as the multiplication of the Helmert matrix $\mathbf{H}_D$ by the CLR transform and the α-CT respectively, it is sufficient to prove that the α-CT tends to the CLR transform when $\alpha \to 0$. From the definition of the geometric mean, we get $\ln g(\mathbf{x}) = D^{-1} \sum_{i=1}^{D} \ln x_i$, which does exist since $\mathbf{x} \in \mathbb{S}^D$. Notice that $\mathbf{G}_D\, \mathbf{x}^\alpha = \mathbf{x}^\alpha - m_\alpha(\mathbf{x}) \mathbf{1}_D$, where $\mathbf{1}_D$ is a D-vector of ones and where the α-mean $m_\alpha(\mathbf{x})$ is $m_\alpha(\mathbf{x}) = 1/D \sum_{i=1}^{D} x_i^\alpha$. From (8) we get

$$\begin{aligned}\lim_{\alpha\to 0} u_{i,\alpha-CT} &= \lim_{\alpha\to 0} \frac{x_i^\alpha - m_\alpha}{\alpha} = \lim_{\alpha\to 0} \left(\frac{x_i^\alpha - 1}{\alpha} - \frac{m_\alpha(\mathbf{x}) - 1}{\alpha} \right) \\ &= \ln(x_i) - \ln(g(\mathbf{x})) = \ln\left(\frac{x_i}{g(\mathbf{x})}\right) = \mathrm{clr}(\mathbf{x})_i.\end{aligned}$$

Hence, we have

$$\lim_{\alpha\to 0} \mathbf{u}_{\alpha-CT} = \mathrm{clr}(\mathbf{x})$$

and

$$\lim_{\alpha\to 0} \mathbf{z}_{\alpha-IT} = \lim_{\alpha\to 0} \mathbf{H}_D\, \mathbf{u}_{\alpha-CT} = \mathbf{H}_D\, \mathrm{clr}(\mathbf{x}) = \mathrm{ilr}(\mathbf{x}).$$

□

This result establishes that the α-CT and the α-IT are a generalization of the CLR and ILR transform, respectively. Notice that, when $\alpha = 1$, these transformations boil down to linear transformations of the compositions. On $[0, 1]$, the parameter α thus offers a modeling flexibility that allows to interpolate smoothly between linear and log-transformations and therefore, hopefully, better adapt to the data. Other positive values of α are also possible. Negative values of α can also be considered for the α-IT and α-CT, provided that the compositions are in $\mathbb{S}^D$, i.e., that no 0-values occur. From now on, we will restrict the use of the α-CT and α-IT to $\alpha \geq 0$.

It is worth emphasizing that the α-CT and the α-IT (with $\alpha > 0$) can be applied to any composition of $\mathbb{S}_0^D$, including those lying on the border of the simplex. This point is an important improvement with respect to the usual CLR and ILR transforms that cannot be applied to compositions with one or more 0s. This point will be further discussed when analyzing the Copernicus Land Cover dataset in Section 5.3.

3.2 Inverse of the Isometric α-transformation

Let us consider a vector $\mathbf{z}$ belonging to the codomain of $A_{\alpha-IT}$, with $\alpha > 0$. In order to obtain the composition $\mathbf{x}$ such that $\mathbf{z} = A_{\alpha-IT}(\mathbf{x})$, one must solve Equation (9). Multiplying both sides of the equation by $\mathbf{H}_D^\top$, we get that the composition $\mathbf{x}$ solves

$$\mathbf{H}_D^\top \mathbf{z} = \alpha^{-1} \mathbf{G}_D \mathbf{x}^\alpha, \tag{12}$$

owing to the fact that $\mathbf{H}_D^\top \mathbf{H}_D = \mathbf{G}_D$. Since $\mathbf{G}_D$ is not invertible, Equation (12) cannot be solved directly, but the solution is unique. Indeed, let us consider that $\mathbf{x}$ and $\mathbf{y}$ are solutions to (12). Then $\mathbf{G}_D \mathbf{x}^\alpha = \mathbf{G}_D \mathbf{y}^\alpha$ if and only if $y_i^\alpha = x_i^\alpha + c$, for $i = 1, \ldots, D$, with $c = m_\alpha(\mathbf{y}) - m_\alpha(\mathbf{x})$. Without loss of generality, let us suppose that $c \geq 0$. Given that both $\mathbf{x}$ and $\mathbf{y}$ belong to $\mathbb{S}^D$, and that $y_i = (x_i^\alpha + c)^{1/\alpha}$, it must hold that

$$1 = \sum_{i=1}^{n} (x_i^\alpha + c)^{1/\alpha} = \sum_{i=1}^{n} x_i.$$

The latter can only be verified if $c = 0$, since for $\mathbf{x} \in \mathbb{S}^D$, $c \geq 0$ and $\alpha > 0$, $(x_i^\alpha + c)^{1/\alpha} \geq x_i$ for $i = 1, ..., D$. It follows that $\mathbf{y} = \mathbf{x}$ and that the solution is unique.

In order to solve (12), let us define $Q_{\mathbf{z}}(\mathbf{y}) = \left\| \mathbf{H}_D^\top \mathbf{z} - \alpha^{-1} \mathbf{G}_D \mathbf{y}^\alpha \right\|$. Then,

$$\mathbf{x} = A_{\alpha-IT}^{-1}(\mathbf{z}) = \arg\min_{\mathbf{y} \in \mathbb{S}_0^D} Q_{\mathbf{z}}(\mathbf{y}).$$

When $\mathbf{z}$ belongs to the codomain of $A_{\alpha-IT}$, the minimum of $Q_{\mathbf{z}}$ is 0 and $\mathbf{x} \in \mathbb{S}^D$. Otherwise, the minimum $Q_{\mathbf{z}}$ is larger than 0 and it is achieved on the border of $\mathbb{S}_0^D$, i.e., with at least one part of $\mathbf{x}$ being equal to 0. In practice, the minimum is found by using the function `nlminb` in `R` on $\mathbf{w}$, with $\mathbf{y}(\mathbf{w}) = \exp \mathbf{w} / C(\exp \mathbf{w})$, thus guaranteeing the positivity of all components of $\mathbf{y}$ and hence of $\mathbf{x}$.

The α-IT only admits a numerical form for the inverse transform, but, as seen above, this does not hamper its use. If we wanted to have a closed-form of the inverse transformation, but still admit 0-values in the compositions, we should use either the α-transformation (7) introduced in Tsagris et al. (2011) (which admits an explicit inverse as mentioned in Section 2.3) or the ALR Box-Cox transformation $\mathbf{z}_{\text{alr}-BC} = [z_{i,\text{alr}-BC}]^\top$, introduced by Barcel et al. (1996), inspired by the ALR transform, $\text{alr}(\mathbf{x}) = \left[\ln \frac{x_1}{x_D}, \cdots, \ln \frac{x_{D-1}}{x_D}\right]^\top$ (Aitchison, 1986), and defined, for each $i = 1, \ldots, D-1$, as

$$z_{i,\text{alr}-BC} = \frac{(x_i/x_D)^\alpha - 1}{\alpha}. \tag{13}$$

The classical ALR transform is recovered from Equation (13) as $\alpha \to 0$. Although the two proposed Box-Cox-like transformations admit an inverse transform, the α-transformation seems harder than necessary, since it introduces the closure operator (2) in the direct transform, and the ALR Box-Cox transform has the same issue of the classical ALR transform: it depends on the chosen denominator x_D and cannot deal with 0-values in x_D.

One may readily see similarities between the α-CT and the Box-Cox transformation, defined for any vector $\mathbf{x}$ in $\mathbb{R}^D$ as $\mathbf{u}_{BC} = A_{BC}(\mathbf{x}) = \frac{\mathbf{x}^\alpha - \mathbf{1}_D}{\alpha}$. The difference between $\mathbf{u}_{BC}$ and $\mathbf{u}_{\alpha-CT}$ relies in the constant which is subtracted to $\mathbf{x}^\alpha$, namely $\mathbf{1}_D$ in the case of the Box-Cox and the α-mean

$m_\alpha(\mathbf{x})\,\mathbf{1}_D$ for the α-CT. However, when multiplying $\mathbf{u}_{BC}$ and $\mathbf{u}_{\alpha-CT}$ by the Helmert matrix $\mathbf{H}_D$, the transformed composition is projected on the hyper-plane orthogonal to vectors colinear to $\mathbf{1}_D$. Hence, the Isometric coordinates $\mathbf{H}_D\,\mathbf{u}_{BC}$ and $\mathbf{H}_D\,\mathbf{u}_{\alpha-CT}$ coincide.

Notice that a number of analyses are invariant on the rescaling of the data by a real constant c (e.g., clustering). In these cases, having fixed α, the results of the analysis with the α-IT are equivalent to those obtained with the transformation $\mathbf{H}_D\,\mathbf{x}^\alpha$. A similar argument holds also with the α-transformation by Tsagris et al. (2011). However, the consistency of the family of α-IT with the CoDa approach is preserved only when the factor α^{-1} is included in the definition, precisely as in Definition 1.

In Figure 1, we plot the codomains of the α-IT (with a 'shield' shape) along with the codomains of the α-transformation (with a triangular shape), when these transformations are applied on the 3-dimensional simplex $\mathbb{S}^3$ for different values of α. When $\alpha = 0$, the codomains of the two transformations coincide with $\mathbb{R}^2$. For the other values of α, one can notice that the codomains of the α-IT appear 'smaller in size' than those of the α-transformation. This just reflects on a different scale for the transformed data if using the α-IT or the α-transformation, which has no relevant impact on their usability in practice.

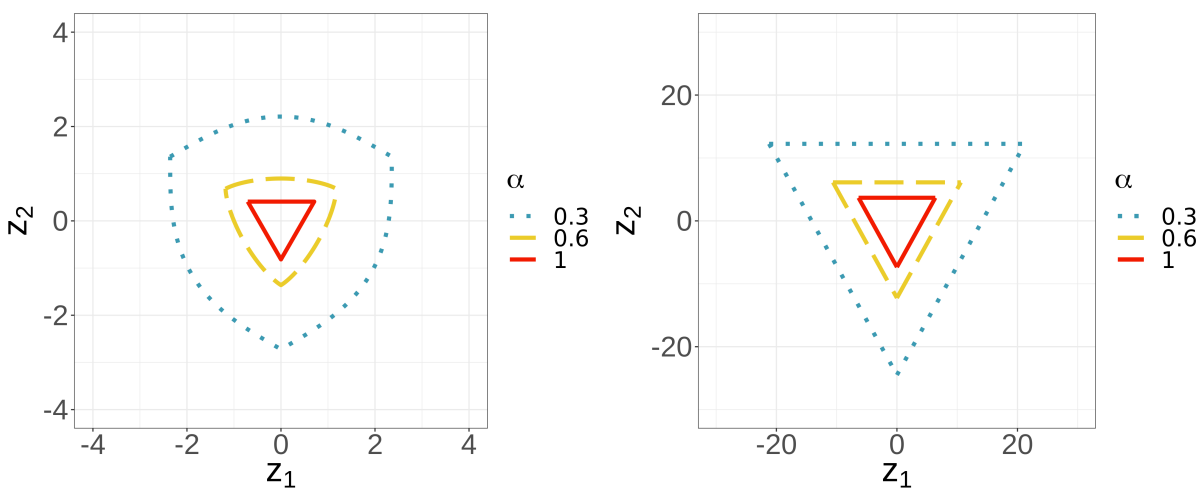

Figure 1: Codomains of the α-IT (left) and the α-transformation (right) for different values of α.

3.3 The α-IT-metric

The Isometric α-transformation of Equation (8) leads to a natural distance between observations $\mathbf{x}, \mathbf{y} \in \mathbb{S}_0^D$, denoted by $d_{\alpha-IT}(\mathbf{x}, \mathbf{y})$, and hereafter referred to as the *α-IT-metric*. For all values of $\alpha > 0$, $d_{\alpha-IT}(\mathbf{x}, \mathbf{y})$ is defined as

$$d_{\alpha-IT}(\mathbf{x}, \mathbf{y}) = \|A_{\alpha-IT}(\mathbf{x}) - A_{\alpha-IT}(\mathbf{y})\|\,, \tag{14}$$

where $\|\cdot\|$ denotes the usual Euclidean distance (in $\mathbb{R}^{D-1}$). When $\alpha \to 0$ we retrieve the Aitchisons distance (4), while when $\alpha = 1$ we retrieve the Euclidean distance $d_1(\mathbf{x}, \mathbf{y}) = [\sum_{i=1}^D (x_i - 1/D - y_i + 1/D)^2]^{1/2} = [\sum_{i=1}^D (x_i - y_i)^2]^{1/2}$.

Note that one can define measures of central tendency based on the α-IT-metric, using the concept of Fréchet mean – similarly as in Pawlowsky-Glahn and Egozcue (2001) and Tsagris et al. (2011). Indeed, a measure of central tendency for a compositional dataset $\mathbf{x}_1, ..., \mathbf{x}_n \in \mathbb{S}_0^D$ can be defined as

$$\mathbf{m} = \arg\min_{\boldsymbol{\mu} \in \mathbb{S}_0^D} \left\{ \frac{1}{n} \sum_{i=1}^{n} d_{\alpha-IT}(\mathbf{x}_i, \boldsymbol{\mu}) \right\}. \tag{15}$$

This is equivalent to back-transforming the sample mean of the transformed data, that is

$$\mathbf{m} = A_{\alpha-IT}^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} A_{\alpha-IT}(\mathbf{x}_i) \right). \tag{16}$$

3.4 Spatial structure of the Centered and Isometric α-transformations

As already mentioned, the classical way of performing a geostatistical analysis is to work on the coordinates of the transformation (Mateu-Figueras et al., 2011; Pawlowsky-Glahn et al., 2015). As we have seen in Section 2.2, we can directly relate the cross-covariance matrix of the ILR transformed data with the variation matrix of the composition. This section aims at deriving analogous results for the Isometric α-transformation.

Definition 2. Consider a second-order LR stationary compositional random field $\mathbf{X}(s)$. The α-IT *covariance matrix* is the $(D-1, D-1)$ matrix $\mathbf{\Phi}_\alpha(h) = [\phi_{\alpha,ij}(h)]_{i,j=1}^{D-1}$, given by

$$\mathbf{\Phi}_\alpha(h) = \mathrm{Cov}\left[\mathbf{Z}_{\alpha-IT}(s), \mathbf{Z}_{\alpha-IT}(s+h)\right] = \mathbf{H}_D \, \mathbf{\Xi}_\alpha(h) \, \mathbf{H}_D^\top, \tag{17}$$

where $\mathbf{Z}_{\alpha-IT}(s)$ is the Isometric α-transformation of $\mathbf{X}(s)$ and $\mathbf{\Xi}_\alpha(h) = [\xi_{\alpha,ij}(h)]_{i,j=1}^{D}$ is the α-*CT covariance matrix* whose elements are

$$\xi_{\alpha,ij}(h) = \mathrm{Cov}\left[u_{i,\alpha-CT}(s), u_{j,\alpha-CT}(s+h)\right]. \tag{18}$$

In Equation (18), $u_{i,\alpha-CT}$ and $u_{j,\alpha-CT}$ are two components of the vector of the Centered α-transformation of $\mathbf{X}(s)$ (Equation (8)).

Proposition 2. *The CLR covariance matrix $\mathbf{\Xi}(h) = [\xi_{ij}(h)]_{i,j=1}^{D}$ can be retrieved by taking the limit for $\alpha \to 0$ of the α-CT covariance matrix $\mathbf{\Xi}_\alpha(h)$ of Equation (18). The ILR covariance matrix $\mathbf{\Phi}(h) = [\phi_{ij}(h)]_{i,j=1}^{D-1}$ can be retrieved by taking the limit for $\alpha \to 0$ of the α-IT covariance matrix $\mathbf{\Phi}_\alpha(h)$ of Equation (17).*

Proof. $\mathbf{\Phi}(h)$ (respectively $\mathbf{\Phi}_\alpha(h)$) is the multiplication of the Helmert matrix $\mathbf{H}_D$ and its transpose by the (D, D) matrix $\mathbf{\Xi}(h)$ (respectively $\mathbf{\Xi}_\alpha(h)$). Hence, it is sufficient to verify that the formula of the CLR covariance matrix $\mathbf{\Xi}(h)$ is retrieved by taking the limit of $\mathbf{\Xi}_\alpha(h)$ for $\alpha \to 0$:

$$\begin{aligned}
\lim_{\alpha \to 0} \xi_{\alpha,ij}(h) &= \mathrm{Cov}\left[\lim_{\alpha \to 0} u_{i,\alpha-CT}(s), \lim_{\alpha \to 0} u_{j,\alpha-CT}(s+h)\right] \\
&= \mathrm{Cov}\left[\ln X_i(s) - \ln g(\mathbf{X}(s)), \ln X_j(s+h) - \ln g(\mathbf{X}(s+h))\right] \\
&= \mathrm{Cov}\left[\ln \frac{X_i(s)}{g(\mathbf{X}(s))}, \ln \frac{X_j(s+h)}{g(\mathbf{X}(s+h))}\right] = \xi_{ij}(h),
\end{aligned}$$

where we use the fact that $\lim_{\alpha\to 0} h_\alpha(\mathbf{x}) = \ln h_0(\mathbf{x}) = \ln g(\mathbf{x})$. Finally,

$$\lim_{\alpha\to 0} \phi_{\alpha,ij}(h) = \phi_{ij}(h).$$

□

Proposition 2 shows that $\mathbf{\Xi}_\alpha(h)$ is a generalization of the CLR covariance matrix. An analogous result holds for the α-IT covariance matrix $\mathbf{\Phi}_\alpha(h)$ with respect to the ILR covariance matrix $\mathbf{\Phi}(h)$.

When analyzing Equation (18), the difference between $u_{i,CT}(s) - u_{j,CT}(s+h)$ that is involved in the computation of the covariance matrix can be interpreted as a contrast; as α tends to 0, this contrast tends to a contrast of logs, i.e., a log-ratio. This observation echoes with the need for working on ratios put forward by Aitchison (1986) and by the compositional community in Pawlowsky-Glahn and Olea (2004), Tolosana-Delgado and van den Boogaart (2013) and Tolosana-Delgado et al. (2009). One could argue that, more generally, one needs to work on α-contrasts as those reported in Equation (18). In this way, not only the Aitchison geometry and in particular the ILR transformation leads to a well-defined spatial structure, but also the geometry derived from the Isometric α-transformation.

3.5 Geostatistics with the Isometric α-transformation

In the spatial analysis of compositional data, we follow the approach of the compositional community which makes use of the 'Principle of working in coordinates' (Mateu-Figueras et al., 2011), but we ground the analysis upon the α-IT. Having transformed the data, we apply prediction methods as cokriging within the Euclidean codomain of the α-IT. The results obtained in $\mathbb{R}^{D-1}$ are finally back-transformed to the simplex through numerical inversion of the α-IT (see Section 3.2). We discuss on the choice of the value of α to use in the α-IT in Section 3.6.

An important remark is that the restriction of the codomain of the α-IT with $\alpha > 0$ to a subset of $\mathbb{R}^{D-1}$ (see Figure 1) does not reflect on serious limitations for its use in practice. One could argue that sometimes the geostatistical predictions would fall out of the codomain of the α-IT, implying the actual impossibility to back-transform the results. Although possible in principle, this would rarely happen in kriging – and, in fact, it never happened during the numerical experiments carried out in this work. Recall that ordinary kriging is an interpolation technique based on a linear combination of the data with coefficients adding up to 1. When all weights are positive, the result of kriging is a convex combination of the data – and it is thus guaranteed to lie within the 'shield' domain of the α-IT, for all α. When some weights are negative, kriging could give interpolation results out of the convex hull of the data. This can be seen as a limitation of the approaches based on the α-IT (as well as of the α-transformation of Tsagris et al. (2011)), which however can be faced, e.g., by constraining the kriging weights to be positive (Cressie, 1993). This point is not investigated further in this paper, as predictions outside the codomain have never been observed.

3.6 Estimating the parameter α

Kriging is the Best Linear Unbiased predictor in the L^2 sense and it is the overall optimal L^2 predictor when the data being analyzed are Gaussian (Cressie, 1993). In order to achieve good

prediction performances, the parameter α of the transformation will thus be estimated using maximum likelihood, assuming that the transformed values are independent multivariate Gaussian vectors. We first consider that all compositions have positive parts, i.e., $\mathbf{x}_k \in \mathbb{S}^D, k = 1, \ldots, n$. Compositional vectors with at least one null part will be considered later. Since it is equivalent and more convenient to maximize the log-likelihood, the maximum likelihood estimator is thus $\hat{\alpha} = \arg\max_\alpha \mathcal{L}(\alpha; \mathbf{x}_1, \ldots, \mathbf{x}_n)$, with

$$\mathcal{L}(\alpha; \mathbf{x}_1, \ldots, \mathbf{x}_n) = -\frac{n}{2} \ln |\hat{\boldsymbol{\Sigma}}| - \frac{1}{2} \sum_{k=1}^{n} (\mathbf{z}_k - \hat{\boldsymbol{\mu}})^\top \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{z}_k - \hat{\boldsymbol{\mu}}) + \sum_{k=1}^{n} \ln |\boldsymbol{J}(\mathbf{x}_k)|, \tag{19}$$

where $|\boldsymbol{J}(\mathbf{x})|$ is the determinant of the Jacobian of the transformation $A_{\alpha-IT}$ and $\mathbf{z}_k = A_{\alpha-IT}(\mathbf{x}_k)$ is a $(D-1)$ Gaussian vector vector with expectation $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ – both being common to all the observations because of stationarity. The Jacobian $\boldsymbol{J}(\mathbf{x})$ is the $(D-1, D-1)$ matrix whose elements are

$$J_{ij}(\mathbf{x}) = \frac{\partial z_i}{\partial x_j} = H_{ij} x_j^{\alpha-1} - H_{iD}(1 - x_1 \cdots - x_{D-1})^{\alpha-1} = H_{ij} x_j^{\alpha-1} - H_{iD} x_D^{\alpha-1}, \tag{20}$$

with $1 \leq i, j \leq D-1$. Here, the index k was dropped for convenience of notation. In (19), $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are the ML estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ respectively, given the multi-Gaussian vectors $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$.

We now consider the case where some of the components of $\mathbf{x}$ are equal to 0, implying that the Jacobian in Equation (20) is undefined when $\alpha \in [0, 1)$. In this case, the D dimensional Gaussian distribution is actually concentrated on a simplex with dimension $D' < D$. The proper way of dealing with this situation is thus to consider that the composition lies in $\mathbb{S}^{D'}$ and that the α-IT vector $\mathbf{z}'$ is given by $\mathbf{z}' = \alpha^{-1} \mathbf{H}_{D'} \mathbf{x}'$, where $\mathbf{x}'$ is the D' dimensional subvector of $\mathbf{x}$ with positive parts only.

The full log-likelihood to be maximized is then

$$\mathcal{L}_0 = \mathcal{L} + \sum_{i=1}^{D} \mathcal{L}_i + \sum_{i=1}^{D} \sum_{j=i+1}^{D} \mathcal{L}_{i,j} + \cdots . \tag{21}$$

where $\mathcal{L}$ is the log-likelihood of the data lying in $\mathbb{S}^D$, given in Equation (19). $\mathcal{L}_i$ is the log-likelihood of the $(D-1)$-dimensional data obtained by removing the coordinate $x_i = 0$ and such that $x_j > 0$, with $j \neq i$. $\mathcal{L}_{i,j}$ is the log-likelihood of the $(D-2)$-dimensional data obtained by removing the coordinates $x_i = x_j = 0$ and such that $x_k > 0$, with $k \notin \{i, j\}$. The sum continues with decreasing dimensions until the lowest possible dimension $D' = 2$, corresponding to edges, with $\mathbb{S}^2 = (0, 1)$. If $D = 3$, there are 4 terms to be considered in $\mathcal{L}_0$. When $D = 4$, the number of terms increases to 11. Notice that each log-likelihood in Equation (21) necessitates at least D' data to be properly computed since $\hat{\boldsymbol{\Sigma}}$ must be of full rank. If there are less than D' data, the contribution of this sub-simplex is simply ignored.

Note that, if the Gaussian hypothesis on the multivariate random field $\mathbf{Z}(s)$ generating the data is in force, the transformed data move away from Gaussianity as $|\alpha - \hat{\alpha}|$ increases, making kriging sub-optimal, in the sense that there exists a non-linear predictor improving on kriging in the L^2 sense. In this sense, $\hat{\alpha}$ is expected to be the value of α optimizing the performance of kriging.

Pattern		$\alpha_0 = 0$	$\alpha_0 = 0.2$	$\alpha_0 = 0.6$	$\alpha_0 = 1$
Center	$\tilde{\mathbf{z}}$	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, 0)$
	σ	1	0.50	0.15	0.065
Border	$\tilde{\mathbf{z}}$	$(-2.3, 1)$	$(-2.3, 1)$	$(-2.3, 1)$	$(-2.3, 1)$
	σ	1	0.50	0.15	0.065
Corner	$\tilde{\mathbf{z}}$	$(4, -3)$	$(4, -3)$	$(4, -3)$	$(4, -3)$
	σ	1	0.38	0.11	0.045

Table 1: Shifting and scaling parameters used to create patterns of data in the center, at the border and in a corner of the simplex.

4 A simulation study

The aim of this section is to analyze the Isometric α-transformation introduced in Section 3.1, by applying it to simulated spatial compositional data. Here, we will evaluate the performances of this approach in comparison to the classical log-ratio transformations introduced by Aitchison (1986) and Egozcue et al. (2003).

4.1 Simulation process

We focus here on spatial compositional data in $\mathbb{S}^D$, $D = 3$, but the same approach could be applied to any dimension $D \in \mathbb{N}$, up to some minor modifications. The spatial domain $\mathcal{D}$ is the square $[0, 10] \times [0, 10]$ filled with 2000 random uniform locations $\{s_j : j = 1, \ldots, 2000\}$. At these locations, bivariate Gaussian values $\{\mathbf{Z}(s_j)\}$ are simulated from a parsimonious bivariate Whittle-Matrn Model (Gneiting et al., 2010) with covariance $C_{ij}(h) = c_{ij} W_{\nu_{ij}}(h/s_{ij})$. Each function $W_{\nu_{ij}}$ is given by

$$W_{\nu_{ij}}(r) = \frac{2^{1-\nu_{ij}}}{\Gamma(\nu_{ij})} r^{\nu_{ij}} K_{\nu_{ij}}(r), \tag{22}$$

where $\nu_{ij} > 0$ and K_ν is the modified Bessel function of second kind. $B = 100$ different realizations from this bivariate Whittle-Matrn model are drawn using the `R` package `RandomFields` with the following parameters: $c_{ii} = 1$, $c_{ij,i\neq j} = 0.8$, $\nu_{ii} = \nu_{ij,i\neq j} = 0.5$, $s_{ii} = s_{ij,i\neq j} = 1$.

Using shifting and scaling of the bivariate data according to $\mathbf{Z}_{new}(s_j) = \sigma(\mathbf{Z}(s_j) - \tilde{\mathbf{z}}), \forall j$, where σ and $\tilde{\mathbf{z}}$ are a scaling and shifting parameter respectively, various scenarios are built from these realizations. The values σ and $\tilde{\mathbf{z}}$ are chosen such that for a given α_0 (hereafter set to $\alpha_0 = 0.2$ or $\alpha_0 = 0.6$) the simulated sets lie within the shield-shaped codomain of the α-IT. Compositional data are then computed using the inverse α-IT presented in Section 3.2 for each vector $\mathbf{Z}_{new}(s_j)$, with $j = 1, \ldots, 2000$. Thanks to the different scaling and shifting scenarios, different patterns of compositional data are created in the simplex.

Figure 2 shows the shifted and scaled data located in the center, at the border and in the corner of the shield-shaped codomain, along with the corresponding compositional datasets once the data have been back-transformed to the simplex with α_0. The specific values of α_0, σ and $\tilde{\mathbf{z}}$ that will be used throughout this work are given in Table 1.

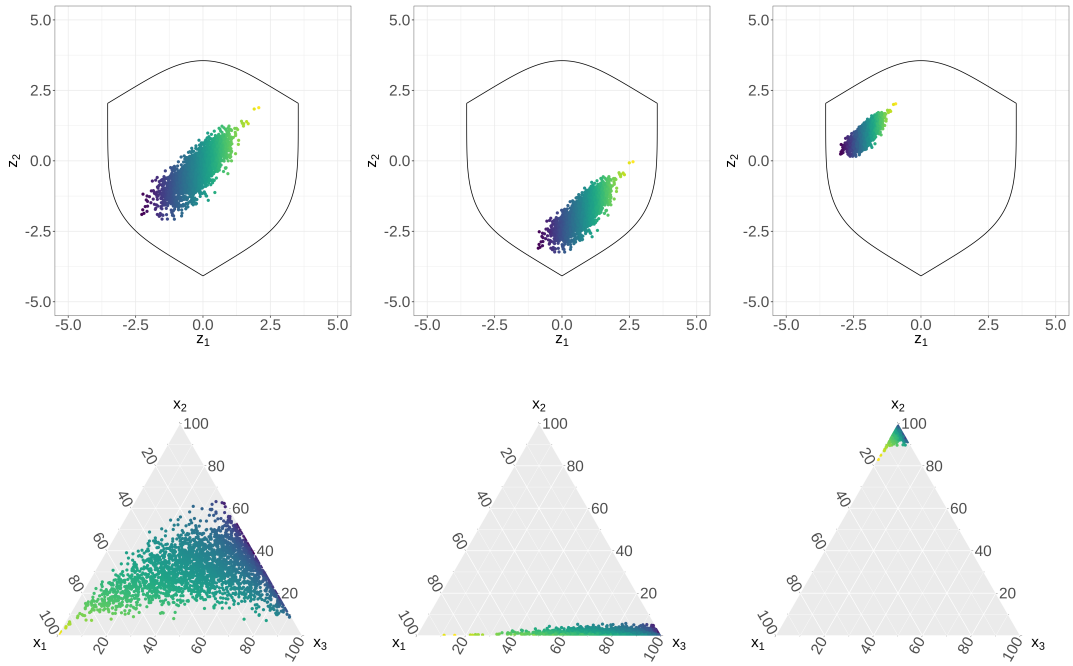

Figure 2: Top: bivariate data after shifting and scaling. The continuous line shows the limit of the codomain for $\alpha_0 = 0.2$. Bottom: corresponding compositional datasets once transformed with the inverse α_0-IT, as in Equation 12. Shades of color follow the z_1 axis.

To evaluate the kriging performances, each of the compositional datasets plotted at the bottom of Figure 2 is then divided into a training set of $n = 200$ data randomly sampled from the realization and a test set made of the remaining $N = 1800$ data. A bivariate proportional model is estimated on the training set and kriging is performed to predict at the test locations. Predicted values are then back-transformed into the simplex. On each of the $B = 100$ realizations, this procedure is applied for a sequence of values of α ranging in $[0, 1]$. Note that the locations of the training and test sets are fixed among all the B different realizations.

4.2 Distances and prediction scores

Since a composition in $\mathbb{S}^D$ can be seen as a discrete distribution over $\{1, \ldots, D\}$, we will use distances between distributions to assess the prediction performances. Note that, even though we here consider scores based on distances between distributions, other measures of discrepancy could be used for the same purpose, such as those based on the concept of divergence, e.g., Kullback-Leibler divergence (Kullback and Leibler, 1951) or the α-divergence (Medak and Cressie, 1991). Specifically, we will use the Hellinger distance and the Total Variation distance. These distances allow us to look at the prediction performances of compositional data in an objective way that does not depend upon a specific transformation of the data. These distances are defined between probability measures but for our purposes we here introduce their definition when discrete distributions are concerned. Consider two discrete probability distributions $P = (p_1, \ldots, p_D)$ and $Q = (q_1, \ldots, q_D)$.

The Hellinger distance, d_H, is defined as

$$d_H(P,Q) = \frac{1}{\sqrt{2}}\sqrt{\sum_{i=1}^{D}(\sqrt{p_i}-\sqrt{q_i})^2}, \tag{23}$$

and the Total Variation distance, d_{TV}, is

$$d_{TV}(P,Q) = \frac{1}{2}\sum_{i=1}^{D}|p_i - q_i|. \tag{24}$$

Classical inequalities between L_1 and L_2 norms provide

$$d_H^2(P,Q) \leq d_{TV}(P,Q) \leq \sqrt{2}d_H(P,Q).$$

For each simulation b, with $b = 1,\ldots,B$, the prediction scores will be the average of the Hellinger (respectively Total Variation) distances between the kriging prediction $\mathbf{x}_j^*$ at location $s_j \in \mathcal{D}$ and the corresponding true value, $\mathbf{x}_j$, with $j = 1,\ldots,N$. These average distances, denoted δ_H^b and δ_{TV}^b, are thus

$$\delta_H^b = \frac{1}{N}\sum_{j=1}^{N} d_H(\mathbf{x}_{j,b}^*, \mathbf{x}_{j,b}), \qquad \delta_{TV}^b = \frac{1}{N}\sum_{j=1}^{N} d_{TV}(\mathbf{x}_{j,b}^*, \mathbf{x}_{j,b}). \tag{25}$$

Finally, we also use the α-IT-distance defined in Equation (14) and we compute

$$\delta_\alpha^b = \left[\frac{1}{N}\sum_{j=1}^{N} d_{\alpha-IT}(\mathbf{x}_{j,b}^*, \mathbf{x}_{j,b})^2\right]^{1/2}. \tag{26}$$

We recall that the parameter α used to compute this distance is not necessarily equal to the parameter α_0.

4.3 Results

First, the Maximum Likelihood Estimator (MLE) is computed following the method detailed in Section 3.6. Table 2 reports the averages and standard deviations of the MLE estimates computed from $B = 100$ realizations with $n_S = 500$ sample points, in several situations and for 4 different values of α_0, including $\alpha_0 = 0$ which corresponds to an ILR model and $\alpha_0 = 1$ which corresponds to conditional Gaussian compositions. Overall, the MLE shows good performances. The parameter α tends to be slightly underestimated, at the exception of $\alpha_0 = 0$. In this case, the true value is at the border of the possible values for α, so that estimation errors are one-sided. The standard deviation increases with α_0. For $\alpha_0 = 0.6$ and border data, the average standard deviation have also been computed with increasing numbers of sample points, i.e. $n_S = 500, 1000, 2000$. The averages on $B = 100$ realizations of independent random vectors were $(0.591, 0.592, 0.615, 0.601)$. The standard deviation of $\hat{\alpha}$ was equal to $(0.208, 0.128, 0.082, 0.062)$ indicating that it decreases roughly at the $n^{-1/2}$ rate. Similar results have been obtained on further simulations (not reported here). In presence of spatial dependence, the rate of convergence was found to be slower under increasing domain asymptotics. Convergence under infill asymptotics has not been explored, this

Pattern	$\alpha_0 = 0$	$\alpha_0 = 0.2$	$\alpha_0 = 0.6$	$\alpha_0 = 1$
Center	0.025 (0.03)	0.181 (0.09)	0.563 (0.24)	0.957 (0.37)
Border	0.051 (0.07)	0.190 (0.08)	0.574 (0.26)	0.940 (0.41)
Corner	0.053 (0.07)	0.178 (0.05)	0.587 (0.18)	0.997 (0.36)

Table 2: Average (standard deviation) of $\hat{\alpha}$ for different configurations of data (as described in Table 1). $B = 100$ realizations with $n_S = 500$ samples.

point being left to further research. An important issue was to check whether MLE is able to estimate correctly α when the data correspond to an ILR transformation ($\alpha = 0$) or to a linear transformation ($\alpha = 1$). Our findings provide evidence that this is indeed the case.

Starting from the compositional datasets, we then explore the kriging performance for different values of α, pretending that the true parameter α_0 is not known. Specifically, for a given value $\alpha \in [0, 1]$, the α-IT is applied to the compositional training set, thereby defining $\mathbf{z}_\alpha = (\mathbf{z}_{1,\alpha}, \mathbf{z}_{2,\alpha})$ in $\mathbb{R}^2$. Notice that when $\alpha \neq \alpha_0$, $\mathbf{z}_\alpha$ will be different from the original simulated values, which are only retrieved when $\alpha = \alpha_0$. A Linear Model of Coregionalization (LMC) is then fitted on $\mathbf{z}_\alpha$ using the `R` package `RGeostats`. Cokriging is then applied at the N locations of the test set. Predicted values are then back-transformed to $\mathbb{S}^3$ with the parameter α. Finally, the prediction scores, presented in Section 4.2, are computed. This process is repeated for several values of α ranging from 0 to 1.

We first assess the performance in the (transformed) Euclidean space by computing the average kriging RMSE, which is nothing but the average over the B simulations of the metric δ_α^b defined in (26). Results are plotted in Figure 3 for different values of α. Each point represents the average metric $\overline{\delta}_\alpha = \sum_{n=1}^{B} \delta_\alpha^b / B$, normalized by σ, which is the empirical standard deviation computed on the transformed vectors $\mathbf{z}_\alpha$. In the left panel, the data are originated from the inverse α-IT with $\alpha_0 = 0.2$, while the right panel represents the data generated with $\alpha_0 = 0.6$. We note that the minimum of the kriging RMSE is attained around $\alpha = \alpha_0$ in the two cases. As α gets away from α_0, the RMSE becomes larger.

In Figure 4 we plot the average of the kriging prediction scores (Total Variation and Hellinger metric defined in Equation (25)) computed on the data in the simplex after inverse α-IT over the $B = 100$ simulations for the center data with $\alpha_0 = 0.2$ and $\alpha_0 = 0.6$. Even though the difference is less marked, this plot does indicate that a value of $\alpha = \alpha_0$ achieves the best scores. The parameter α resulting in minimal kriging error is thus close to the value α_0 used to simulate the data and also close to the ML estimates reported in Table 2. These findings provide good support for the use of ML estimation for estimating the parameter α when analyzing data.

Prediction scores based on Total Variation and Hellinger metrics are less variable with respect to α than the kriging errors computed in the Euclidean space $\mathbb{R}^2$. This is due to the fact that the inverse transform shrinks the data (both observed and kriged) into a smaller area of the simplex causing the absolute error to decrease.

Figure S1 in the Supplementary Material shows the same result as Figure 4 for the extreme cases $\alpha_0 = 0$ (data simulated from inverse ILR) and $\alpha_0 = 1$ (data simulated from linear transformation). As expected, the plot shows that the Total Variation metric and the Hellinger metrics recover $\alpha = 0$

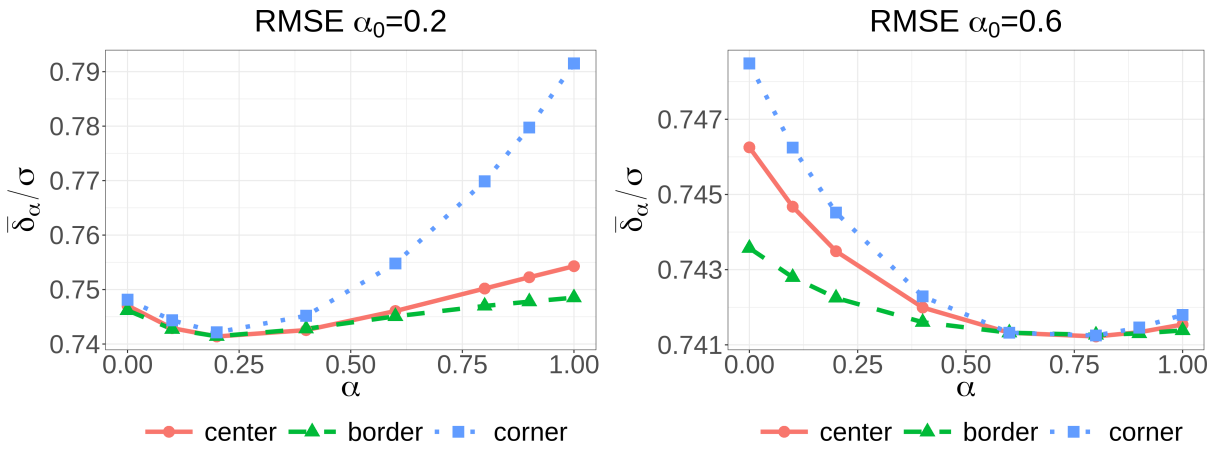

Figure 3: Average kriging RMSE $\bar{\delta}_\alpha/\sigma$ over $B = 100$ tests for center, border and corner data in $\mathbb{R}^2$ before inverse α-IT for $\alpha_0 = 0.2$ (left) and $\alpha_0 = 0.6$ (right).

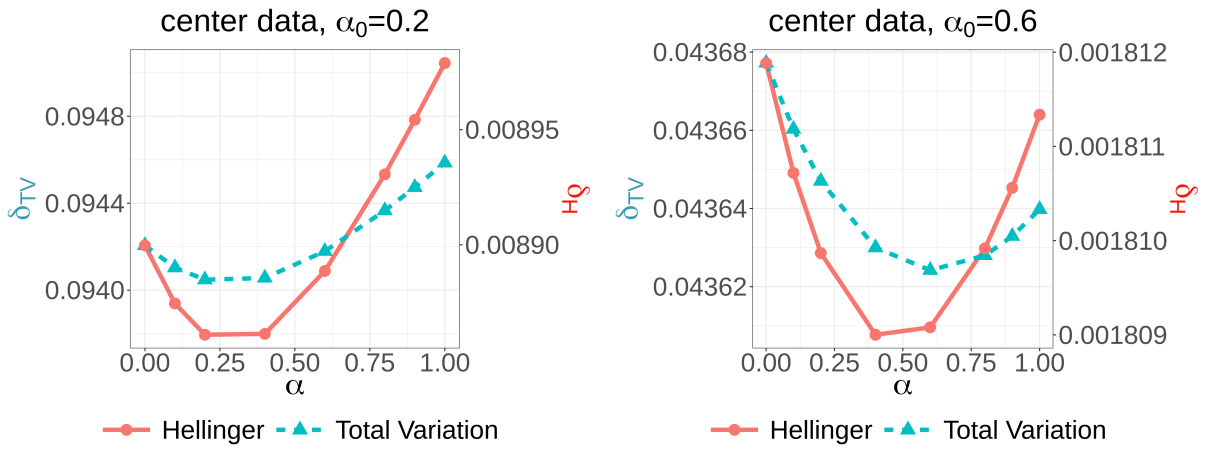

Figure 4: Average error over $B = 100$ tests for center data after α-IT with different α and $\alpha_0 = 0.2$ (left) and $\alpha_0 = 0.6$ (right): Total Variation metric in blue dashed lines and Hellinger metric in red continuous lines.

(respectively $\alpha = 1$) not far from the best value of α for kriging.

Figure 5 shows the normalized averages of the kriging errors of the center data with $\alpha_0 = 0.2$ and $\alpha_0 = 0.6$ for different Isometric α-metrics. The result shown in the plot is less straightforward to interpret, since the minimal kriging error computed with the normalized α-IT metric is attained close to the value α used in the metric instead of being close to α_0 as one would expect. This implies that the choice of the metric influences the result of the analysis in a sort of self-confirming way and that it is thus safer to use metrics that are not related to the transformation of the compositional data, as it is the case for the Total Variation or the Hellinger metrics. In particular, this result indicates that one should not use the Aitchison metric, corresponding to $\alpha = 0$, in all circumstances, irrespective of data at hand.

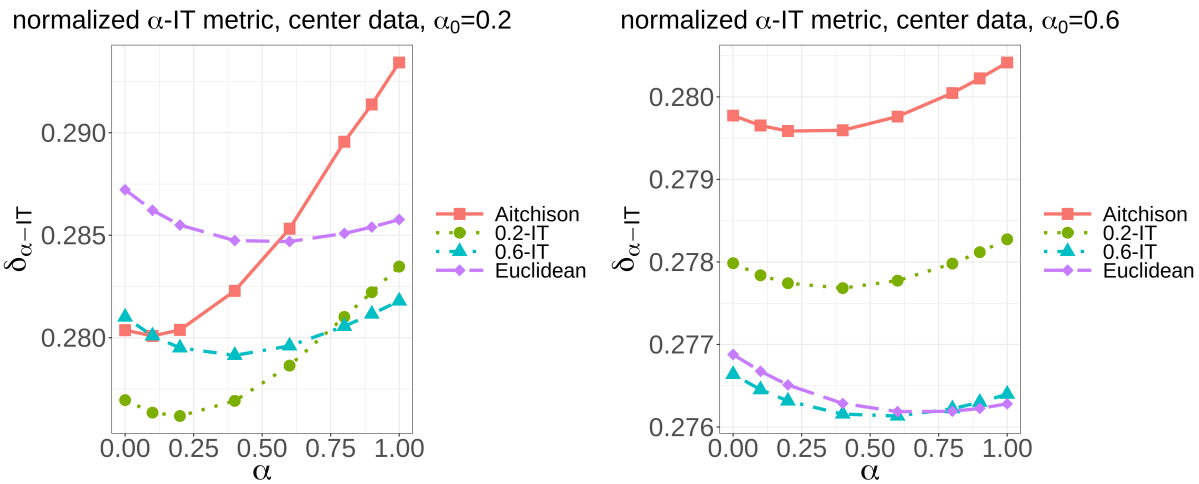

Figure 5: Average of normalized α-IT error over $B = 100$ tests for center data after α-IT with different α and $\alpha_0 = 0.2$ (left) and $\alpha_0 = 0.6$ (right).

Finally, Figure S2 in the Supplementary Material shows the same scores of Figure 4, but for border and corner data. We remark that the evolutions of the errors are very similar to the center case. We note that in none of the plots (center, border or corner) the value $\alpha = 0$, which corresponds to the classical ILR transform, leads to the lowest error in kriging.

5 Application to the Copernicus Land Cover Map

A geostatistical analysis of a spatial compositional dataset is now conducted following the lines presented above: cokriging of the transformed data by α-IT followed by the back-transformation into the simplex. Using Monte Carlo cross-validation, we shall assess the performance of our approach for several values of α, including the value α^* maximizing the likelihood (21), even though the transformed variables are not necessarily multivariate Gaussian.

5.1 The dataset

The Copernicus Dynamic Land Cover Map (CGLS-LC100) (Buchhorn et al., 2019) delivers a global land cover map at 100 m spatial resolution, offering a primary land cover scheme. The different proportions of covers sum to 1 for every pixel, since each of them represents the proportion of the pixel covered by that given land cover. These proportions can thus be considered as compositional data since they are positive and sum to 1. Scientists in spatial analysis of compositional data have been using land cover datasets to apply prediction techniques; recent works can be found in Lungarska and Chakir (2018), Nguyen et al. (2021) and Thomas-Agnan et al. (2021).

For our study, we consider only a part of the Copernicus dataset that corresponds to the Po' Valley in Italy, a major geographical feature of Northern Italy. It extends approximately 650 km in an E-W direction, with an area of $46,000\text{km}^2$. As standard practice in geostatistics, latitude/longitude coordinates are transformed into the corresponding UTM coordinate system. In a set of 2000 pixels, randomly picked in the spatial domain, we select the 4 main parts of the compositional samples, i.e., those which explain the highest proportion of the composition, and create a new compositional dataset with 4 parts, *crops, shrub, grass* and *tree*, belonging to the simplex $\mathbb{S}^4_0$. Since some land covers are not found in each pixel, some of the parts are equal to 0, which makes this dataset an interesting case study for assessing how performances vary with the amount of data containing parts equal to 0.

The predictive performances will be measured using a Monte Carlo cross-validation approach, in which the following procedure is repeated $B = 20$ times on the set of 2000 pixels: for each $b = 1, \ldots, 20$, a random sample of size 500 is selected to be the training set and the remaining 1500 compositions form the validation set; an LMC is fitted to the α-IT transformed data and cokriging is performed for prediction at the validation locations. Scores are computed using the Total Variation and the Hellinger metrics introduced in Equation (25).

5.2 Spatial analysis with positive parts

We first consider data belonging to $\mathbb{S}^4$, i.e., data with positive values for all parts. With this setting, a fair comparison is possible with the classical ILR transform. In Figure 6 (right), the set of 2000 data is represented in stacked barplots, where each bar represents a compositional sample. The same data are mapped in Figure 7, where each of the four components varies between 0 and 1. Visual inspection confirms that the area is quite homogeneous, and that stationarity can reasonably be assumed for the purpose of this study.

The log-likelihood (21) is maximized with respect to α for each of the B training samples, thereby providing B estimates of $\alpha^\star$. The average of these estimates is equal to 0.120 and their standard deviation is 0.048, showing a small variability around the average value. It is noticeable that $\alpha = 0$ lies outside the 2-standard deviations confidence interval, a fact that provides strong evidence that setting $\alpha = 0$ for an ILR transform is not supported by the data. For the rest of this Section, we shall thus consider that $\alpha^* = 0.12$ corresponds to the ML estimates of the α-IT transform.

Figure 8 displays the scatterplots of the 2000 transformed data after the α-IT with $\alpha = 0$, $\alpha^\star = 0.12$ and $\alpha = 1$. When $\alpha = 1$ the triangular shapes reflect that a linear transformation has

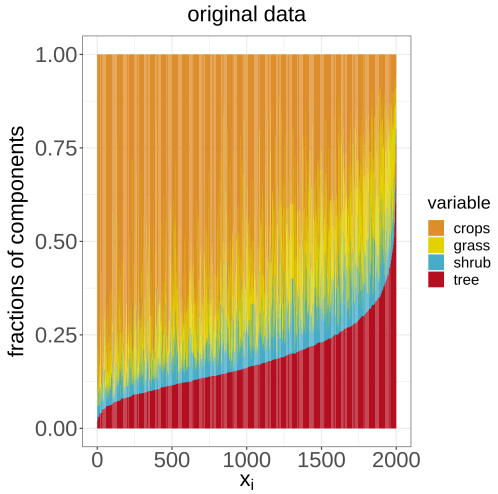

Figure 6: Stacked barplots for the compositional set of size 2000 excluding compositions with 0s.

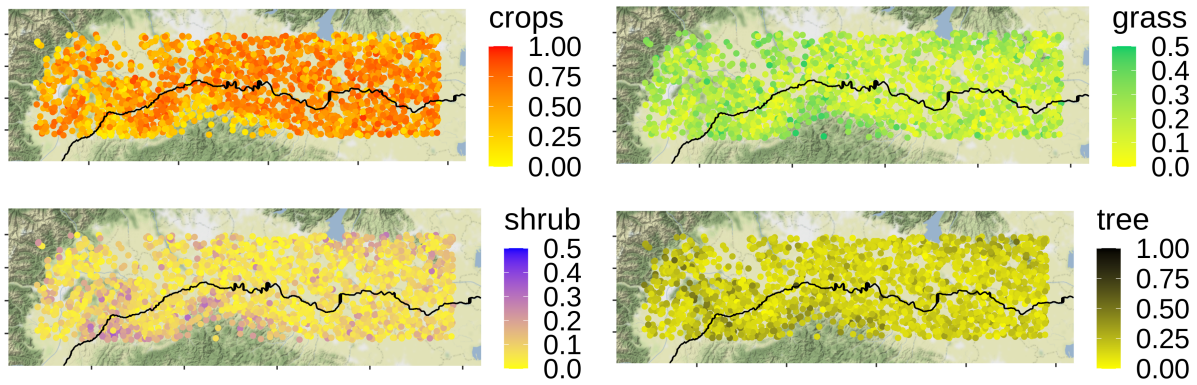

Figure 7: Compositional set of size 2000 excluding compositions with 0s.

been applied on the data originated from the simplex $\mathbb{S}^4$. When $\alpha = 0$, one can notice that groups of data are slightly separated from the main cluster of data. We shall return to this in Section 5.3.

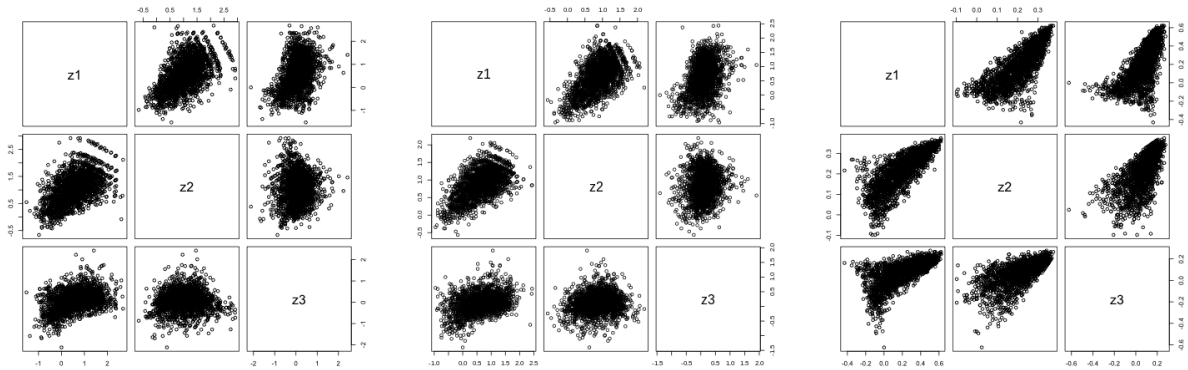

Figure 8: Scatterplot of z_1, z_2 and z_3 after Isometric α-transformation with $\alpha = 0$ (left), $\alpha^\star = 0.12$ (center) and $\alpha = 1$ (right).

Once the data have been transformed to the Euclidean space $\mathbb{R}^3$ for each value of $\alpha \in A = \{0, \alpha^\star, 0.3, 0.5, 0.75, 1\}$, the empirical (co)variograms are computed and an LMC is fitted to the transformed data. Figure 9 shows the empirical and fitted (co-)variograms for $\alpha^\star = 0.12$. One can observe the presence of a significant nugget effect – larger than the partial sill in every direct or cross-variogram – entailing a possible effect on the quality of kriging predictions. Similar fit were obtained for all values of α. The shapes were identical, but the sill decreased as α increases, owing to the associated reduction of the codomain in $\mathbb{R}^3$ (see Figure 1).

The average of the kriging errors at the validation locations across the B simulations for different values of α computed for the two metrics (Hellinger metric and Total Variation metric) presented in Equation (25) are reported in Figure 10. According to the Total Variation metric, $\alpha^\star = 0.12$ leads to the best result compared to all other values of α, including $\alpha = 0$ and $\alpha = 1$. Regarding the Hellinger metric, the minimum of the kriging error among the 6 values of α is obtained for $\alpha = 0.3$, which is not exactly the value that corresponds to the maximal proximity to Gaussianity (i.e., $\alpha^\star = 0.12$), but it is close to this value. Notice that the true minimum could be attained for some other value within the range $(0.12, 0.50)$, but this has not been evaluated here. Figure 11 shows the back-transformed proportions of the kriging values over a grid made of 10000 points after the α-IT with $\alpha^\star = 0.12$ given 500 observed points. Notice that at each location, these predicted proportions are non-negative and that their sum is equal to 1.

To summarize, this analysis has shown that the values $\alpha = 0$ and $\alpha = 1$, corresponding to the Aitchison and the linear transformation respectively, are not optimal for both the Hellinger and the Total Variation metrics. An intermediate geometry, corresponding to the α-IT with $\alpha^\star = 0.12$, seems to better describe the dataset in a geostatistical setting.

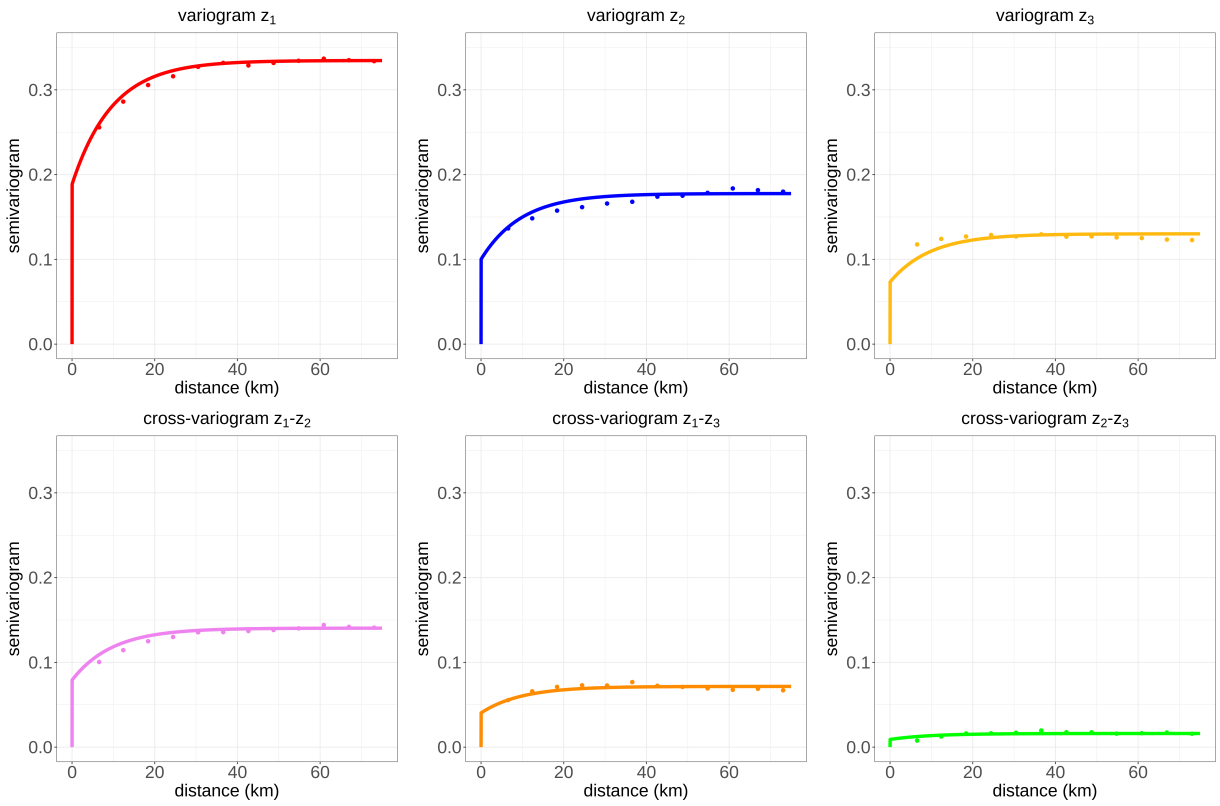

Figure 9: Variograms and cross-variograms for z_1, z_2, z_3 of the dataset without 0s after α-IT with $\alpha^\star = 0.12$.

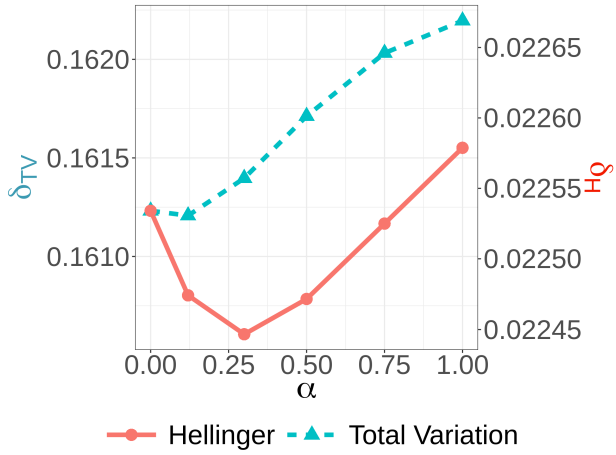

Figure 10: Average of the kriging errors: Total Variation metric in blue dashed lines and Hellinger metric in red continuous lines.

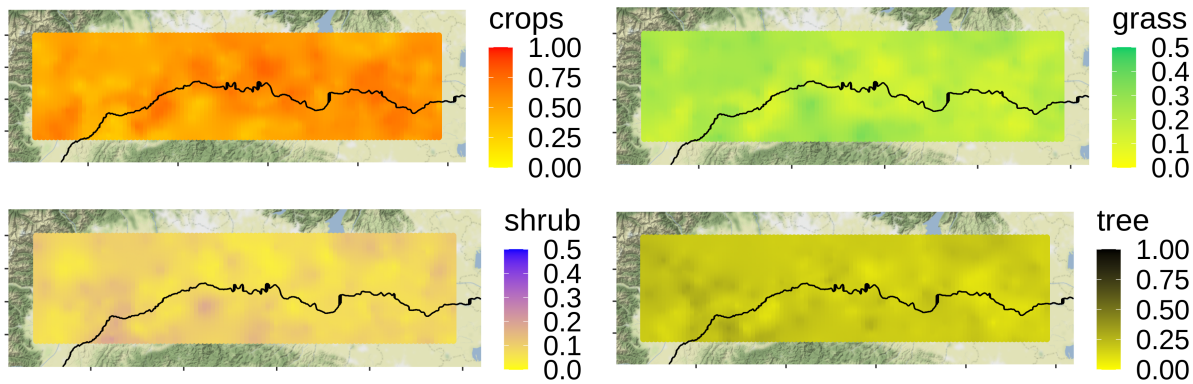

Figure 11: Kriging proportions over a grid of 10000 points with $\alpha^\star = 0.12$. Proportions are positive and sum to 1.

5.3 Spatial analysis with 0s

In the complete Copernicus dataset the land cover proportion is actually equal to 0 for at least one of the four parts in about half of the pixels (53% of compositions without 0s, 31% with 1 zero, 12% with 2 zeros, 4% with 3 zeros). A compositional geostatistical analysis thus requires to address the issue raised by null values in the vector of proportions. Dealing with 0s is a challenge for the log-ratio approach. In Tolosana-Delgado et al. (2009), the authors propose to use the undersampled cokriging, where, even if some coordinates of the observation vectors may be missing, it is possible to look for both an interpolation of full vectors at unsampled locations and the completion of the missing variables at the sampled locations. Martn-Fernndez et al. (2003); Palarea-Albaladejo and Martn-Fernndez (2007); Martn-Fernndez et al. (2011); Templ et al. (2016) propose various methods to impute 0s with small values before the statistical analysis of the data; as an alternative, these authors suggest to remove the samples with at least one 0.

As discussed in Section 3.1, the α-IT (with $\alpha > 0$) can be applied to compositions that include one or more components equal to 0, thus overcoming the theoretical limitations of the log-ratio approach. In order to test our approach on such datasets, the Monte Carlo cross-validation assessment presented in Section 5.2 is here repeated on the data containing 0s. Two settings are considered. In the first setting, a moderate proportion of data (set to 10%) presents at least one part being equal to 0. Among those, the proportion of data with one and two parts equal to 0 reflects the proportion seen in the data. In the second setting, the sample of size 2000 is drawn from the general dataset regardless of the number of null compositions. Hence, on average, about half of the data contain at least one part equal to 0. In both settings, data with a single positive part (thus equal to 1) are excluded.

In the first setting, the average of the 20 ML estimates for α is equal to 0.120 and the standard deviation is 0.036. In the second setting, the average of ML estimates is equal to 0.089 and the standard deviation is 0.040. Notice that $\alpha = 0$ lies outside the confidence interval in all cases. The α-IT is performed with $\alpha \in A' = \{0.01, \alpha^\star, 0.3, 0.5, 0.75, 1\}$, where 0.01 was chosen as an approximation of 0 (0 being not acceptable). Figure 12 represents the Total Variation and the

Hellinger scores of the kriged compositions as a function of α. The errors are maximal when α is close to 0. There is a local minimum at 0.12 and 0.30 in the first and second setting respectively. Then, the scores reach a floor when $\alpha > 0.5$ and they remain almost equal up to $\alpha = 1$. The same behavior can be observed on the normalized α-IT prediction error (in the Euclidean space) instead of Total Variation or Hellinger distances (in the simplex). For example, in the second setting, $\bar{\delta_\alpha}/\sigma$ is equal to (1.64, 1.58, 1.49, 1.56, 1.42 and 1.41) as α goes from 0.01 to 1 in A'.

The results above thus show that even though the ML estimates are close to $\alpha = 0.1$, the prediction scores are minimized close to $\alpha = 1$, suggesting that absence of transformation of the data is a good option when it comes to kriging in presence of compositions with null parts.

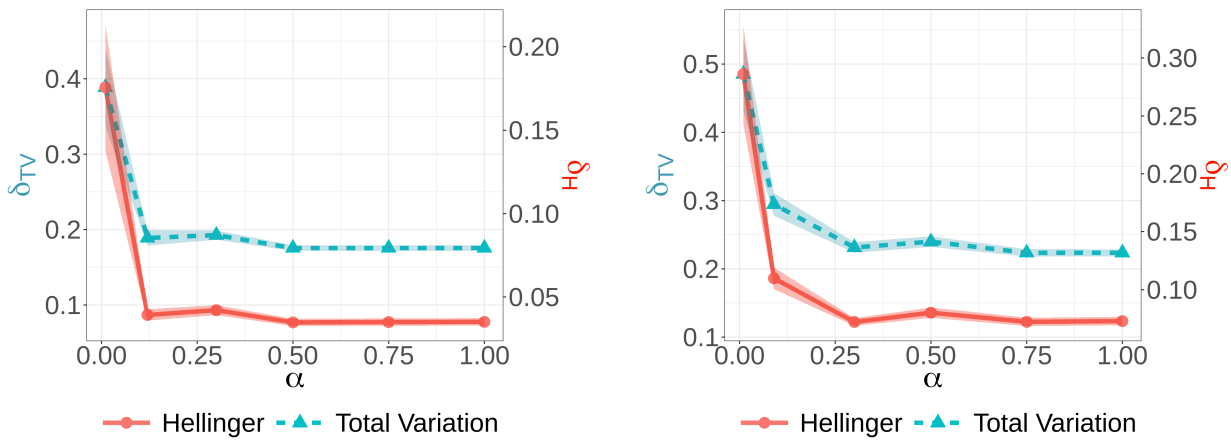

Figure 12: Hellinger and Total Variation metrics computed on prediction errors for 20 random sample datasets with 0s. Thick line: average; light shadow : $\pm 2\sigma$ envelopes. Left: moderate (10%) amount of 0s. Right: large amount of 0s.

In order to understand this counter-intuitive behavior, Figure 13 offer a 3D representation of the set of data with a large proportions of compositions with null parts, along with the kriged vectors in the Simplex (left) and in the 3D Euclidean space (right). Two values of α are considered: in the top row, α is the ML estimate $\alpha^\star$, and in the bottom row, $\alpha = 1$ is used. It is clearly visible that the α-IT produces clusters of data in the Euclidean space. These clusters are associated to different groups of data. The most numerous group lies in the center, and corresponds to data with positive values for all parts. Then, there are four clusters corresponding to the four possible groups of data with exactly one part equal to 0, i.e. data belonging to one of the facets of the tetrahedron. Finally, there are six smaller groups corresponding to data with two parts equal to 0. These data lie on the edges of the tetrahedron. When $\alpha = 1$, the transformation is linear, and these clusters are as close to each other as in the original dataset. As α gets smaller, the distance between these clusters increases. When α approaches 0, the distance between clusters tends to infinity, in relation with the fact that the logarithm of 0 is undefined. The likelihood in Equation (21) is the sum of partial likelihoods computed within each group, but it does not account for the distance between groups. The maximum of the likelihood is thus reached when the groups are the closest to Gaussianity, regardless of the clustering effect.

When kriging is computed at a target location, data in the neighborhood can belong to different

clusters (some data with only positive parts, some others with one or two 0s, possibly not for the same parts). In this case, kriging, which is the weighted average of values belonging to different clusters, can be located in an area of the Euclidean space where there are no data. This is the case for the light blue points represented in Figure 13. The kriging RMSE in the α-IT geometry will then increase as α decreases. Once back-transformed, these points will end up in regions in the simplex where there are no data, implying high values of the Total Variation and Hellinger metrics. Figure S3 represents the same plots for a moderate proportion of compositions with null parts.

In conclusion, the clustering effect due to compositions with 0s is amplified as α decreases, entailing poor kriging performances both within the α-IT geometry and for metrics in the simplex. Note that a similar problem is addressed in Compositional Data Analysis when analyzing the so-called *essential zeros*, i.e., 0s which are not due to rounding or low sensitivity of the measurement instrument. In these cases, zero-replacement is clearly not appropriate. A strategy of analysis proposed by some authors (e.g., Aitchison and Kay (2003)) consists of two separate steps, namely (i) analyzing the patterns of zeros (i.e., the data clusters) and (ii) analyzing, within clusters, the non-zero parts according to the Aitchison geometry. A similar strategy can be envisioned in the setting of α-IT, including an explicit modeling of the clusters induced by the presence of 0s. A relevant advantage here would be that the geometry induced by the α-IT is well-defined over all the compositions, thus allowing modeling across clusters, and not only within clusters.

6 Discussion and conclusion

A new class of α-transformations, named α-IT, has been proposed, to allow for the geostatistical analysis of compositional data. The transformation has been proved to converge to the Isometric Log-Ratio transformation as α approaches 0, while it reduces to a linear transformation when $\alpha = 1$. In this sense, the α-IT represents a compromise between the Aitchison geometry ($\alpha = 0$) and the Euclidean one ($\alpha = 1$). Nonetheless, the presence of the parameter α controlling the degree of transformation applied to the data offers the relevant advantage of letting the problem suggest the most appropriate transformation to use on the data. As far as kriging is concerned, we proposed a maximum likelihood estimator for α, which maximizes the Gaussianity of the transformed data – hence the kriging performances. On simulations of compositions originating from Gaussian random fields with inverse α-IT, our results show that the ML estimator provides close to unbiased estimates for α, i.e., it allows to correctly identify the α-IT yielding to data Gaussianity and optimizing kriging performances.

It is worth mentioning that, unlike alternative classes of α-transformations nowadays available (Tsagris et al., 2011), the α-IT also allows for an explicit characterization of the covariance structure of the field according to the geometry induced by the transformation. In particular, this enables one to formulate a generalization of the so-called ILR covariance matrix (see Section 2.2) – widely-used in compositional data analysis – while drawing a direct connection with the latter as α approaches 0.

Simulation results and data analyses on Copernicus land cover data confirm that the Aitchison geometry – retrieved for $\alpha = 0$ – may not be optimal for kriging, as it proved to be outperformed

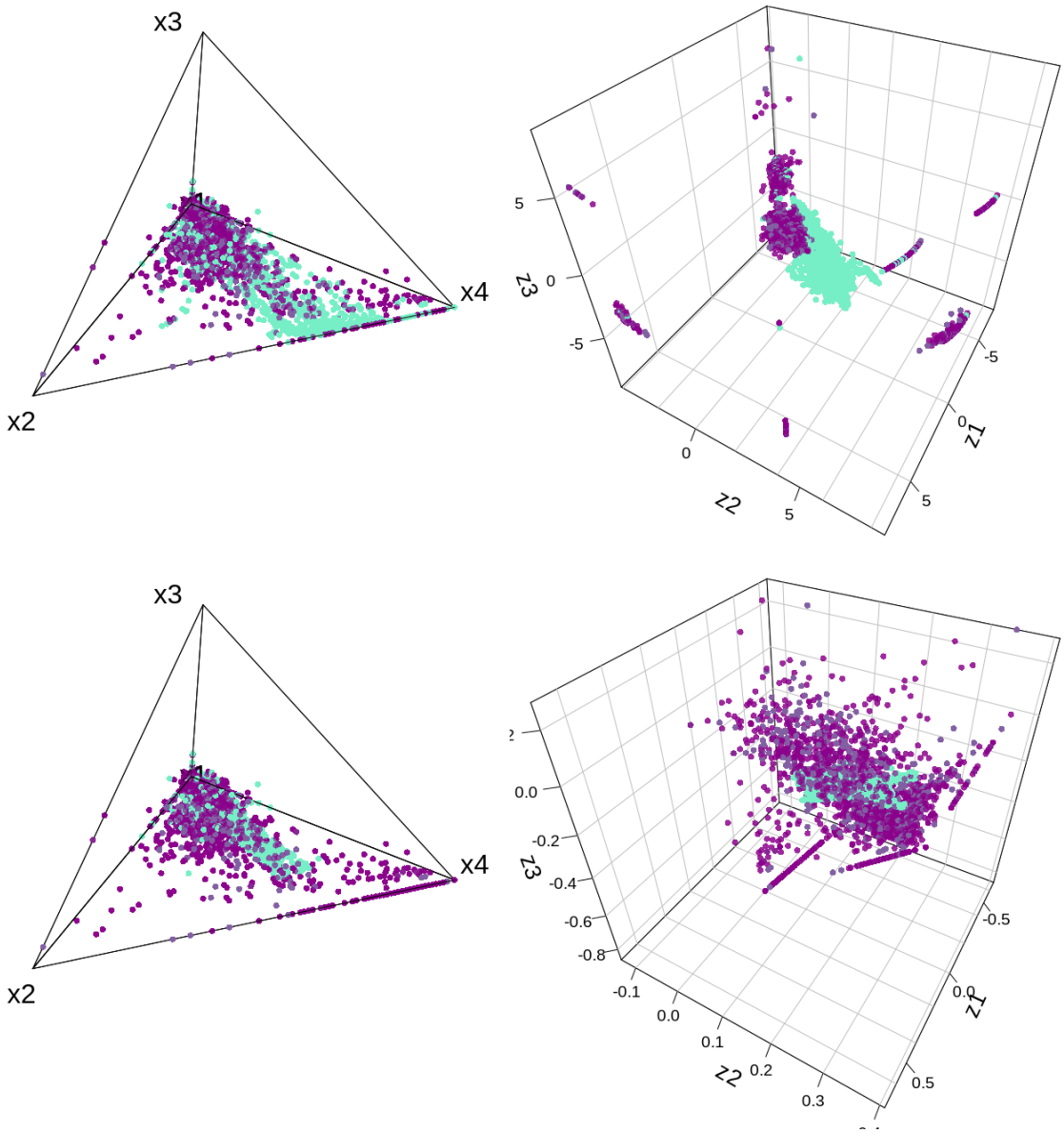

Figure 13: Original data (violet) and kriging predictions (light blue) with a high (about 50%) fraction of data with 0s (second setting). Left: simplex. Right: Euclidean space. Top: $\alpha^\star = 0.089$. Bottom: $\alpha = 1$.

by the α-IT when α was set through ML. On the other hand, the α-IT allows one to explicitly deal with compositions with 0s, for which the log-ratio approach is not well-defined. Our investigation on this point provides evidence that the use of a small value for α may not be appropriate in the presence of null parts, because this amplifies the grouping structure merely induced by the 0 themselves, with detrimental effect on kriging performances. Our results indeed suggest that the use of a geometry close to the Euclidean one ($\alpha \simeq 1$) may be more appropriate in these cases instead. These findings complement, from a different perspective, the theoretical results established in Allard and Marchant (2018), which state that linear combinations of the data (and thus kriging with $\alpha = 1$) are the only unbiased central tendency characteristics satisfying a small set of axioms, namely continuity, reflexivity, and marginal stability.

In this work, we computed the cokriging of the transformed vector $\mathbf{Z}(s_0)$ at unsampled locations s_0, conditional on the vectors $\mathbf{Z}(s_j) = A_{\alpha-IT}(\mathbf{X}(s_j))$, $j = 1, \ldots, n$. Ultimately, the goal is to predict the compositional vector $\mathbf{X}(s_0)$ given the same information. However, by back-transforming the predictions from $\mathbb{R}^{D-1}$ to $\mathbb{S}^D$ with a non-linear transformation $A^{-1}_{\alpha-IT}$, one actually introduces a bias even though the prediction of $\mathbf{Z}(s_0)$ is unbiased. This issue is closely related to the above mentioned results in Allard and Marchant (2018). However, it should be noted that the definition of the bias is tightly related to the geometry under consideration. Implicitly, the geometry in use when evaluating the bias on the prediction of $\mathbf{X}(s_0)$ is the Euclidean geometry in the simplex, whilst the geometry implied by the α-IT distance defined in Section 3.3 is equal to the Euclidean distance in the transformed space. Defining and correcting the bias is an interesting yet very broad issue worth to be further investigated. The problem of dealing with missing data, when some of the compositional vectors are incomplete, has not been tackled here. It is also left for future work.

More in general, we remark that, while this work focuses on the analysis of spatial compositional data with a focus on spatial prediction (kriging), the α-IT transformation can be used for any type of compositional analysis, ranging from exploratory analysis, to classification, inference and regression. In these settings, one may use the α-IT to convert the original compositional data into Euclidean data, analyze them according to the Euclidean geometry, and then transform back the results to the simplex, with the same process as that demonstrated in this work. While we generally envision a clear path of research to this scope, further research is needed to export the idea here presented to further settings, in the direction of, e.g., establishing criteria to set the optimal value for α. Nonetheless, our α-IT can be a promising alternative to other transformations for the very same reasons that led us to its development in the spatial setting.

Acknowledgments

The authors are grateful to four anonymous reviewers for their very careful reading and the many valuable comments that helped to improve the manuscript. `R` scripts reproducing our analyses can be found at `http://github.com/luciclar/alphaIT_spatial_compositional`.

References

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Chapman & Hall Ltd.

Aitchison, J. and Kay, J. (2003). Possible solution of some essential zero problems in compositional data analysis. In Thi-Henestrosa, S. and Martn-Fernndez, J. A., editors, *Proceedings of CoDaWork'03, The 1st Compositional Data Analysis Workshop*. University of Girona: Girona.

Allard, D. and Marchant, T. (2018). Means and covariance functions for geostatistical compositional data: an axiomatic approach. *Mathematical Geosciences*, 50(3):299–315.

Barcel, C., Pawlowsky, V., and Grunsky, E. (1996). Some aspects of transformations of compositional data and the identification of outliers. *Mathematical Geology*, 28:501518.

Billheimer, D., Guttorp, P., and Fagan, W. F. (2001). Statistical interpretation of species composition. *Journal of the American Statistical Association*, 96(456):1205–1214.

Buccianti, A., Mateu-Figueras, G., and Pawlowsky-Glahn, V. (2006). *Compositional Data Analysis in the Geosciences: From Theory to Practice*. Geological Society London.

Buchhorn, M., Smets, B., Bertels, L., Lesiv, M., Tsendbazar, N.-E., Herold, M., and Fritz, S. (2019). Copernicus global land service: Land cover 100m; epoch 2015; globe. dataset of the global component of the copernicus land monitoring service. Technical report.

Cressie, N. (1993). *Statistics for Spatial Data*. John Wiley & Sons, Inc.

Egozcue, J. J., Pawlowsky-Glahn, V., and Mateu-Figueras, G. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, 35:279300.

Gneiting, T., Kleiber, W., and Schlater, M. (2010). Matrn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 105:1167–1177.

Greenacre, M. (2009a). Log-ratio analysis is a limiting case of correspondence analysis. *Mathematical Geosciences*, 42:129–134.

Greenacre, M. (2009b). Power transformations in correspondence analysis. *Computational Statistics & Data Analysis*, 53(8):3107–3116.

Greenacre, M. (2011). Measuring subcompositional incoherence. *Mathematical Geosciences*, 43:681693.

Kullback, S. and Leibler, R. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86.

Lungarska, A. and Chakir, R. (2018). Climate-induced land use change in france: Impacts of agricultural adaptation and climate change mitigation. *Ecological Economics*, 147C.

Martn-Fernndez, J., Barcel-Vidal, C., and Pawlowsky-Glahn, V. (2003). Dealing with zeros and missing values in compositional data sets using nonparametric imputation. *Mathematical Geology*, 350:253278.

Martn-Fernndez, J., Palarea-Albaladejo, J., and Olea, R. A. (2011). *Compositional Data Analysis: Theory and Applications, chapter Dealing with zeros.* Wiley, Chichester.

Mateu-Figueras, G., Pawlowsky-Glahn, V., and Egozcue, J. J. (2011). The principle of working on coordinates. *Pawlowsky-Glahn and Buccianti (2011)*, pages 31–42.

Medak, F. and Cressie, N. (1991). Confidence regions in ternary diagrams based on the power divergence statistics. *Mathematical Geology*, 23:1045–1057.

Nguyen, T. H. A., Thomas-Agnan, C., Laurent, T., and Ruiz-Gazen, A. (2021). A simultaneous spatial autoregressive model for compositional data. *Spatial Economic Analysis*, 16(2):161–175.

Palarea-Albaladejo, J. and Martn-Fernndez, J. (2007). A parametric approach for dealing with compositional rounded zeros. *Computational. Geoscience*, 39:625645.

Pawlowsky-Glahn, V. and Egozcue, J. J. (2001). Geometric approach to statistical analysis on the simplex. *Stochastic Environmental Research and Risk Assessment*, 15:384–398.

Pawlowsky-Glahn, V. and Egozcue, J. J. (2016). Spatial analysis of compositional data: A historical review. *Journal of Geochemical Exploration*, 164:28–32.

Pawlowsky-Glahn, V., Egozcue, J. J., and Tolosana-Delgado, R. (2015). *Modeling and Analysis of Compositional Data.* John Wiley & Sons, Ltd.

Pawlowsky-Glahn, V. and Olea, R. A. (2004). *Geostatistical Analysis of Compositional Data.* Oxford University Press, Inc.

Scealy, J. L. and Welsh, A. H. (2014). Colours and cocktails: Compositional data analysis. *Australian & New Zealand Journal of Statistics*, 56(2):145–169.

Templ, M., Hron, K., Filzmoser, P., and Gardlo, A. (2016). Isometric logratio transformations for compositional data analysis. *Chemometrics and Intelligent Laboratory Systems*, 155:183190.

Thomas-Agnan, C., Laurent, T., Ruiz-Gazen, A., Nguyen, T.-H.-A., Chakir, R., and Lungarska, A. (2021). Spatial simultaneous autoregressive models for compositional data: Application to land use. In *Advances in Compositional Data Analysis*, pages 225–249. Springer, Cham.

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, J. J., Snchez-Arcilla, A., and Gomez, J. (2010). Classifying wave forecasts with model-based geostatistics and the aitchison distribution. *Stochastic Environmental Research and Risk Assessment*, 25:1091–1100.

Tolosana-Delgado, R., Pawlowsky-Glahn, V., and Egozcue, J. J. (2008). Indicator kriging without order relation violations. *Mathematical Geosciences*, 40(3):327–347.

Tolosana-Delgado, R. and van den Boogaart, K. G. (2013). Joint consistent mapping of high-dimensional geochemical surveys. *Mathematical geosciences*, 45(8):1167–1177.

Tolosana-Delgado, R., van den Boogaart, K. G., and Pawlowsky-Glahn, V. (2009). Estimating and modeling variograms of compositional data with occasional missing variables in r. *Project: Compositional data analysis methods and applications*.

Tolosana-Delgado, R., van den Boogaart, K. G., and Pawlowsky-Glahn, V. (2011). *Geostatistics for Compositions*, chapter 6, pages 73–86. John Wiley & Sons, Ltd.

Tsagris, M. (2015). Regression analysis with compositional data containing zero values. *Chilean Journal of Statistics*, 6(2):47–57.

Tsagris, M., Preston, S., and Wood, A. (2011). A data-based power transformation for compositional data. *Proceedings of the 4th Compositional Data Analysis Workshop, Girona, Spain*.

Tsagris, M., Preston, S., and Wood, A. (2016). Improved classification for compositional data using the α-transformation. *Journal of Classification*, 33:243–261.

Tsagris, M. and Stewart, C. (2020). A folded model for compositional data analysis. *Australian & New Zealand Journal of Statistics*, 62(2):249–277.

Wackernagel, H. (2003). *Multivariate geostatistics: an introduction with applications*. Springer Science & Business Media.

Walvoort, D. J. and de Gruijter, J. J. (2001). Compositional kriging: a spatial interpolation method for compositional data. *Mathematical Geology*, 33(8):951–966.

7 Supplementary Material

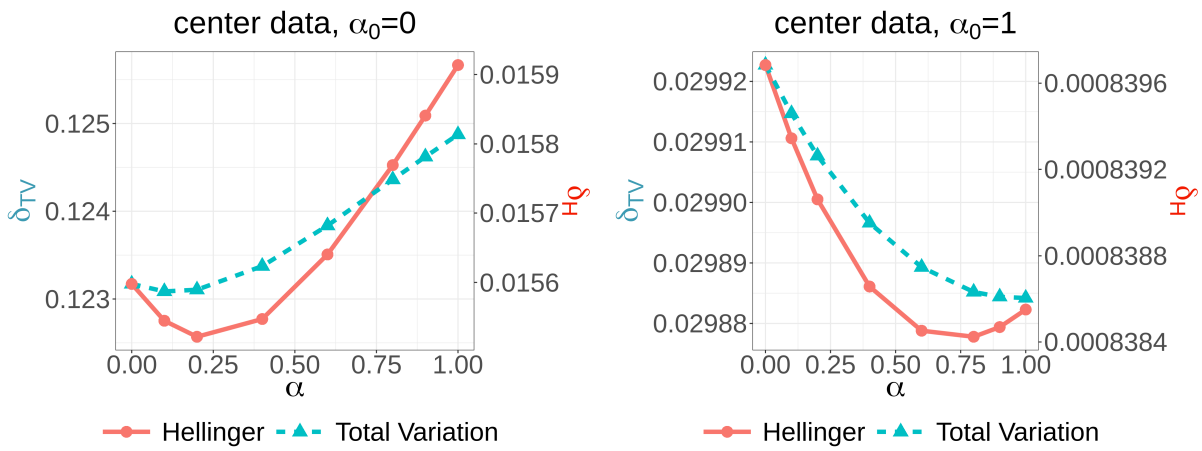

Figure S1: Average error over $B = 100$ tests for center data after α-IT with different α and $\alpha_0 = 0$ (left) and $\alpha_0 = 1$ (right): Total Variation metric in blue dashed lines and Hellinger metric in red continuous lines.

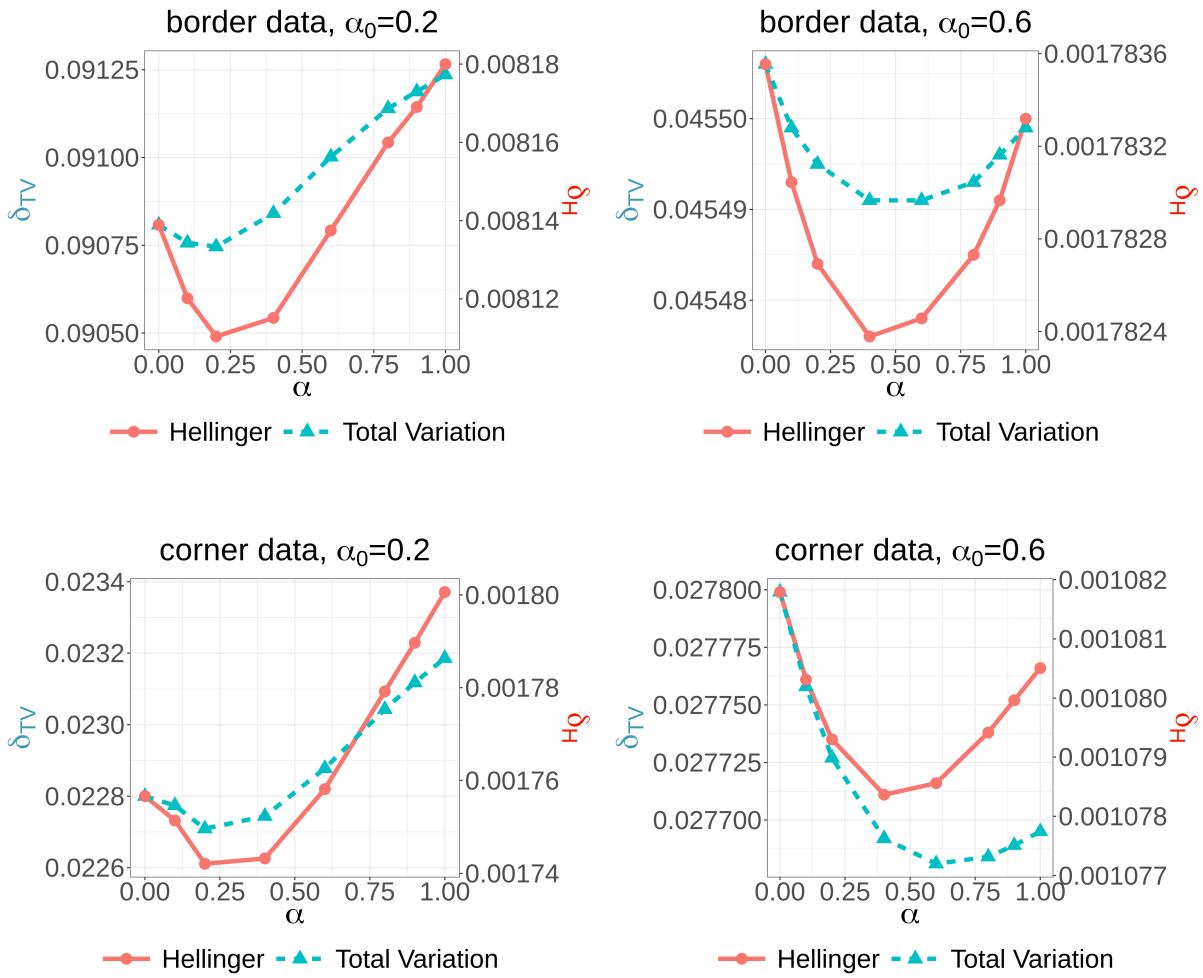

Figure S2: Average error over $B = 100$ tests for border (up) and corner (down) data after α-IT with different α and $\alpha_0 = 0.2$ (left) and $\alpha_0 = 0.6$ (right): Total Variation metric in blue dashed lines and Hellinger metric in red continuous lines.

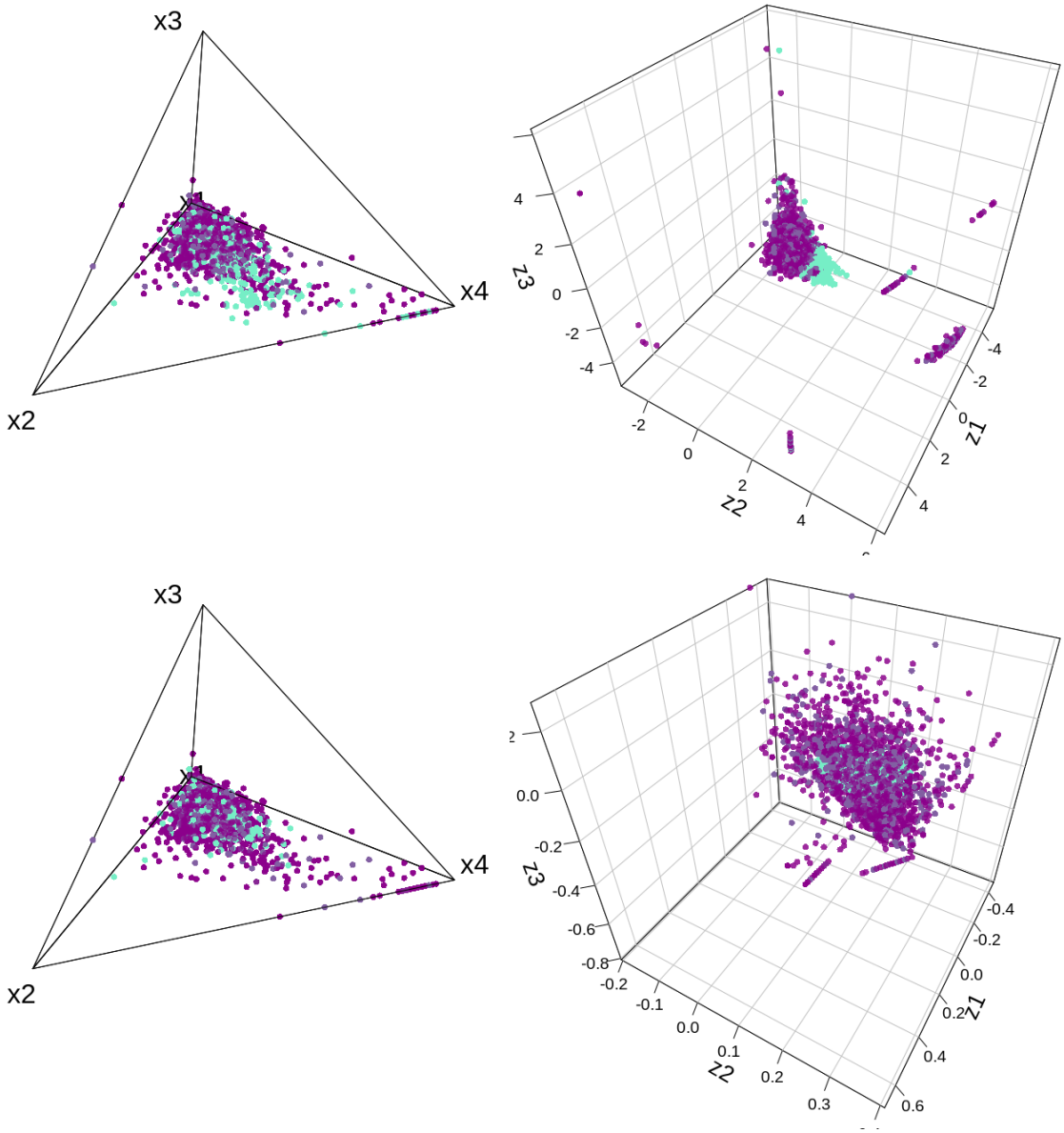

Figure S3: Original data (violet) and kriging predictions (light blue) with a moderate (10%) fraction of data with 0s (first setting). Left: simplex. Right: Euclidean space. Top: $\alpha^\star = 0.12$. Bottom: $\alpha = 1$.