AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE

10.1002/2015WR018369

Key Points:
- We derive a new projection strategy for the stochastic simulation of particle-size curves
- Algorithms to generate conditional realizations of the entire particle-size curve are introduced
- The procedure is demonstrated on particle-size curves sampled in a heterogeneous aquifer

Correspondence to:
A. Menafoglio,
alessandra.menafoglio@polimi.it

Citation:
Menafoglio, A., A. Guadagnini, and P. Secchi (2016), Stochastic simulation of soil particle-size curves in heterogeneous aquifer systems through a Bayes space approach, *Water Resour. Res.*, *52*, 5708–5726, doi:10.1002/2015WR018369.

Received 11 NOV 2015
Accepted 21 JUN 2016
Accepted article online 24 JUN 2016
Published online 2 AUG 2016

Stochastic simulation of soil particle-size curves in heterogeneous aquifer systems through a Bayes space approach

A. Menafoglio[1], A. Guadagnini[2,3], and P. Secchi[1]

[1]MOX, Department of Mathematics, Politecnico di Milano, Milano, Italy, [2]Department of Civil and Environmental Engineering, Politecnico di Milano, Milano, Italy, [3]Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, Arizona, USA

Abstract We address the problem of stochastic simulation of soil particle-size curves (PSCs) in heterogeneous aquifer systems. Unlike traditional approaches that focus solely on a few selected features of PSCs (e.g., selected quantiles), our approach considers the entire particle-size curves and can optionally include conditioning on available data. We rely on our prior work to model PSCs as cumulative distribution functions and interpret their density functions as functional compositions. We thus approximate the latter through an expansion over an appropriate basis of functions. This enables us to (a) effectively deal with the data dimensionality and constraints and (b) to develop a simulation method for PSCs based upon a suitable and well defined projection procedure. The new theoretical framework allows representing and reproducing the complete information content embedded in PSC data. As a first field application, we demonstrate the quality of unconditional and conditional simulations obtained with our methodology by considering a set of particle-size curves collected within a shallow alluvial aquifer in the Neckar river valley, Germany.

1. Introduction

Characterization of natural heterogeneity of aquifer bodies relies on diverse types of observations. These include, for example, direct measurements/estimates of hydraulic parameters such as hydraulic conductivity and porosity and data enabling us to infer a classification of soil types. Merging all available information within a unique theoretical and operational framework would form the basis for a robust system characterization. A stochastic approach is nowadays recognized as a viable tool to quantify how uncertainty propagates from incomplete knowledge of the properties of the host porous medium (in terms of spatial distribution of geomaterials and associated parameters) to state variables of interest (including, e.g., groundwater fluxes and chemical concentrations).

Here we present a new way according to which the information content embedded in particle-size curves (PSCs) can be employed to assist the stochastic characterization of a natural aquifer. These types of data are routinely available in field studies performed in diverse settings. They are usually obtained through relatively simple and inexpensive methods, such as traditional grain sieve analysis, sedigraph or laser diffraction methods. The information content which can be extracted from PSCs includes a set of representative particle diameters that are defined as average soil particle sizes corresponding to given quantiles of the PSC. Representative diameters can then be employed within existing empirical formulations relating them to aquifer parameters such as porosity and/or saturated hydraulic conductivity [e.g., *Rosas et al.*, 2014; *Vienken and Dietrich*, 2011; *Vukovic and Soro*, 1992]. In a few other cases [e.g., *Rogiers et al.*, 2012] a site-specific model is proposed to assess the possibility of estimating saturated hydraulic conductivity from the complete data set characterizing the PSCs. These can also be employed for the purpose of soil textural classification, according to a variety of approaches [e.g., *Riva et al.*, 2006; *Martìn et al.*, 2005, and references therein]. In this sense, texture data consisting in percentage values of sand, silt, and clay (which can be inferred from PSCs) can be employed together with other quantities, including, e.g., bulk density of soil, as input to pedotransfer functions to estimate soil hydraulic properties [e.g., *Rawls et al.*, 1982; *Pachepsky et al.*, 2006; *Schaap et al.*, 2001; *Schaap*, 2013, and references therein]. An alternative approach is grounded on concepts of similar media scaling [e.g., *Miller and Miller*, 1956; *Vogel et al.*, 1991] to exploit the dependence of hydraulic

properties on pore size and key geometrical descriptors of the pore space. The latter approach enables one to scale hydraulic properties of multiple soils to unique reference water retention curves and partially saturated relative hydraulic conductivity functions [e.g., among others, *Tuli et al*., 2001; *Das et al*., 2005; *Nasta et al*., 2013].

In this broad framework, hydrogeological investigations commonly employ a number of discrete quantiles of an available PSC which are subject to geostatistical analysis and then (a) mapped onto a spatial grid through Kriging or (b) employed in a numerical Monte Carlo setting to generate multiple realizations of the spatial distribution of aquifer properties and/or textural composition [e.g., *Riva et al*., 2006, 2008, 2010; *Hu et al*., 2009; *Bianchi et al*., 2011]. As recently pointed out by *Menafoglio et al*. [2014, 2015], these standard approaches suffer from two major drawbacks: (a) they require the joint geostatistical analysis of various characteristic particle diameters with an ordering constraint, entailing, e.g., calibration of multiple variogram and cross-variogram models, and (b) they do not fully exploit the richness of information associated with available PSCs.

It is then clear that having at our disposal advanced techniques for generating geostatistically based Monte Carlo realizations of an entire particle-size distribution instead of selected quantiles would dramatically improve our ability to represent and reproduce the complete information content embedded in PSC data. This technology has the clear potential to yield improved characterizations of the spatial variability and uncertainty associated with structural features of geomaterials forming natural aquifers, and can thus effectively support studies of groundwater flow and chemical transport. To the best of our knowledge, the challenging problem of performing geostatistical simulations of random spatial fields of soil particle-size distributions has not yet been explored in the literature.

The aim of this work is to provide the theoretical basis and the associated computational algorithms to generate Monte Carlo realizations of spatial distributions of PSCs. These can optionally be conditional on available observed PSCs at a set of discrete locations in the system. We do so by advancing our previous work [*Menafoglio et al*., 2014, 2015], within which we developed and applied a Functional Compositional Kriging predictor model for interpolating PSC data. We demonstrate here our new stochastic simulation approach through a field-scale analysis grounded on observed PSCs. We obtain (conditional and unconditional) realizations of PSC maps, which can readily be included in Monte Carlo simulations of groundwater flow and transport in randomly heterogeneous aquifer systems.

We ground our theoretical developments on a nonparametric framework, which combines the point of view of geostatistics [*Chilès and Delfiner*, 1999], functional data analysis (FDA) [*Ramsay and Silverman*, 2005] and compositional data analysis (CoDa) [*Pawlowsky-Glahn et al*., 2015]. Consistent with the concepts first introduced by *Menafoglio et al*. [2014, 2015], we model PSCs as cumulative distribution functions and interpret their densities, termed particle-size densities (PSDs), as functional compositions (FCs) belonging to a Bayes space [*Egozcue et al*., 2006]. FCs are positive functions constrained to integrate to a constant (e.g., probability density functions). They represent the infinite-dimensional (i.e., functional) counterpart of compositional data. The latter are positive multivariate data that represent proportions (or per cent amounts) of a total (e.g., unity or 100). We ensure the associated constraints (positivity, integration to one), through a log-ratio approach, which is well established in the multivariate setting and reflects the observation that the relevant information embedded in constant-sum objects is conveyed by (log-)ratios among components, rather than by their absolute values (i.e., the concept of *relative* information, see, e.g., *Aitchison* [1986]). We treat our FCs in a corresponding so-called Bayes space. The latter was designed to properly represent the data constraints (e.g., positivity, constant sum), and generalizes the Aitchison geometry to the functional case. In this context, we develop our stochastic simulation method for PSDs by relying upon a suitable and well defined projection strategy for FCs in Bayes spaces. This enables us to (a) reduce the dimensionality of the problem by guaranteeing a high level of precision and (b) characterize and simulate PSDs via an approximated multivariate problem.

The work is organized as follows. Section 2 describes the field data that are employed as a test bed to illustrate our methodology. Section 3 illustrates our stochastic simulation strategy in the unconditional and conditional settings. Section 4 describes our results obtained at the target field site. Section 5 concludes the work. The basic notions on Bayes space theory are given in Appendix A, and additional theoretical and algorithmic details are provided in Appendices B and C.

2. Experimental Site and Available Data

We consider here a data set obtained at the Lauswiesen site, located in the Neckar river valley near the city of Tübingen, Germany. The subsurface system in the area has been characterized through extensive information obtained at a number of boreholes, which are employed to perform sedimentological as well as hydraulic analyses. A relatively regular upper clay layer with a thickness of 1–2 m overlies a conductive Quaternary sand and gravel deposit. The latter rests on a layer of Keuper marl which is considered to define an impervious bedrock boundary of the aquifer hosted in the Quaternary sand and gravel system. The saturated thickness of the aquifer we are considering is approximately 5 m. All boreholes penetrate the aquifer down to bedrock. Details of site hydrogeology are given by *Riva et al*. [2006] and references therein. Available pumping test data have been employed by *Neuman et al*. [2007] for the stochastic analysis of late-time drawdowns and by *Panzeri et al*. [2015] for the application of data assimilation techniques based on the concept of Moment Equation Ensemble Kalman Filter.

Of specific relevance to our study are the available 406 PSCs sampled along 12 fully penetrating vertical boreholes. The data set was employed by *Riva et al*. [2006, 2008, 2010], *Barahona-Palomo et al*. [2011], and *Riva et al*. [2014] in the context of stochastic modeling studies aimed at (a) providing a probabilistic analysis of solute residence times within well capture zones, (b) interpreting an observed tracer test in a numerical Monte Carlo framework, (c) assessing the link between the spatial covariance functions of the (natural) logarithm of hydraulic conductivity and of soil particle representative diameters, and (d) characterizing the correlation between hydraulic conductivity values estimated through impeller flowmeter downhole measurements and by way of empirical formulations based on PSC representative diameters.

The available PSCs were measured on soil samples of characteristic length ranging from 5 to 26.5 cm. A number of 12 sieve diameters (i.e., 0.063, 0.125, 0.25, 0.50, 1.0, 2.0, 4.0, 8.0, 16.0, 31.5, 63.0, and 100.0 mm) were employed in the sieve analysis procedure. Figure 1c depicts a sketch of the borehole network and sampling locations at the site. Applying traditional empirical relationships between characteristic soil diameters and permeability indicates that the site is mainly constituted by heterogeneous and conducive deposits of alluvial origin.

Particle-size curves associated with one of the available boreholes (borehole B5 in Figure 1) have been employed by *Menafoglio et al*. [2014] to perform a geostatistical analysis of PSCs through the corresponding densities, interpreted as functional compositions (FCs). These authors embed this latter concept within the geostatistical framework of *Menafoglio et al*. [2013] through which they provide Kriging estimates of the full PSC on a computational grid, together with the associated Kriging variance. The geostatistical setting of *Menafoglio et al*. [2014] has been extended by *Menafoglio et al*. [2015] to characterize the complete set of PSCs at the site and to properly account for the information content related to the local occurrence of diverse soil types (or textural classes). The key result of the authors is the formulation of an original theoretical framework according to which one can take full advantage of the complete set of information embedded in measured PSCs to (a) classify PSCs into clusters which represent the occurrence at a site of diverse soil types, (b) characterize the spatial distribution of each identified textural class, and (c) provide Kriging estimates of the heterogeneous distribution of PSCs within each region which contributes to form the internal architecture of the geological system.

Menafoglio et al. [2014, 2015] analyze available PSDs by resorting to a smoothing procedure suitable for PSCs. This enabled them to obtain the smooth estimates of PSDs from raw data (Figures 1a and 1b), and to embed these in their analyses leading to Kriging predictors of the PSDs at unsampled locations. Note that data considered in this work refer to the particle-size distribution within the domain of available observation, i.e., associated with the grain dimensions between the minimum and the maximum sieve diameters. For the purpose of illustration, we here consider a subset of the smoothed data of *Menafoglio et al*. [2015], as detailed in section 4; the reader is referred to *Menafoglio et al*. [2015] for further details on data preprocessing.

3. A Projection Strategy for the Stochastic Simulation of Particle-Size Densities

This section illustrates the theoretical elements underpinning our approach. The key idea is to generate stochastic realizations of spatial functional data. We are not interested in simulating on a computational grid only a discrete number of points of these functions but rather the full function. We then ground our method on a projection of these functions onto a suitable functional basis and develop generation algorithms that consider the coefficient of such a basis expansion. For simplicity, we describe here the general points of the approach and devote Appendices A–C to the details of the mathematical developments.

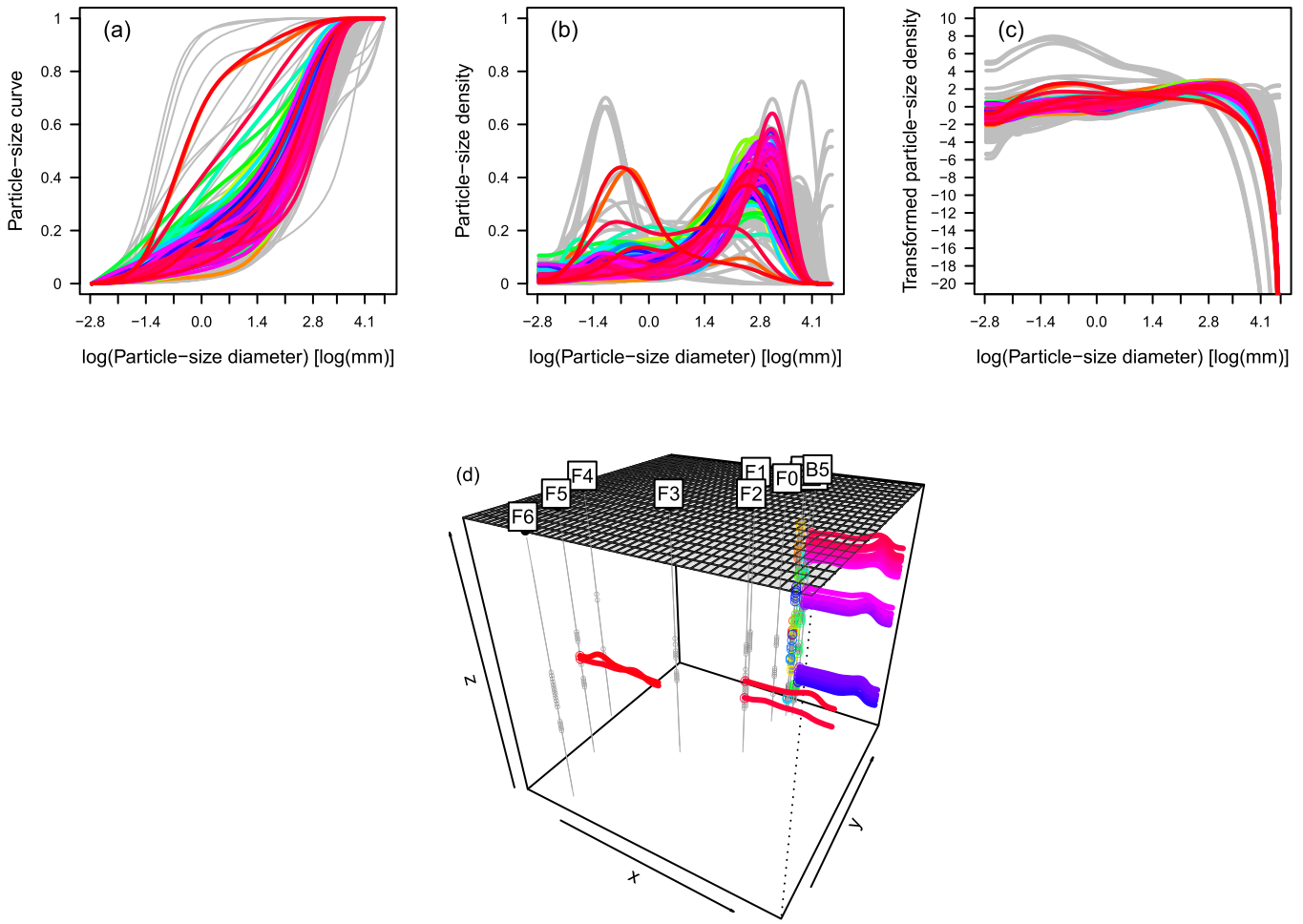

Figure 1. Smoothed conditional (a) PSCs, (b) PSDs, and (c) centered log-ratio transform of PSD used to perform computations (colored curves are associated with the subsample employed for our demonstration); (d) sketch of the sampling scheme at the site and PSDs along boreholes B5, F2, and F5. In Figure 1d, *x* coordinates range in [3508459, 3508712], *y* coordinates in [5377622, 5377779], and *z* coordinates in [300.331, 308.924].

3.1. Notation and Background

We denote by $D \subset \mathbb{R}^3$ the three-dimensional aquifer domain. Let $\mathcal{X}_{\boldsymbol{s}}$ be a (random) particle-size curve, associated with location $\boldsymbol{s}$ in D: for any soil particle size t in the closed interval $\mathcal{T}=[t_{\min}, t_{\max}]$, $\mathcal{X}_{\boldsymbol{s}}(t)$ denotes the random proportion of particles having size smaller than or equal to t. We denote by $\{\mathcal{X}_{\boldsymbol{s}}, \boldsymbol{s} \in D\}$ the random field of PSCs, that is a collection of random functional elements (i.e., the PSCs), indexed by the continuous spatial variable $\boldsymbol{s}$ in D. In this sense, $\{\mathcal{X}_{\boldsymbol{s}}, \boldsymbol{s} \in D\}$ is a *functional* random field.

Let $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$ be sampling/measurement locations in D. Given the observation of $\mathcal{X}_{\boldsymbol{s}_1}, \ldots, \mathcal{X}_{\boldsymbol{s}_n}$ at these locations, our goal is to provide a collection of stochastic simulations (or realizations) of the PSC $\mathcal{X}_{\boldsymbol{s}_0}$ at a given location $\boldsymbol{s}_0$ in D. These simulations may be either *unconditional* or *conditional*. The former are realizations sampled from the (estimated) distribution of the field $\{\mathcal{X}_{\boldsymbol{s}}, \boldsymbol{s} \in D\}$, whereas the latter are realizations from the (estimated) conditional distribution of $\{\mathcal{X}_{\boldsymbol{s}}, \boldsymbol{s} \in D\}$ given the observations $\mathcal{X}_{\boldsymbol{s}_1}, \ldots, \mathcal{X}_{\boldsymbol{s}_n}$. Conditional simulations reproduce the actual data at the measurement locations.

If in each location in D we only considered a quantile (or the mean) of the distribution (e.g., the median particle-size $X_{\boldsymbol{s}}$ of the PSC $\mathcal{X}_{\boldsymbol{s}}$) we would work with a real-valued random field $\{X_{\boldsymbol{s}}, s \in D\}$, for which multivariate geostatistical methods of analysis, estimation and simulation are well known [e.g., among others, *de Marsily*, 1986; *Deutsch and Journel*, 1998; *Chilès and Delfiner*, 1999].

Stochastic simulations of the PSCs may rely on the discretization of each particle-size distribution as a list of point evaluations of the PSC $(\mathcal{X}_{\boldsymbol{s}}(t_1), \ldots, \mathcal{X}_{\boldsymbol{s}}(t_K))$ (or a list of quantiles). Accordingly, one can then produce

either unconditional or conditional realizations of these quantities, by employing, for example, well-established Gaussian cosimulation methods [e.g., *Deutsch and Journel*, 1998; *Remy et al*., 2009].

Note that refining the discretization (i.e., increasing the number of particle sizes at which the PSC is evaluated), leads to improved approximations of the PSC through the corresponding vector $(\mathcal{X}_{\boldsymbol{s}}(t_1),\ldots,\mathcal{X}_{\boldsymbol{s}}(t_K))$. Nevertheless, this approach suffers from two critical drawbacks: (i) the so-called curse of dimensionality and (ii) the ordering constraints. With reference to the former, we note that an increased refinement in the discretization of a PSC is associated with a corresponding increase of the computational burden to produce a random realization. Ideally, if one aimed at obtaining a realization of the entire distribution function (i.e., of the entire PSC), cosimulation of an infinity of point values would be needed. With reference to the latter point, it is clear that PSCs are associated with an ordering relation, i.e., by definition, $\mathcal{X}_{\boldsymbol{s}}(t_i) < \mathcal{X}_{\boldsymbol{s}}(t_j)$, for $i<j$. These constraints should be properly considered to obtain admissible results. This is especially critical when dealing with a fine grid of evaluation in the domain $[t_{\min}, t_{\max}]$, since in this case the violation of the ordering constraints is likely to take place, because of the closeness of the values taken by the PSC at two consecutive points of evaluations, i.e., of $\mathcal{X}_{\boldsymbol{s}}(t_i)$, $\mathcal{X}_{\boldsymbol{s}}(t_{i+1})$, for $i=1,\ldots,K-1$.

Our approach is grounded on the idea that one can tackle the curse of dimensionality by interpreting PSC data as functions that can be approximated as a combination of a low number of functional components. The projection of the curves onto these components allows reducing functions to vectors of coefficients (i.e., coordinates associated with the basis expansion). The latter can be then analyzed via multivariate methods, including, e.g., frequently used standard techniques for Gaussian cosimulation in a geostatistical framework. Constraints imposed by the nature of the data analyzed are treated by selecting a proper functional space, together with a suitable set of functional basis elements.

We tackle these challenges by following the approach of *Menafoglio et al*. [2014, 2015], which is a generalization to the functional setting of the strategy of *Tolosana-Delgado et al*. [2008]. These authors rely upon the so-called Aitchison geometry [*Pawlowsky-Glahn and Egozcue*, 2001] to analyze spatial compositional data, i.e., vectors whose components express proportions or percent amount of a whole (e.g., discrete probability density functions). We work here within the theory of Bayes spaces [*Egozcue et al*., 2006; *van den Boogaart et al*., 2014] through which the Aitchison geometry can be generalized to functional compositions (FCs), i.e., positive functions integrating to a constant (e.g., probability density functions). Note that particle-size densities (PSDs)—i.e., the derivative of PSCs—can be interpreted within this framework. From a mathematical viewpoint, functional compositions are points in Bayes spaces, where proper notions of sum and product by a constant, inner product and norm are defined, in agreement with the so-called principles of compositional data analysis. For the purpose of our discussion, we do not present the complex mathematical constructions involved in the introduction of Bayes spaces. We limit our illustration to mention that it is possible to introduce a transformation based on logarithms, that allows preserving the constraints of PSDs (i.e., positivity and integration to 1). Let us denote by $\{\mathcal{Y}_{\boldsymbol{s}}, \boldsymbol{s}\in D\}$ the random field whose generic element $\mathcal{Y}_{\boldsymbol{s}}$ is the PSD associated with the PSC $\mathcal{X}_{\boldsymbol{s}}$ (i.e., $\mathcal{Y}_{\boldsymbol{s}}$ is the derivative of $\mathcal{X}_{\boldsymbol{s}}$ with respect to t). We consider for each element $\mathcal{Y}_{\boldsymbol{s}}$ of this field its *centered log-ratio* (clr) transformation $\mathcal{Z}_{\boldsymbol{s}}$

$$\mathcal{Z}_{\boldsymbol{s}}(t)=clr(\mathcal{Y}_{\boldsymbol{s}})(t)=\ln\left(\mathcal{Y}_{\boldsymbol{s}}(t)\right)-\frac{1}{t_{\max}-t_{\min}}\int_{t_{\min}}^{t_{\max}}\ln\left(\mathcal{Y}_{\boldsymbol{s}}(\tau)\right)d\tau,\quad t\in\mathcal{T}. \tag{1}$$

Transformation (1) enables us to overcome issues related to the data constraints by mapping the original constrained problem (PSDs must be positive and integrate to 1) to an unconstrained problem (the clr-transformations can take arbitrary values). Note that preserving the positivity constraint of the PSDs leads to honor the ordering relation among quantiles (or evaluations) of the associated PSCs. From the mathematical viewpoint, transformation (1) maps FCs from a particular Bayes space to the space $L^2(\mathcal{T})$ of square-integrable functions on $\mathcal{T}$, endowed with the usual notion of sum and product by a constant, inner product and norm (see Appendix B for further details).

In the remaining part of this section, we introduce the mathematical construction for the field $\{\mathcal{Z}_{\boldsymbol{s}}, \boldsymbol{s}\in D\}$ of curves in L^2, obtained through the transformation in (1). We show in Appendix B that it is possible to express the entire construction in the geometry of the Bayes space, without necessarily invoking a clr-transformation.

3.2. Mathematical Construction

We assume that $\{\mathcal{Z}_{\boldsymbol{s}}, \boldsymbol{s}\in D\}$ is a stationary Gaussian random field of in L^2, with constant spatial mean m

$$m(t)=\mathbb{E}[\mathcal{Z}_{\boldsymbol{s}}(t)], \quad t \in \mathcal{T}, \boldsymbol{s} \in D, \tag{2}$$

and stationary cross-covariance operator C:

$$C(\boldsymbol{s}_1-\boldsymbol{s}_2)x=\mathbb{E}\left[\left(\left(\int_{\mathcal{T}}(\mathcal{Z}_{\boldsymbol{s}_1}(\tau)-m(\tau))x(\tau)d\tau\right)(\mathcal{Z}_{\boldsymbol{s}_2}-m)\right], \quad x \in L^2, \boldsymbol{s}_1, \boldsymbol{s}_2 \in D. \tag{3}$$

For the field $\{\mathcal{Z}_{\boldsymbol{s}}, \boldsymbol{s} \in D\}$, we consider the following truncated expansion over an orthonormal functional basis $\{v_k, k \geq 0\}$ (i.e., a basis of L^2 such that $\int_{\mathcal{T}} v_j(\tau)v_k(\tau)d\tau=1$ if $j=k$, 0 otherwise)

$$\mathcal{Z}_{\boldsymbol{s}}^K=m+\sum_{k=1}^{K}\xi_k(\boldsymbol{s})v_k, \tag{4}$$

where $\xi_k(\boldsymbol{s})=\int_{\mathcal{T}}(\mathcal{Z}_{\boldsymbol{s}}(\tau)-m(\tau))v_k(\tau)d\tau$ is the (random) projection of $\mathcal{Z}_{\boldsymbol{s}}^K$ onto the basis function v_k, and K is a given truncation order. In this setting, each element $\mathcal{Z}_{\boldsymbol{s}}$ can be represented through a K-dimensional vector of coefficients $\boldsymbol{\xi}(\boldsymbol{s})=(\xi_1(\boldsymbol{s}), \ldots, \xi_K(\boldsymbol{s}))^T$, with respect to the truncated basis $\{v_k, 1 \leq k \leq K\}$.

Given a truncation order K, we denote by $\{\mathcal{Z}_{\boldsymbol{s}}^K, \boldsymbol{s} \in D\}$ the random field whose elements are given by (4). The distributional properties of this field are determined by m and by those of the zero-mean multivariate random field $\{\boldsymbol{\xi}(\boldsymbol{s}), \boldsymbol{s} \in D\}$. It is noted that (a) $\{\boldsymbol{\xi}(\boldsymbol{s}), \boldsymbol{s} \in D\}$ is a Gaussian random field in $\mathbb{R}^K$ by virtue of the Gaussian assumption on the field $\{\mathcal{Z}_{\boldsymbol{s}}, \boldsymbol{s} \in D\}$; (b) the covariance operator of the field $\mathcal{Z}_{\boldsymbol{s}}^K$ can be expressed in terms of the covariograms and cross-covariograms of the multivariate field of coordinates $\{\boldsymbol{\xi}(\boldsymbol{s}), \boldsymbol{s} \in D\}$ (see Appendix B for details).

Our strategy to obtain either conditional or unconditional simulations of the field $\{\mathcal{Z}_{\boldsymbol{s}}, \boldsymbol{s} \in D\}$ is to resort to approximation (4) for an appropriate order K and then to perform simulations of the multivariate random field $\{\boldsymbol{\xi}(\boldsymbol{s}), \boldsymbol{s} \in D\}$.

For any given tolerance, one can determine a truncation order K such that $\mathcal{Z}_{\boldsymbol{s}}^K$ approximates $\mathcal{Z}_{\boldsymbol{s}}$ with a desired precision (in the mean square sense), uniformly in D (see Appendix B for details). In principle, setting a large value for parameter K would be preferable, to obtain improved approximations of $\mathcal{Z}_{\boldsymbol{s}}$ through $\mathcal{Z}_{\boldsymbol{s}}^K$. However, the value of K has a dramatic effect on the computational cost which is required for the simulation because it controls the dimensionality of the field $\{\boldsymbol{\xi}(\boldsymbol{s}), \boldsymbol{s} \in D\}$. Thus, one needs to consider a balance between limited computational power and accuracy.

We also note that the quality of a Kth order approximation in (4) varies according to the basis $\{v_k, k \geq 1\}$ which is employed. The best Kth order approximation (in the mean square sense) is attained when considering as a functional basis the set of the first K eigenfunctions of $C(\boldsymbol{0})$, $w_1, \ldots, w_K$ (called functional principal components, FPCs). The latter are obtained by solving the eigenequations

$$C(\boldsymbol{0})w_k=\lambda_k w_k, \quad k=1, 2, \ldots \tag{5}$$

where $\lambda_1 > \lambda_2 > \ldots > \lambda_K > \ldots$ are the eigenvalues of $C(\boldsymbol{0})$. The eigenvalue λ_k $(k=1, 2, \ldots, K)$ then represents the proportion of the total variability of the data which is captured by projecting the data along direction w_k. As in multivariate principal component analysis, one can then set the truncation order K as the minimum order that allows explaining a given amount of the total variability (e.g., 90% or 95%). Otherwise, depending on the case analyzed, K can be identified as the minimum order at which an elbow starts to appear in the so-called scree plot, that displays the proportion of variability explained by the first K eigenfunctions as a function of K.

In most studies, the zero-lag covariance operator is not known a priori. In this case, one can apply an empirical version of the proposed strategy, i.e., (a) estimate from available data the zero-lag covariance operator $C(\boldsymbol{0})$ through the empirical estimator

$$Sx=\frac{1}{n}\sum_{i=1}^{n}\left[\int_{\mathcal{T}}(\mathcal{Z}_{\boldsymbol{s}_i}(\tau)-\hat{m}(\tau))x(\tau)d\tau\right](\mathcal{Z}_{\boldsymbol{s}_i}-\hat{m}), \quad x \in L^2, \tag{6}$$

$\hat{m}=\frac{1}{n}\sum_{i=1}^{n}\mathcal{Z}_{\boldsymbol{s}_i}$ denoting the sample mean, (b) compute the eigen-pairs $(\hat{\lambda}_k, \hat{w}_k)$, $k=1, \ldots, n-1$, of this estimate, and (c) project the observations on the first K eigenfunctions (or empirical functional principal components, EFPCs) [*Ramsay and Silverman*, 2005] of S to obtain the representation

$$\mathcal{Z}_{\boldsymbol{s}_i} \approx \hat{m} + \sum_{k=1}^{K} \hat{\xi}_k(\boldsymbol{s}_i)\hat{w}_k. \tag{7}$$

Here $\hat{\xi}_k(\boldsymbol{s}_i) = \int_{\mathcal{T}} (\mathcal{Z}_{\boldsymbol{s}_i}(\tau) - \hat{m}(\tau))\hat{w}_k(\tau)d\tau$ is called *score* and is the projection of $(\mathcal{Z}_{\boldsymbol{s}_i} - \hat{m})$ along the kth EFPC $\hat{w}_k$. Note that the (empirical) FPCA is the infinite-dimensional counterpart of principal components analysis, which is widely employed in the multivariate framework to perform optimal dimensionality reduction of a multivariate data set. In general, most of the techniques that are commonly employed in the multivariate framework to identify and interpret principal components can be extended to the functional setting, as shown by *Ramsay and Silverman* [2005]. We also remark that EFPCA is equivalent to perform an empirical functional principal component analysis in the Bayes space, called simplicial functional principal component analysis (SFPCA) [see *Hron et al*., 2016]. In particular, the back transformation of the EFPCs $\hat{w}_k$ via the inverse clr

$$e_k(t) = clr^{-1}(\hat{w}_k)(t) = \frac{\exp(\hat{w}_k(t))}{\int_{\mathcal{T}} \exp(\hat{w}_k(\tau))d\tau}, \quad t \in \mathcal{T}, \quad k = 1, \ldots, K, \tag{8}$$

defines the functional components upon which the PSDs are actually projected in the Bayes space, and can be employed for interpretation purposes, as detailed in section 4.

Given the optimal expansion (7), one can then employ multivariate techniques [e.g., *Chilès and Delfiner*, 1999; *Mariethoz and Caers*, 2015] to perform unconditional or conditional (geostatistical) stochastic simulations of the K-dimensional vectors of scores $\hat{\boldsymbol{\xi}}(\boldsymbol{s}_i) = \left(\hat{\xi}_1(\boldsymbol{s}_i), \ldots, \hat{\xi}_K(\boldsymbol{s}_i)\right)^T$. Here we illustrate the application of our approach to a field case by employing the multivariate Gaussian simulator available in the package gstat [*Pebesma*, 2004] of software R [*R Core Team*, 2013]. Conditional simulations of section 4.4 are based on the sequential Gaussian method of *Abrahamsen and Benth* [2001]. It is remarked that any multivariate simulation method could be employed as well, without substantial modifications to the overall strategy here proposed.

The (random) realization $\mathcal{Z}^*_{\boldsymbol{s}_0}$ is obtained as

$$\mathcal{Z}^*_{\boldsymbol{s}_0} = \hat{m} + \sum_{k=1}^{K} \hat{\xi}^*_k(\boldsymbol{s}_0)\hat{w}_k, \tag{9}$$

$\hat{\xi}^*_k(\boldsymbol{s}_0)$ denoting the kth element of a realization of vector $\hat{\boldsymbol{\xi}}(\boldsymbol{s}_0)$. The realization $\mathcal{Y}^*_{\boldsymbol{s}_0}$ of the PSD at a target location $\boldsymbol{s}_0$ is finally obtained by mapping back $\mathcal{Z}^*_{\boldsymbol{s}_0}$ to $\mathcal{Y}^*_{\boldsymbol{s}_0}$ FCs, through the inverse of the clr-transformation

$$\mathcal{Y}^*_{\boldsymbol{s}_0}(t) = clr^{-1}(\mathcal{Z}^*_{\boldsymbol{s}_0})(t) = \frac{\exp(\mathcal{Z}^*_{\boldsymbol{s}_0}(t))}{\int_{\mathcal{T}} \exp(\mathcal{Z}^*_{\boldsymbol{s}_0}(\tau))d\tau}, \quad t \in \mathcal{T}. \tag{10}$$

4. Example of Application: Simulation of Particle-Size Densities at the Lauswiesen Test Site

We illustrate here our methodology for the simulation of PSDs on the basis of field data presented in section 2. As a test bed, we consider the subset of the complete data set depicted in Figure 1, formed by 100 PSDs randomly sampled from the set of data belonging to the second cluster singled out by *Menafoglio et al*. [2015]. As a first step, we transform the data via the clr-transformation (1) and obtain the curves depicted in Figure 1c. We thus apply EFPCA to this data set in section 4.1 and obtain the best empirical basis for the representation of the (transformed) data. In the following sections we illustrate the results of unconditional and conditional simulation at the site.

4.1. Functional Principal Component Analysis of PSDs at the Field Site

Following the approach based on clr transform described in section 3, we perform EFPCA of the data set depicted in Figure 1c. For the sake of simplicity, we estimate the mean m via the sample estimator $\hat{m} = \frac{1}{n}\sum_{i=1}^{n} \mathcal{Z}_{\boldsymbol{s}_i}$; more refined estimates may be employed (e.g., via generalized least squares) [*Menafoglio et al*., 2013, 2014]. Figure 2 depicts the key results of the analysis. For ease of interpretation, in Figures 2e–2h, we depict the results back-transformed to PSDs (i.e., to the Bayes space) through equation (8).

Based on the scree plot in Figure 2a and on the scores boxplots in Figure 2b, we set the truncation order to $K = 4$. This choice enables us to explain 97% of the total variability of the data set. The first $K = 4$ EFPCs $\{\hat{w}_1, .., \hat{w}_4\}$ and their PSD counterparts $\{\hat{e}_1, \ldots, \hat{e}_4\}$ are depicted in Figures 2c and 2d, respectively. Figures 2e–2h depict (in the space of densities) the mean function plus/minus the eigenfunctions multiplied by twice the standard deviation along the corresponding direction, i.e., $\hat{m} \pm 2\sqrt{\hat{\lambda}_k}\hat{w}_k$, $k = 1, \ldots, 4$. The curves in Figures 2e–2h are representative of the patterns characterizing the observations presenting high/low scores along the corresponding EFPCs, when compared with the mean $\hat{m}$. In this sense, the first EFPC, $\hat{w}_1$ (or SFPC $\hat{e}_1$), captures the variability in the position of the mode and in the mass concentration around it. High scores along EFPC $\hat{w}_1$ are represented by the blue curve in Figure 2e, which depicts a PSD with larger mode and higher mass concentration than the mean, the opposite behavior being depicted as a red curve in Figure

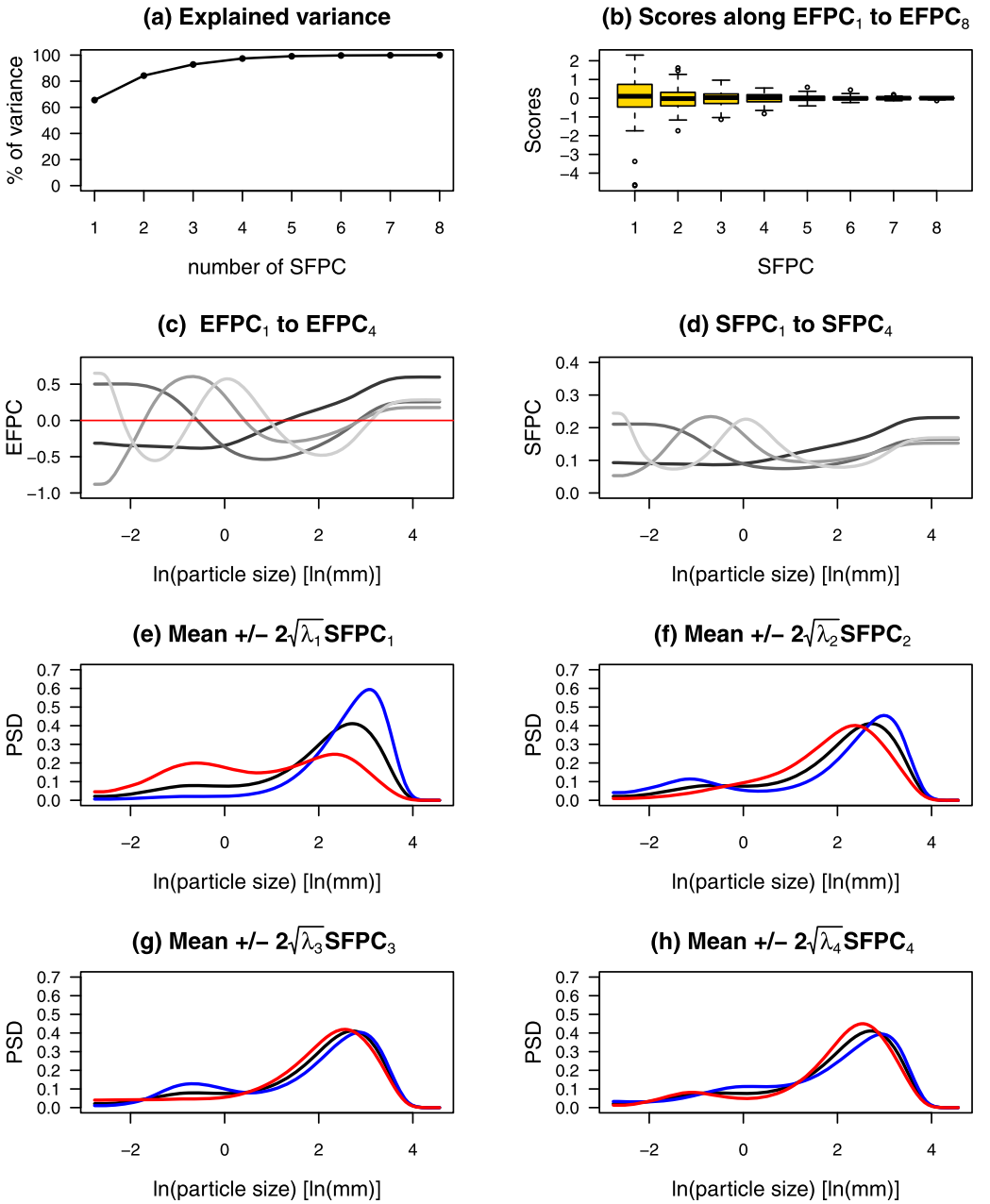

Figure 2. Results of EFPCA on the data set of PSDs. (a) Scree plot. (b) Boxplots of the scores along the first eight EFPCs. (c) First four EFPCs, $\hat{w}_1, \ldots, \hat{w}_4$. (d) First four SFPCs, $\hat{e}_1, \ldots, \hat{e}_4$. (e-h) The solid black curve indicates the mean function, the red and blue curves respectively indicate, in the space of densities, the mean plus or minus the EFPCs multiplied by twice the square root of the corresponding eigenvalue.

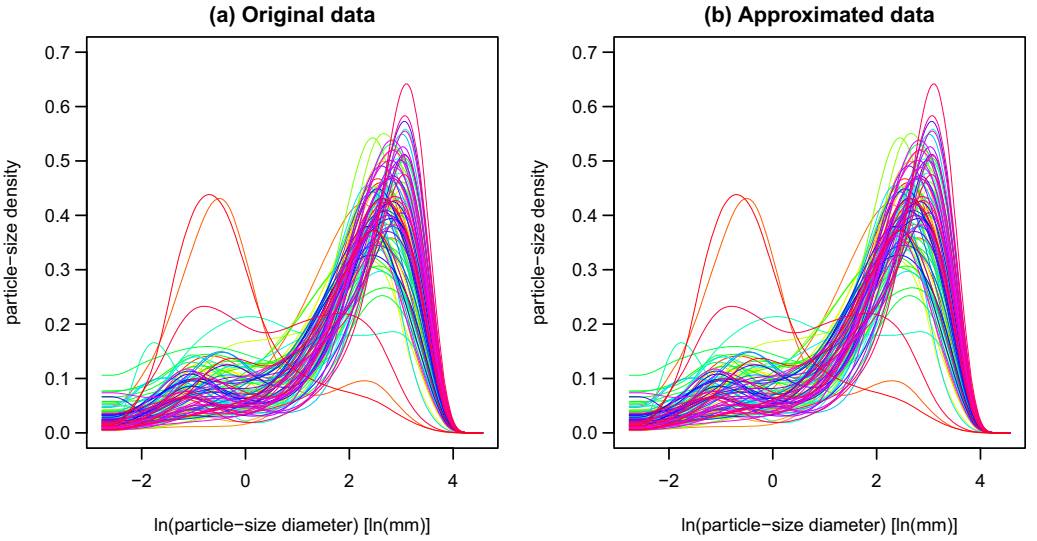

Figure 3. Original data set and approximated PSDs obtained via (7).

2e. The second EFPC, $\hat{w}_2$ (or SFPC $\hat{e}_2$), is interpreted in terms of the modality of the distribution (Figure 2f): high scores along the EFPC $\hat{w}_2$ are registered for bimodal densities (blue curve), whereas low scores are associated with unimodal distributions. A correspondingly strong interpretation for the remaining EFPCs is not emerging as clearly as for the first two EFPCs.

Figures 3a and 3b compare the original data and their approximation based on the truncated expansion (7) with $K = 4$. Inspection of Figure 3 allows recognizing that the approximated curves provide a viable reproduction of all the main features of the original densities.

Finally, Figure 4 depicts the scatter plot of the scores along the retained functional components (colors are consistent with the curves in Figure 3).

4.2. Geostatistical Modeling of the Scores

Once the approximation (7) has been obtained, stochastic simulation of a PSD $\mathcal{Y}_{\boldsymbol{s}_0}$ at a target location $\boldsymbol{s}_0$ in D requires the geostatistical characterization of the vectors of scores $\hat{\xi}(\boldsymbol{s}_1), \ldots, \hat{\xi}(\boldsymbol{s}_n)$. Consistent with the assumption of section 3, we consider $\hat{\xi}(\boldsymbol{s}_1), \ldots, \hat{\xi}(\boldsymbol{s}_n)$ to be a partial observation of a K-dimensional stationary Gaussian random field $\{\hat{\xi}(\boldsymbol{s}), \boldsymbol{s} \in D\}$. Following *Menafoglio et al.* [2015], we consider a geometric anisotropy at the site, characterized by anisotropy ratio of $R = 0.04$ between the horizontal and vertical directions. Thus, hereafter we refer all our estimates and simulated quantities to an isotropic spatial domain obtained by dilation of the actual vertical coordinate by a factor $1/R = 25$. In this context, omnidirectional variograms are estimated. Figure 5 depicts the variograms and cross-variograms estimated from the scores $\hat{\xi}(\boldsymbol{s}_1), \ldots, \hat{\xi}(\boldsymbol{s}_n)$. We fit a valid model to these estimates by employing a Linear Model of Coregionalization (LMC) [e.g., *Chilès and Delfiner*, 1999] based on an exponential model with nugget. We note that speed up of computations could be achieved upon employing simplifying assumptions on the vector of scores, e.g., by modeling the fields $\{\hat{\xi}_k(\boldsymbol{s}), \boldsymbol{s} \in D\}$, $k = 1, \ldots, K$, as uncorrelated. This simplifying assumption might be considered as a viable approximation at the site on the basis of the results depicted in Figure 5. For the sake of completeness, in our application described in the following sections, we prefer to consider the complete LMC estimated as in Figure 5.

4.3. Unconditional Simulation of PSDs

We illustrate an example of unconditional simulation of PSDs by considering a two-dimensional computational grid $D_0 \subset D$ which comprises 625 points, at a fixed elevation of 300 m asl. Based on the LMC estimated in section 4.2, we perform unconditional Gaussian cosimulation of the K-dimensional vectors $\hat{\xi}(\boldsymbol{s}_0)$, $\boldsymbol{s}_0 \in D_0$. Figure 6 depicts a selected realization simulated on the grid D_0 according to the proposed methodology.

We test the quality of the simulation by generating $NMC = 1000$ Monte Carlo replicates of the field on D_0. The CPU time required for the computations based on the R package gstat, within R version 3.0.2 was approximately 70′55″ (CPU time refers to an Intel® Core™ i7-3517U CPU @ 1.90 GHz). We then compute the empirical

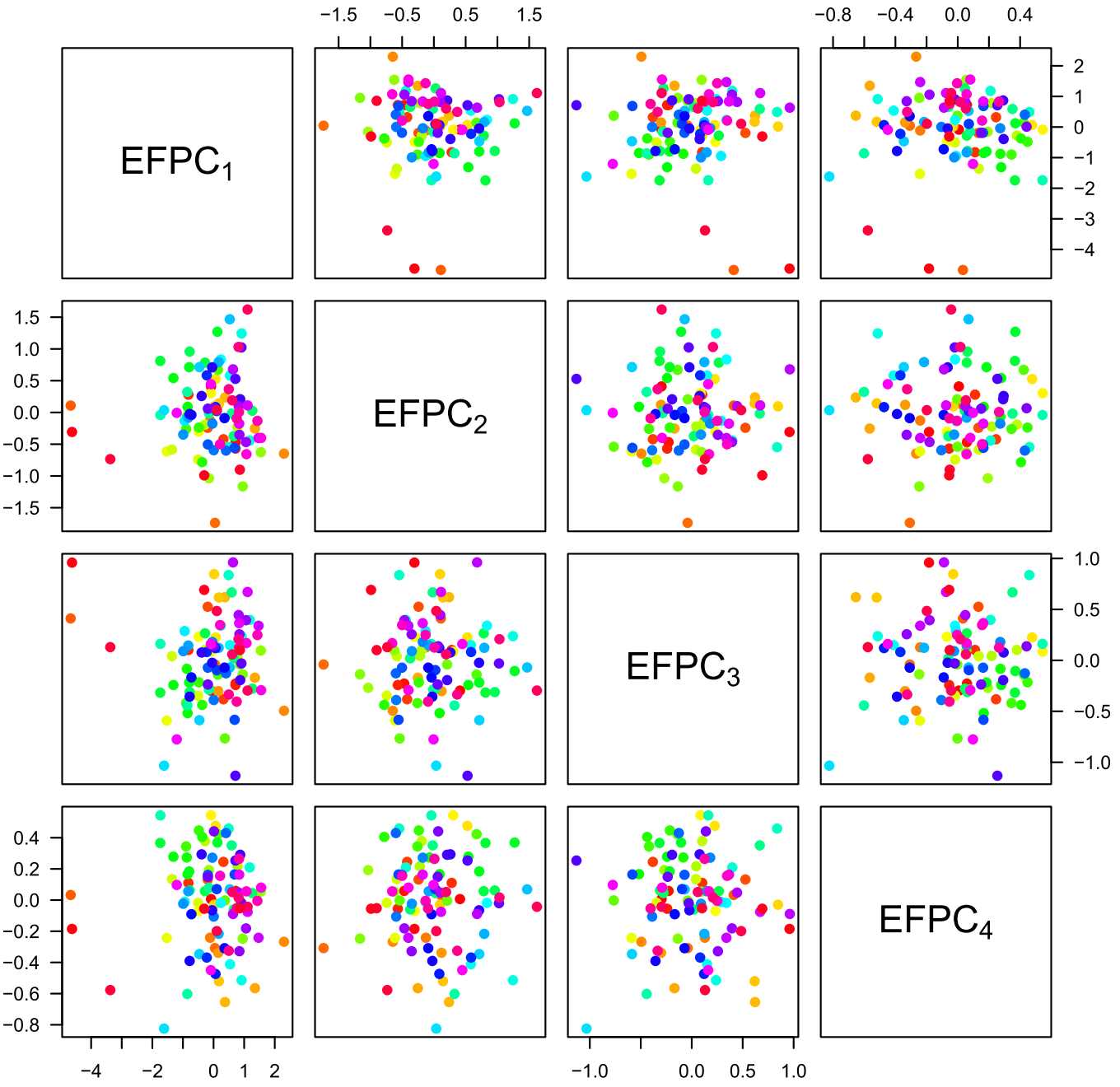

Figure 4. Scores along the estimated EFPCs, $\hat{w}_1, \ldots, \hat{w}_4$. Colors of the symbols are consistent with those of the curves in Figure 3.

variogram associated with each realization as well as directional sample variogram based on the collection of the *NMC* generated fields. Figure 7 depicts the generating semivariogram models together with the *NMC* semivariograms associated with (i) the generated fields and (ii) the sample semivariogram calculated along two mutually normal directions for a reference point located at the center of the simulation domain. Visual inspection of the results suggests that the generating (semi)variogram models are always fairly reproduced in an ensemble sense. Results of corresponding quality are obtained for other reference points in the system (not shown).

As an additional test, we repeat the same analysis by considering the trace-semivariogram of the field of (transformed) PSDs, defined in this setting as

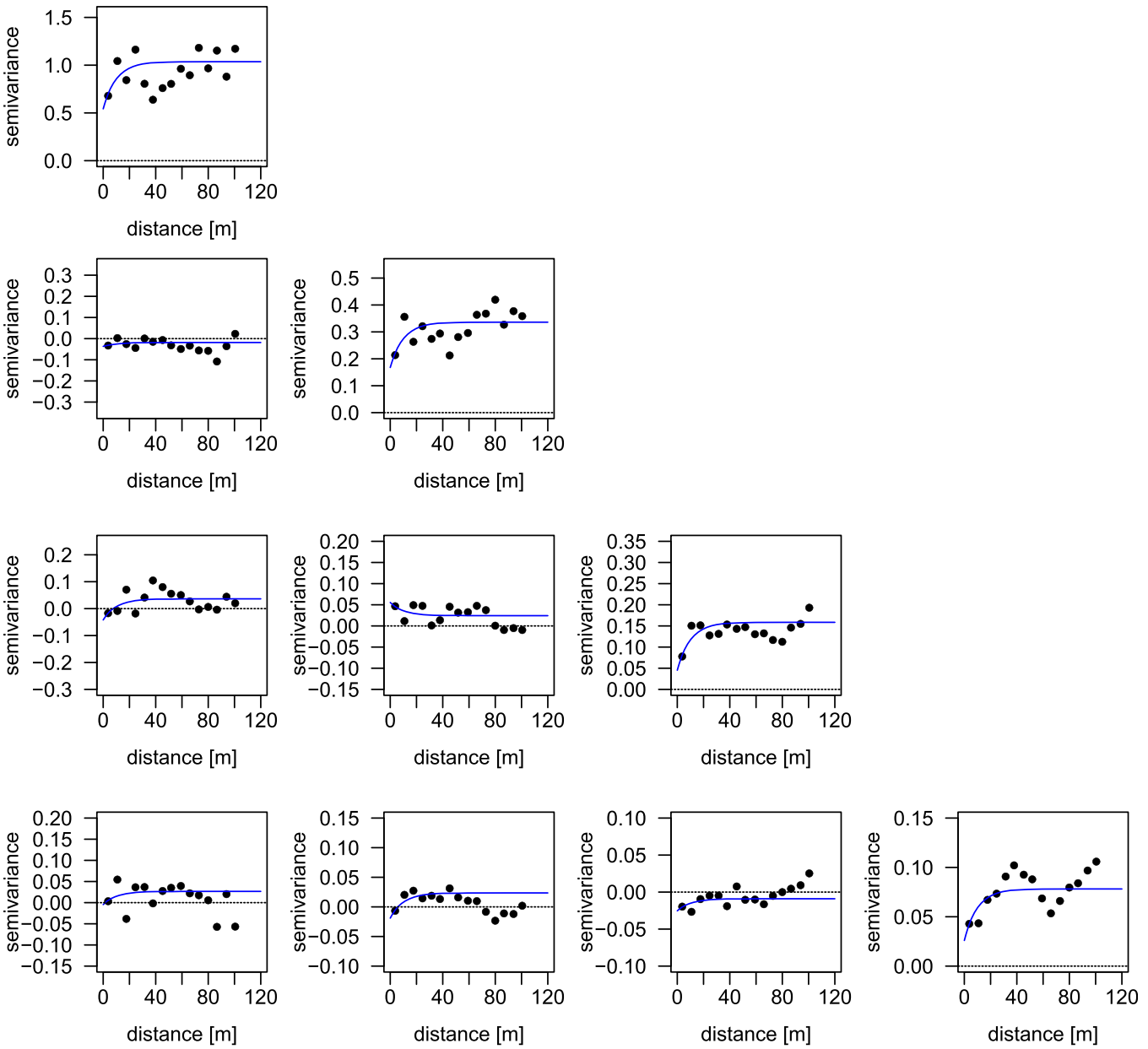

Figure 5. Omnidirectional semivariograms and cross-semivariograms estimated from the scores $\hat{\xi}_{s_1}, \ldots, \hat{\xi}_{s_n}$.

$$\gamma_{tr}(||\boldsymbol{s}_i - \boldsymbol{s}_j||) = \mathbb{E}\left[\int_{\mathcal{T}} (\mathcal{Z}_{\boldsymbol{s}_i}(\tau) - \mathcal{Z}_{\boldsymbol{s}_j}(\tau))^2 d\tau\right], \quad \boldsymbol{s}_i, \boldsymbol{s}_j \in D. \tag{11}$$

The trace-semivariogram is a global measure of spatial dependence undertaking, in the functional context, the same role as its finite-dimensional counterpart [see e.g., *Menafoglio et al.*, 2013, 2014, and references therein].

The quality of the results of this analysis depicted in Figure 8 further corroborates our conclusions, thus imbuing us with confidence about the potential of the generation method and results.

4.4. Conditional Simulation of Particle-Size Densities at the Lauswiesen Field Site

Here we illustrate an example of conditional simulation at the field site. For the purpose of our illustration, we consider a one-dimensional grid $D_1 \subset D$ of 250 points taken along borehole B5 at the site. Simulations are here performed conditional to the set of approximated PSDs obtained in section 4.1.

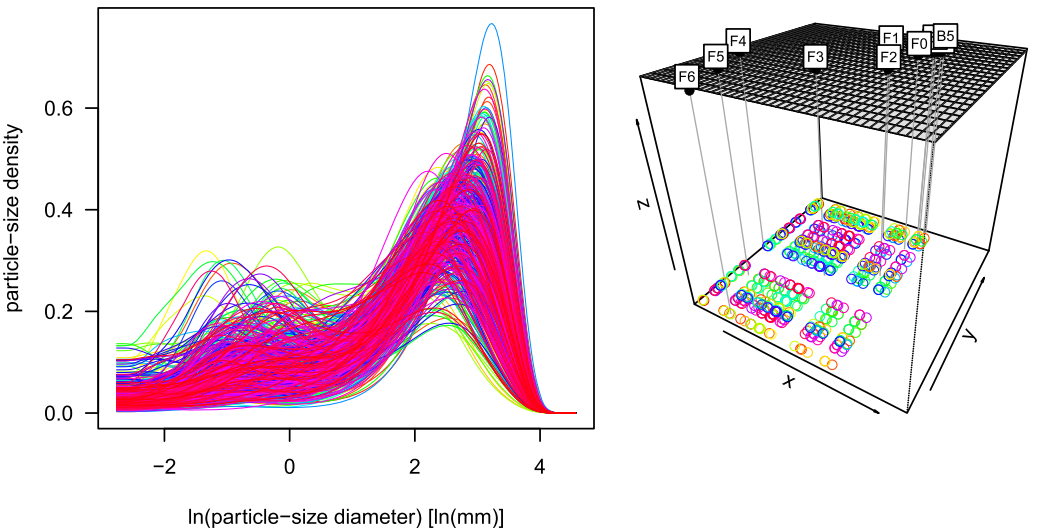

Figure 6. An example of unconditional realization: (left) spatially dependent PSDs and (right) the simulation grid.

Figure 9 depicts a selected realization on grid D_1, obtained by conditionally simulating the K-dimensional vectors of scores $\hat{\boldsymbol{\xi}}(\boldsymbol{s}_0)$, for $\boldsymbol{s}_0$ in D_1, according to the LMC of Figure 5. The CPU time for the simulation based on the R package gstat, within R version 3.0.2 took approximately 21′53″ (CPU time refers to an Intel® Core™ i7-3517U CPU @ 1.90 GHz). It can be noted that, by construction, the simulation interpolates the approximated clr-transform of PSDs (clr-PSDs for short), i.e., $\mathcal{Z}^K_{\boldsymbol{s}_1},\ldots,\mathcal{Z}^K_{\boldsymbol{s}_n}$, rather than the observed clr-PSDs $\mathcal{Z}_{\boldsymbol{s}_1},\ldots,\mathcal{Z}_{\boldsymbol{s}_n}$. We refer to Appendix C for a strategy to honor the original PSD data—i.e., those prior to EFPCA.

To assess the quality of the prediction, we perform 1000 simulations on the grid D_1. We notice that, for each $\boldsymbol{s}_0 \in D_1$, the ensemble average of the simulations at $\boldsymbol{s}_0$, i.e., $\sum_{j=1}^{1000}\mathcal{Z}^{(j)}_{\boldsymbol{s}_0}$, should approximate the conditional expectation $\mathbb{E}[\mathcal{Z}^K_{\boldsymbol{s}_0}|\mathcal{Z}^K_{\boldsymbol{s}_1},\ldots,\mathcal{Z}^K_{\boldsymbol{s}_n}]$, as simulations $\mathcal{Z}^{(j)}_{\boldsymbol{s}_0}$, $j=1,\ldots,1000$, are draws from the (approximated) conditional distribution of $\mathcal{Z}^K_{\boldsymbol{s}_0}$ given $\mathcal{Z}^K_{\boldsymbol{s}_1},\ldots,\mathcal{Z}^K_{\boldsymbol{s}_n}$. The conditional expectation $\mathbb{E}[\mathcal{Z}^K_{\boldsymbol{s}_0}|\mathcal{Z}^K_{\boldsymbol{s}_1},\ldots,\mathcal{Z}^K_{\boldsymbol{s}_n}]$ can be estimated from available (transformed) smoothed data $\mathcal{Z}^K_{\boldsymbol{s}_1},\ldots,\mathcal{Z}^K_{\boldsymbol{s}_n}$ as

$$\mathcal{Z}^{*K}_{\boldsymbol{s}_0}=\hat{m}+\sum_{k=1}^{K}\hat{\xi}^*_k(\boldsymbol{s}_0)\hat{w}_k \tag{12}$$

where $\hat{\boldsymbol{\xi}}^*(\boldsymbol{s}_0)=(\hat{\xi}^*_1(\boldsymbol{s}_0),\ldots,\hat{\xi}^*_K(\boldsymbol{s}_0))^T$ is the Simple Cokriging prediction of the score vector at $\boldsymbol{s}_0$, based on $\hat{\boldsymbol{\xi}}^*(\boldsymbol{s}_1),\ldots,\hat{\boldsymbol{\xi}}^*(\boldsymbol{s}_n)$ [see e.g., *Menafoglio and Petris*, 2016]. Note that the same argument can be formulated directly for the PSDs (i.e., in the Bayes space). Figures 10a and 10b display, in the space of densities, the ensemble average of the 1000 simulated clr-PSDs and the Kriging prediction based on the variography previously estimated, respectively. From the graphical inspection of Figures 10a and 10b, one can appreciate the high quality of our simulations. This is also confirmed by Figure 10c, which represents, for $J=1,\ldots,K$, the minimum, maximum and mean, over $\boldsymbol{s}_0 \in D_1$, of the squared distance $d(\boldsymbol{s}_0;J)^2=||\sum_{j=1}^{J}\mathcal{Z}^{(j)}_{\boldsymbol{s}_0}-\mathcal{Z}^{*K}_{\boldsymbol{s}_0}||^2$ of the partial ensemble averages $\sum_{j=1}^{J}\mathcal{Z}^{(j)}_{\boldsymbol{s}_0}$ from the Simple Kriging prediction $\mathcal{Z}^{*K}_{\boldsymbol{s}_0}$.

5. Conclusions and Further Research

The theoretical and application-oriented contributions of our work lead to the following key conclusions.

1. A novel strategy has been proposed to address the problem of stochastic simulation of particle-size curves (PSCs) and associated densities (PSDs). The latter constitute a set of (infinite-dimensional) functional data and treating them as functional compositions (i.e., as elements of the Bayes space) is a key feature of the procedure. Our theoretical framework enables us to (a) formulate a Gaussian model for the infinite-dimensional field of PSDs; (b) project the available data onto a truncated orthonormal basis to obtain a finite-dimensional approximation of the (otherwise infinite-dimensional) PSDs via a set of multivariate vectors of coefficients; and (c) perform either unconditional or conditional stochastic simulation, based on the multivariate random

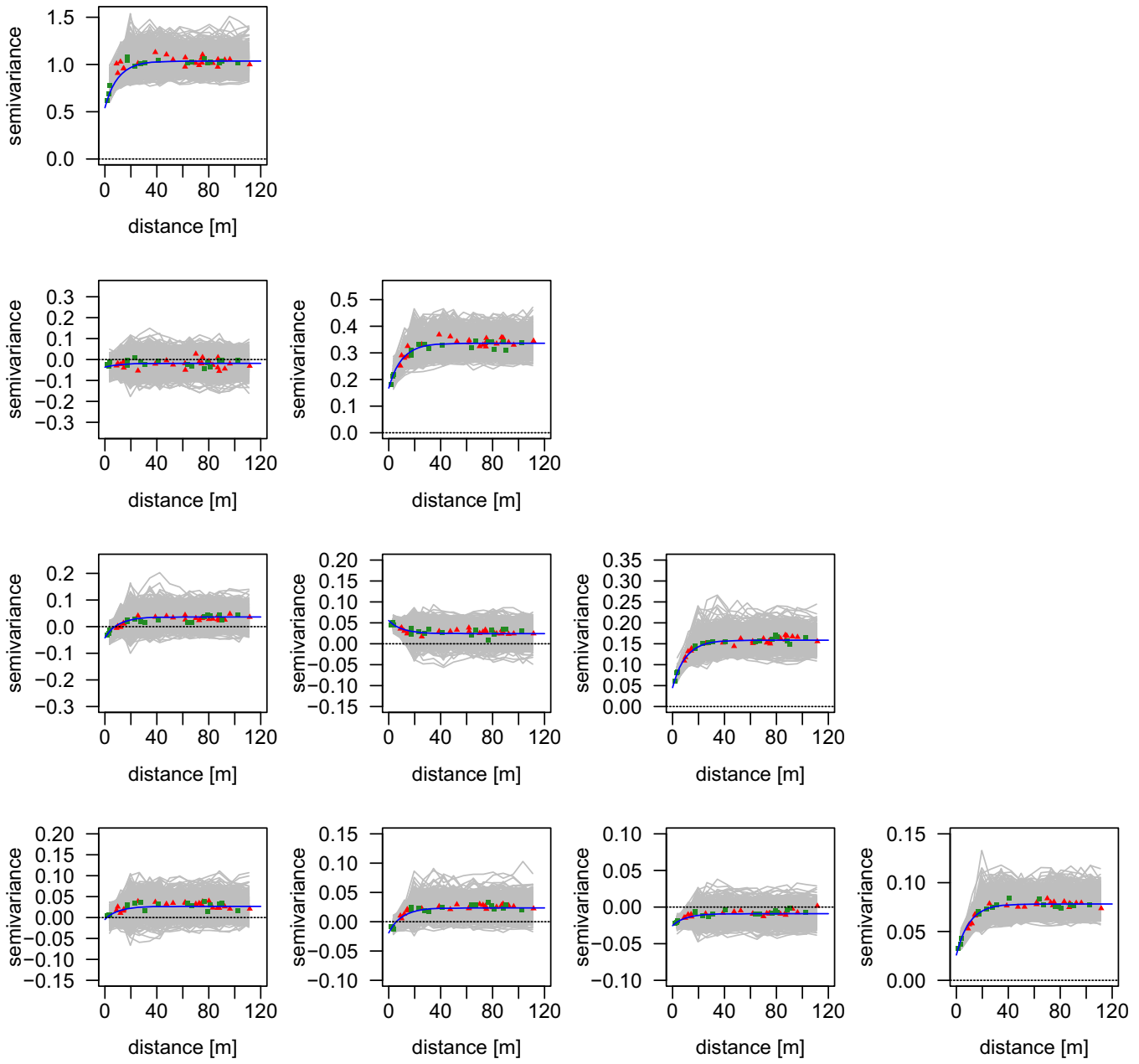

Figure 7. Generating LMC (blue lines), estimated omnidirectional semivariogram and cross-semivariogram in 1000 Monte Carlo simulations (grey curves), average over 1000 simulation of the semivariogram estimated at the central point in direction *x* (red symbols) and *y* (green symbols).

field of coefficients. The latter step can be addressed through the use of any of the available techniques for multivariate stochastic simulation (including, e.g., sequential Gaussian cosimulation).

2. We study the way one can set the dimension of the approximating problem and the functional basis onto which these types of functional data can be projected. Our results suggest that an optimal solution is provided upon relying on a simplicial functional principal component analysis (SFPCA). In this context, one may need to set the dimensionality of the approximated problem according to the available computational resources. As such, key challenges associated with future direct implementation of the approach to field-scale settings are related to improving the computational efficiency required for the simulation of the spatial field of coefficients, a step which still appears to be quite costly.
3. The stochastic simulation procedure has been demonstrated through an extensive Monte Carlo study based on a set of particle-size curves collected within a shallow alluvial heterogeneous aquifer system.

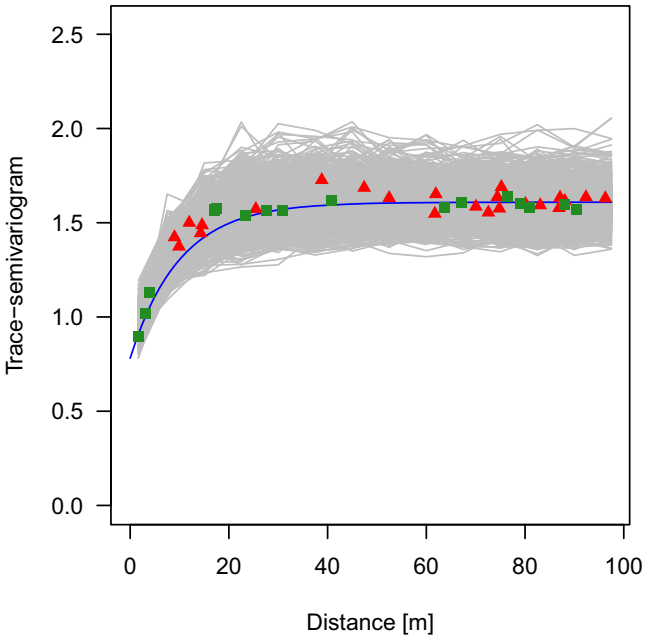

Figure 8. Generating model (blue curves), estimated (omnidirectional) trace-semivariograms in 1000 simulations (grey curves), average over the collection of 1000 simulation of the (omnidirectional) trace-semivariogram estimated at the central point in direction *x* (red symbols) and *y* (green symbols).

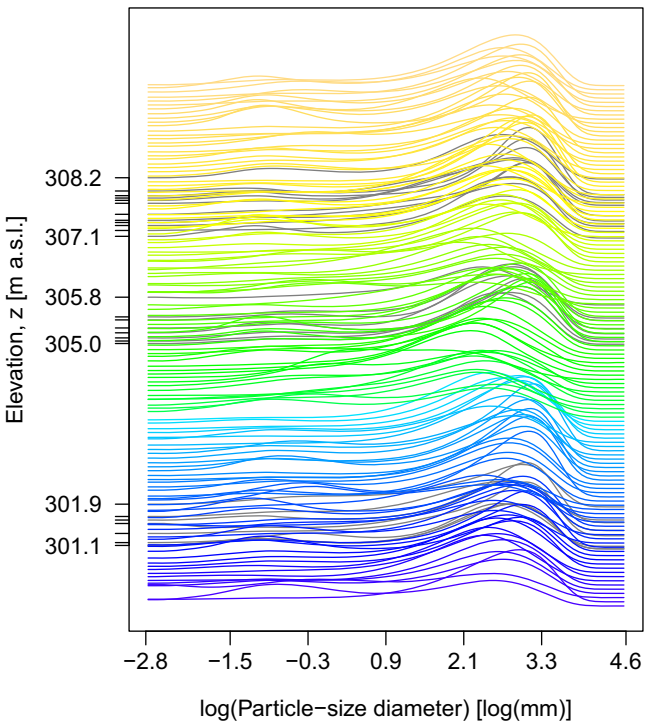

Figure 9. Conditional realization of PSDs at borehole B5 of the investigated field site. Vertical coordinates correspond to the sample/target locations. Elevation is given in meters above sea level (m asl). Simulated PSDs are plotted as colored curves, data as grey curves.

The quality of our results appear to be quite satisfactory in all tested scenarios. While we employ a stationary assumption for the purpose of our demonstration, it is possible to extend the technique to nonstationary settings of the kind arising, e.g., when an aquifer is conceptualized as a composite medium, where diverse nonoverlapping materials form its internal architecture. Work in this direction is currently under way [*Menafoglio et al.*, 2015]. With reference to practical applications, we note that, in contrast to common approaches relying solely on a few features of PSCs (e.g., selected quantiles), our approach yields collections of stochastic realizations of the spatial distribution of the entire PSC, thus contributing to a key improvement of one's ability to characterize the complete information content embedded in PSC data. In this sense, future extensions will consider embedding the approach in the context of site characterization procedures whereas PSCs carry information about the spatial distribution of geomaterials, as well indications on geochemical and hydraulic attributes of soil samples. Having at our disposal rigorous and efficient techniques to project estimates of PSCs on a computational grid via Kriging and/or generate a collection of stochastic realizations of PSCs can also assist inverse modeling of subsurface flow and chemical transport and/or improve the effectiveness of data assimilation techniques such as those based, e.g., on Ensemble Kalman Filter and its variants.

Appendix A: Density Functions as Elements of a Bayes Space

A proper (geo)statistical analysis and simulation of PSDs should account for the peculiar nature of this kind of constrained (compositional) data. The log-ratio approach for the statistical

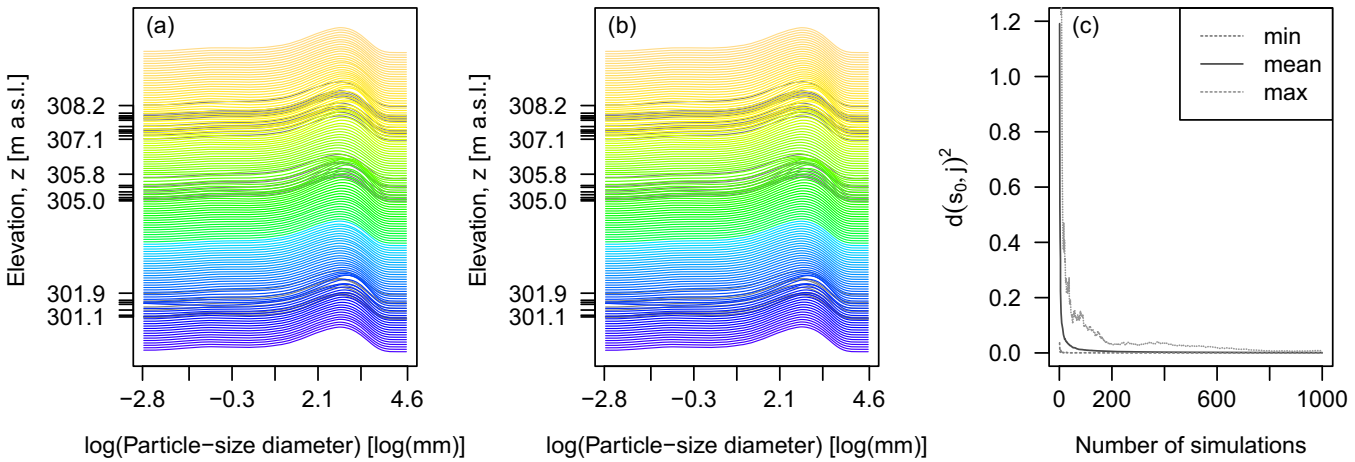

Figure 10. Assessment of the quality of conditional simulations at borehole B5 of Lauswiesen field site. (a) Average of 1000 conditional simulations of PSDs; (b) simple Kriging prediction of PSDs; and (c) squared distance between partial ensemble averages $\sum_{j=1}^{J} \mathcal{Z}_{\mathbf{s}_0}^{(j)}$ and Simple Kriging prediction $\mathcal{Z}_{\mathbf{s}_0}^{*K}$. In Figures 10a and 10b, vertical coordinates correspond to the sample/target locations. Elevation is given in m asl. Simulated PSDs are plotted as colored curves, data as grey curves.

analysis of multivariate compositions was pioneered by *Aitchison* [1986] and *Pawlowsky-Glahn and Egozcue* [2001] and is well established in the statistical literature. It is based on the key observation that constant-sum objects convey only relative information. Indeed, one can readily see that a component (or part) of a compositional vector does not provide information per se, but relative to the measure of the whole—i.e., the constant they sum up to—and to the remaining parts of the composition. The Aitchison geometry then yields a proper setting to perform the statistical analysis, by accounting for the data constraints via the log-ratio approach.

In this setting, density functions, such as PSDs, can be viewed as functional compositions (FCs), i.e., compositional vectors with infinitely many parts, that are constrained to be positive and to integrate to a constant. As such, they inherit the key properties of multivariate compositions. Recent works of *Egozcue et al*. [2006, 2013], *van den Boogaart et al*. [2010], and *van den Boogaart et al*. [2014] extend the Aitchison geometry to the infinite-dimensional setting through the theory of Bayes spaces, with the aim of providing the space of FCs with a geometrical structure consistent with the key properties of compositions and allowing for their statistical analysis. As in *Menafoglio et al*. [2014, 2015], the focus of this work is on continuous FCs defined on the closed interval $\mathcal{T}=[t_{\min}, t_{\max}]$. Two FCs f, g are considered equivalent if they are proportional, i.e., $f=c\cdot g$, for $c>0$. This equivalence relation reflects the so-called *scale invariance* property of FCs upon which the log-ratio approach is grounded: proportional FCs convey the same set of *relative* information, i.e., the measure of the whole is of no interest in a compositional analysis. Here we always consider as representative of an equivalence class of FCs its element integrating to 1.

Call $A^2(\mathcal{T})$ (or A^2 for short) the space of (equivalence classes of) FCs on $\mathcal{T}$, whose logarithms are squared integrable, i.e.,

$$A^2=\left\{f:\mathcal{T}\to(0,+\infty),\ \int_{\mathcal{T}}\log^2(f(\tau))d\tau<+\infty\right\}. \tag{A1}$$

The space A^2 can be equipped with the operations of *perturbation* $\oplus$ and *powering* $\odot$ [*Egozcue et al*., 2006; *van den Boogaart et al*., 2014]

$$f\oplus g=\mathcal{C}(fg);\quad \alpha\odot f=\mathcal{C}(f^\alpha),\quad f,g\in A^2,\ \alpha\in\mathbb{R}, \tag{A2}$$

where $\mathcal{C}(f)=\int_{\mathcal{T}}f$ is the called *closure operation*, and maps a FC in the representative of its equivalence class that integrates to 1. The neutral elements of perturbation and powering (i.e., those playing the role of 0 in the sum and 1 in the product) are $0_\oplus\equiv 1/|\mathcal{T}|$, with $|\mathcal{T}|$ the length of $\mathcal{T}$, and 1, respectively. *Egozcue et al*. [2006] prove that $(A^2,\oplus,\odot)$ is a vector space, perturbation and powering playing the role of sum and product by a constant, respectively. In this setting, we denote by $f\ominus g$ the difference, in the geometry of A^2, between f and g, namely the perturbation of f with the reciprocal of g, i.e., $f\ominus g=\mathcal{C}[f\oplus 1/g]$, for f, g in A^2.

To endow A^2 with a Hilbert space structure, *Egozcue et al.* [2006] equip the vector space $(A^2, \oplus, \odot)$ with the inner product

$$\langle f, g\rangle_{A^2} = \int_{\mathcal{T}} [\log(f(\tau))\log(g(\tau))] - \frac{1}{|\mathcal{T}|}\int_{\mathcal{T}} \log(f(\tau))d\tau \int_{\mathcal{T}} \log(g(\tau))d\tau, \quad f, g \in A^2. \tag{A3}$$

Egozcue et al. [2006] prove that $(A^2, \oplus, \odot, \langle\cdot,\cdot\rangle_{A^2})$ is a Hilbert space, which is called *Bayes* (*Hilbert*) *space*.

The clr-transformation defined in (1) provides an isometric isomorphism (i.e., a bijective relation preserving distances) between the space $A^2(\mathcal{T})$ and the space $L^2(\mathcal{T})$ of (equivalence classes of) square-integrable functions on $\mathcal{T}$. From the computational viewpoint, the use of clr-transforms is convenient, as it allows mapping problems in A^2 into problems in L^2, where most methods of FDA can be applied. Further, one has

$$\mathrm{clr}(f \oplus g) = \mathrm{clr}(f) + \mathrm{clr}(g), \quad \mathrm{clr}(\alpha \odot f) = \alpha \cdot \mathrm{clr}(f), \quad \langle f, g\rangle_{A^2} = \langle \mathrm{clr}(f), \mathrm{clr}(g)\rangle_{L^2}. \tag{A4}$$

The Hilbert space geometry of the space $(A^2, \oplus, \odot, \langle\cdot,\cdot\rangle_{A^2})$ together with the properties of the clr-transformation allows formulating the method devised in section 3 equivalently in the Bayes (Hilbert) space geometry.

Appendix B: A Mathematical Framework in Bayes Spaces for the Stochastic Simulation of PSDs

For any $\boldsymbol{s}$ in D, we denote by $\mu_{\boldsymbol{s}}$ the Fréchet mean of $\mathcal{Y}_{\boldsymbol{s}}$, i.e. [*Fréchet*, 1948],

$$\mu_{\boldsymbol{s}} = \mathbb{E}[\mathcal{Y}_{\boldsymbol{s}}] = \underset{\mathcal{Y}\in A^2(\mathcal{T})}{\arg\inf}\ \mathbb{E}[||\mathcal{Y}_{\boldsymbol{s}} \ominus \mathcal{Y}||^2_{A^2}]. \tag{B1}$$

We assume $\{\mathcal{Y}_{\boldsymbol{s}}, \boldsymbol{s} \in D\}$ to be a stationary Gaussian random field in A^2 [*Bogachev*, 1998; *Bosq*, 2000]. This implies that the mean function $\mu_{\boldsymbol{s}} = \mu$ is spatially constant. We indicate with $\mathcal{C}$ the covariance function of the field $\{\mathcal{Y}_{\boldsymbol{s}}, \boldsymbol{s} \in D\}$, that maps any pair of locations $\boldsymbol{s}_1, \boldsymbol{s}_2$ in D into the cross-covariance operator $\mathcal{C}(\boldsymbol{s}_1 - \boldsymbol{s}_2)$ between the elements of the field at such locations, i.e.,

$$\mathcal{C}(\boldsymbol{s}_1 - \boldsymbol{s}_2)x = \mathbb{E}[\langle \mathcal{Y}_{\boldsymbol{s}_1} \ominus \mu, x\rangle_{A^2} \odot (\mathcal{Y}_{\boldsymbol{s}_2} \ominus \mu)], \quad x \in A^2. \tag{B2}$$

We consider for each $\boldsymbol{s} \in D$ the expansion

$$\mathcal{Y}_{\boldsymbol{s}} = \mu \oplus \bigoplus_{k=1}^{+\infty} \xi_k(\boldsymbol{s}) \odot u_k, \tag{B3}$$

where $\{u_k, k \geq 1\}$ is a given orthonormal basis of A^2, and $\xi_k(\boldsymbol{s}) = \langle \mathcal{Y}_{\boldsymbol{s}} \ominus \mu, u_k\rangle_{A^2}$. The basis $\{u_k, k \geq 1\}$ and the expansion (B3) are well defined by virtue of the Hilbert space structure of the space A^2. Note that random coefficients $\xi_k(\boldsymbol{s})$, $k = 1, \ldots, K$ coincide with those in (4), provided that $v_k = clr(u_k)$, due to the properties of the clr-transformation.

If we could jointly simulate random realizations of all the real (random) coefficients $\{\xi_k(\boldsymbol{s}_0)\}_{k\geq 1}$, we would obtain a random realization of $\mathcal{Y}_{\boldsymbol{s}_0}$ through (B3). However, this is practically unaffordable because the effort required to simulate a multivariate random field increases with its dimensionality. We circumvent this issue by considering, for $\boldsymbol{s}$ in D, the sequence of truncated expansions (equivalent to (4))

$$\mathcal{Y}^K_{\boldsymbol{s}} = \mu \oplus \bigoplus_{k=1}^{K} \xi_k(\boldsymbol{s}) \odot u_k, \quad K \geq 1. \tag{B4}$$

The element $\mathcal{Y}^K_{\boldsymbol{s}}$ associated with a truncation order K yields an approximation of $\mathcal{Y}_{\boldsymbol{s}}$ such that

$$\mathbb{E}[||\mathcal{Y}^K_{\boldsymbol{s}} \ominus \mathcal{Y}_{\boldsymbol{s}}||^2_{A^2}] = \sum_{k=K+1}^{+\infty} \mathbb{E}[|\xi_k(\boldsymbol{s})|^2] = \sum_{k=K+1}^{+\infty} \langle \mathcal{C}(\boldsymbol{0})\mathrm{u_k}, \mathrm{u_k}\rangle, \tag{B5}$$

which approaches 0 as K increases to infinity. Note that the term at the right hand side of (B5) does not depend on the spatial index $\boldsymbol{s}$ in D. Thus, for any given tolerance, one can determine a truncation order K such that $\mathcal{Y}^K_{\boldsymbol{s}}$ approximates $\mathcal{Y}_{\boldsymbol{s}}$ (in the mean square sense) with a desired precision, uniformly in D.

Given a truncation order K, a random field $\{\mathcal{Y}^K_{\boldsymbol{s}}, \boldsymbol{s} \in D\}$ in A^2 whose elements are given by (B4) can be defined. The distributional properties of such a field are determined by μ and by those of the zero-mean multivariate random field $\{\xi(\boldsymbol{s}), \boldsymbol{s} \in D\}$, $\xi(\boldsymbol{s})$ indicating the K-dimensional coefficient vector of the basis expansion (B4) in $\boldsymbol{s}$, i.e.,

$\boldsymbol{\xi}(\boldsymbol{s})=(\xi_1(\boldsymbol{s}),\ldots,\xi_K(\boldsymbol{s}))^T$. Note that both $\mathcal{Y}_{\boldsymbol{s}}^K$ and $\boldsymbol{\xi}(\boldsymbol{s})$ are Gaussian random fields (in A^2 and $\mathbb{R}^K$, respectively) by virtue of the Gaussian assumption on the field $\{\mathcal{Y}_{\boldsymbol{s}}, \boldsymbol{s}\in D\}$. Additionally, the element $\mathcal{Y}_{\boldsymbol{s}}^K$ has mean $\mu_{\boldsymbol{s}}^K=\mu$ by virtue of (B4), and the following matrix representation of the covariance function $\mathcal{C}^K$ of the field $\{\mathcal{Y}_{\boldsymbol{s}}^K, \boldsymbol{s}\in D\}$ holds

$$\mathcal{C}^K(\boldsymbol{h})x=\bigoplus_{j=1}^{K}\bigoplus_{k=1}^{K}(\mathbb{C}_{jk}x_j)\odot u_k, \tag{B6}$$

where $x_j=\langle x,u_j\rangle_{A^2}$ and $\mathbb{C}_{jk}=\langle\mathcal{C}(\boldsymbol{h})u_j,u_k\rangle_{A^2}=\mathbb{E}[\xi_j(\boldsymbol{s})\xi_k(\boldsymbol{s})]$.

The quality of a Kth order approximation of the kind (B4) varies according to the basis $\{u_k, k\geq 1\}$ employed. Given $K\geq 1$, the mean square error of approximating $\mathcal{Y}_{\boldsymbol{s}}$ through the projection (B4) over the first K elements of the basis $\{u_k, k\geq 1\}$ is bounded below by [see e.g., *Horváth and Kokoszka*, 2012, Theorem 3.2]

$$\mathbb{E}[||\mathcal{Y}_{\boldsymbol{s}}^K\ominus\mathcal{Y}_{\boldsymbol{s}}||_{A^2}^2]\geq\sum_{k=K+1}^{+\infty}\lambda_k, \tag{B7}$$

where (λ_k, e_k), $k\geq 1$, represent the eigenpairs of $\mathcal{C}(\boldsymbol{0})$, with eigenvalues ordered in decreasing order $\lambda_1\geq\lambda_2\geq\ldots$. Given K, the basis should be chosen as to attain a mean square error of approximation as close as possible to the lower bound (B7). It can be proved [e.g., *Horváth and Kokoszka*, 2012, Theorem 3.2] that the bound in (B7) is reached when considering $u_1,\ldots,u_K$ to be precisely the set of the first K eigenfunctions of $\mathcal{C}(\boldsymbol{0})$, i.e., $e_1,\ldots,e_K$.

If the zero-lag covariance operator is not known a priori, one can apply the simplicial functional principal component analysis (SFPCA) of *Hron et al.* [2016] to (a) estimate from available data the zero-lag covariance operator $\mathcal{C}(\boldsymbol{0})$ through the empirical estimator

$$\mathcal{S}x=\frac{1}{n}\bigoplus_{i=1}^{n}\langle\mathcal{Y}_{\boldsymbol{s}_i}\ominus\hat{\mu},x\rangle_{A^2}(\mathcal{Y}_{\boldsymbol{s}_i}\ominus\hat{\mu}),\quad x\in A^2, \tag{B8}$$

$\hat{\mu}=\frac{1}{n}\oplus_{i=1}^{n}\mathcal{Y}_{\boldsymbol{s}_i}$ denoting the sample mean, (b) compute the eigen-pairs $(\hat{\lambda}_k,\hat{\boldsymbol{e}}_k)$, $k=1,\ldots,n-1$, of this estimate, and (c) project the observations on the first K eigenfunctions (or simplicial functional principal components, SFPCs) of $\mathcal{S}$ to obtain the representation

$$\mathcal{Y}_{\boldsymbol{s}_i}\approx\hat{\mu}\oplus\bigoplus_{k=1}^{K}\hat{\xi}_k(\boldsymbol{s}_i)\odot\hat{\boldsymbol{e}}_k, \tag{B9}$$

which is the equivalent, in the Bayes space A^2, of (7). The computation of the SFPCs and expansion (B9) can rely on the centered log-ratio (clr) transformation, as shown in section 3 and Appendix A. Note that the basis coefficients $\hat{\xi}_k(\boldsymbol{s})$ appearing in (B9) coincide with those in (7), as $\langle\mathcal{Y}_{\boldsymbol{s}}^K\ominus\hat{\mu},\hat{\boldsymbol{e}}_k\rangle_{A^2}=\langle\mathcal{Z}_{\boldsymbol{s}}^K-\hat{m},\hat{w}_k\rangle_{L^2}$, i.e., the scores computed in L^2 are the same as those in A^2.

Appendix C: Reproducing the Observations in Conditional Simulations

By construction, the conditional simulations obtained through the projection strategy of section 3 are based on the approximated PSDs $\mathcal{Y}_{\boldsymbol{s}_1}^K,\ldots,\mathcal{Y}_{\boldsymbol{s}_n}^K$, $\mathcal{Y}_{\boldsymbol{s}_i}^K=clr^{-1}(\mathcal{Z}_{\boldsymbol{s}_i})$, rather than the observed PSDs $\mathcal{Y}_{\boldsymbol{s}_1},\ldots,\mathcal{Y}_{\boldsymbol{s}_n}$. Here we illustrate a strategy to obtain simulations that honor the actual observations at locations where these are collected.

We call $\mathcal{Y}_{\boldsymbol{s}_0}^K$ the simulated PSD at a target location $\boldsymbol{s}_0\in D$, and denote by $\epsilon_{\boldsymbol{s}_i}^K=\mathcal{Y}_{\boldsymbol{s}_i}\ominus\mathcal{Y}_{\boldsymbol{s}_i}^K$, $i=1,\ldots,n$, the residuals of SFPCA. These residuals are neglected when analyzing and simulating PSDs via approximation (B9) (or (7)). One can embed these in the (conditional) simulation procedure by interpolating them through an appropriate notion of Kriging, and then sum the result to the simulated realization $\mathcal{Y}_{\boldsymbol{s}_0}^K$. In this appendix, we introduce an extension of the method illustrated in section 3 and Appendix B, by using the notation of the Bayes (Hilbert) space A^2. Note that one could work in L^2, by replacing the operations in A^2 $(\oplus,\odot)$, with those in L^2 $(+,\cdot)$, and working with the $\mathcal{Z}$ instead of the $\mathcal{Y}$ variables.

Menafoglio et al. [2014] introduce the notion of Functional Compositional Kriging (FCK) that allows obtaining the best linear unbiased prediction in the sense of linear combination of the data in A^2. We call $\epsilon_{\boldsymbol{s}_0}^{*K}$ the FCK prediction of the residual at $\boldsymbol{s}_0$. This prediction is obtained as the linear combination $\epsilon_{\boldsymbol{s}_0}^{*K}=\oplus_{i=1}^{n}\vartheta_i^*\odot\epsilon_{\boldsymbol{s}_i}^K$ of the residuals $\epsilon_{\boldsymbol{s}_i}^K$, $i=1,\ldots,n$, whose weights minimize the prediction mean square error (MSE). Note that no unbiasedness constraint needs to be imposed, as the residuals $\epsilon_{\boldsymbol{s}_i}^K$ are zero mean by construction. Taking

advantage of the work of *Menafoglio et al.* [2013, 2014], it is possible to show that minimization of the MSE is tantamount to solving the FCK system

$$\Gamma^{\epsilon}\vartheta = \gamma_0^{\epsilon}, \tag{C1}$$

where $\Gamma^{\epsilon}_{i,j} = \mathbb{E}[||\epsilon^K_{\boldsymbol{s}_i} \ominus \epsilon^K_{\boldsymbol{s}_j}||^2_{A^2}]$, $i,j = 1,\dots,n$, $\vartheta = (\vartheta_1,\dots,\vartheta_n)^T \in \mathbb{R}^n$, $(\gamma_0)_i = \mathbb{E}[||\epsilon^K_{\boldsymbol{s}_i} \ominus \epsilon^K_{\boldsymbol{s}_0}||^2_{A^2}]$, $i = 1,\dots,n$. Note that (C1) is a Simple Kriging system, consistent with the observation that residuals are zero mean.

Having computed the prediction $\epsilon^{*K}_{\boldsymbol{s}_0}$, one can finally obtain the desired simulation as $\mathcal{Y}^K_{\boldsymbol{s}_0} \oplus \epsilon^{*K}_{\boldsymbol{s}_0}$.

Acknowledgments
Funding from the European Union's Horizon 2020 Research and Innovation program (Project "Furthering the knowledge Base for Reducing the Environmental Footprint of Shale Gas Development" FRACRISK—grant agreement 640979) is acknowledged. All data used in the paper will be retained by the authors for at least 5 years after publication and will be available to the readers upon request.

References

Abrahamsen, P., and F. Benth (2001), Kriging with inequality constraints, *Math. Geol.*, *33*(6), 719–744, doi:10.1023/A:1011078716252.
Aitchison, J. (1986), *The Statistical Analysis of Compositional Data*, Chapman and Hall, London, U. K.
Barahona-Palomo, M., M. Riva, X. Sánchez-Vila, E. Vázquez-Sune, and A. Guadagnini (2011), Quantitative comparison of impeller flowmeter and particle-size distribution techniques for the characterization of hydraulic conductivity variability, *Hydrogeol. J.*, *19*(3), 603–661.
Bianchi, M., C. Zheng, C. Wilson, G. R. Tick, G. Liu, and S. M. Gorelick (2011), Spatial connectivity in a highly heterogeneous aquifer: From cores to preferential flow paths, *Water Resour. Res.*, *47*, W05524, doi:10.1029/2009WR008966.
Bogachev, V. (1998), *Gaussian Measures*, Am. Math. Soc., USA.
Bosq, D. (2000), *Linear Processes in Function Spaces*, Springer, N. Y.
Chilès, J. P., and P. Delfiner (1999), *Geostatistics: Modeling Spatial Uncertainty*, John Wiley, N. Y.
Das, B. S., N. W. Haws, and P. S. C. Rao (2005), Defining geometric similarity in soils, *Vadose Zone J.*, *4*(2), 264–270.
de Marsily, G. (1986), *Quantitative Hydrogeology*, Academic, N. Y.
Deutsch, C. V., and A. G. Journel (1998), *GSLIB Geostatistical Software Library and User's Guide*, Oxford Univ. Press, N. Y.
Egozcue, J., V. Pawlowsky-Glahn, R. Tolosana-Delgado, M. Ortego, and K. van den Boogaart (2013), Bayes spaces: Use of improper distributions and exponential families, *Rev. Real Acad. Cienc. Exactas Fis. Nat. Ser. A. Mat.*, *107*(2), 475–486.
Egozcue, J. J., J. L. Díaz-Barrero, and V. Pawlowsky-Glahn (2006), Hilbert space of probability density functions based on Aitchison geometry, *Acta Math. Sinica English Ser.*, *22*(4), 1175–1182.
Fréchet, M. (1948), Les éléments Aléatoires de Nature Quelconque dans une Espace Distancié, *Ann. L'Inst. Henri Poincaré*, *10*(4), 215–308.
Horváth, L., and P. Kokoszka (2012), *Inference for Functional Data With Applications*, Springer Ser. Stat., Springer, N. Y.
Hron, K., A. Menafoglio, M. Templ, K. Hruzova, and P. Filzmoser (2016), Simplicial principal component analysis for density functions in Bayes spaces, *Comput. Stat. Data Anal.*, *94*, 330–350.
Hu, B. X., M. M. Meerschaert, W. Barrash, D. W. Hyndman, C. He, X. Li, and L. Guo (2009), Examining the influence of heterogeneous porosity fields on conservative solute transport, *J. Contam. Hydrol.*, *108*(3–4), 77–88.
Mariethoz, G., and J. Caers (2015), *Multiple-Point Geostatistics: Stochastic Modeling With Training Images*, John Wiley, Hoboken, N. J.
Martìn, M. A., J. M. Rey, and F. J. Taguas (2005), An entropy-based heterogeneity index for mass-size distributions in earth science, *Ecol. Modell.*, *182*, 221–228.
Menafoglio, A., and G. Petris (2016), Kriging for Hilbert-space valued random fields: The operatorial point of view, *J. Multivariate Anal.*, *146*, 84–94.
Menafoglio, A., P. Secchi, and M. Dalla Rosa (2013), A Universal Kriging predictor for spatially dependent functional data of a Hilbert Space, *Electron. J. Stat.*, *7*, 2209–2240.
Menafoglio, A., A. Guadagnini, and P. Secchi (2014), A Kriging approach based on Aitchison geometry for the characterization of particle-size curves in heterogeneous aquifers, *Stochastic Environ. Res. Risk Assess.*, *28*(7), 1835–1851.
Menafoglio, A., P. Secchi, and A. Guadagnini (2015), A Class-Kriging predictor for Functional Compositions with application to particle-size curves in heterogeneous aquifers, *Math. Geosci.*, *48*, 463–485, doi:10.1007/s11004-015-9625-7.
Miller, E., and R. Miller (1956), Physical theory for capillary flow phenomena, *J. Appl. Phys.*, *27*(4), 324–332.
Nasta, P., N. Romano, S. Assouline, J. A. Vrugt, and J. W. Hopmans (2013), Prediction of spatially variable unsaturated hydraulic conductivity using scaled particle-size distribution functions, *Water Resour. Res.*, *49*, 4219–4229, doi:10.1002/wrcr.20255.
Neuman, S. P., A. Blattstein, M. Riva, D. M. Tartakovsky, A. Guadagnini, and T. Ptak (2007), Type curve interpretation of late-time pumping test data in randomly heterogeneous aquifers, *Water Resour. Res.*, *43*, W10421, doi:10.1029/2007WR005871.
Pachepsky, Y., W. Rawls, and H. Lin (2006), Hydropedology and pedotransfer functions, *Geoderma*, *131*(3), 308–316.
Panzeri, M., M. Riva, A. Guadagnini, and S. Neuman (2015), Enkf coupled with groundwater flow moment equations applied to Lauswiesen aquifer, Germany, *J. Hydrol.*, *521*, 205–216, doi:10.1016/j.jhydrol.2014.11.057.
Pawlowsky-Glahn, V., and J. J. Egozcue (2001), Geometric approach to statistical analysis in the symplex, *Stochastic Environ. Res. Risk Assess.*, *15*, 384–398.
Pawlowsky-Glahn, V., J. J. Egozcue, and R. Tolosana-Delgado (2015), *Modeling and Analysis of Compositional Data, Statistics in Practice*, John Wiley, Hoboken, N. J.
Pebesma, E. J. (2004), Multivariable geostatistics in S: The gstat package, *Comput. Geosci.*, *30*, 683–691.
R Core Team (2013), *R: A Language and Environment for Statistical Computing*, R Found. for Stat. Comput., Vienna.
Ramsay, J., and B. Silverman (2005), *Functional Data Analysis*, 2nd ed., Springer, N. Y.
Rawls, W., D. Brakensiek, and K. Saxton (1982), Estimation of soil water properties, *Trans. ASAE*, *28*(5), 1316–1320.
Remy, N., A. Boucher, and J. Wu (2009), *Applied Geostatistics With SGeMS: A User's Guide*, Cambridge Univ. Press, Cambridge, U. K.
Riva, M., L. Guadagnini, A. Guadagnini, T. Ptak, and E. Martac (2006), Probabilistic study of well capture zones distributions at the Lauswiesen field site, *J. Contam. Hydrol.*, *88*, 92–118.
Riva, M., A. Guadagnini, D. Fernández-Garcáa, X. Sánchez-Vila, and T. Ptak (2008), Relative importance of geostatistical and transport models in describing heavily tailed breakthrough curves at the Lauswiesen site, *J. Contam. Hydrol.*, *101*, 1–13.
Riva, M., L. Guadagnini, and A. Guadagnini (2010), Effects of uncertainty of lithofacies, conductivity and porosity distributions on stochastic interpretations of a field scale tracer test, *Stochastic Environ. Res. Risk Assess.*, *24*, 955–970.
Riva, M., X. Sanchez-Vila, and A. Guadagnini (2014), Estimation of spatial covariance of log-conductivity from particle-size data, *Water Resour. Res.*, *50*, 5298–5308, doi:10.1002/2014WR015566.

Rogiers, B., D. Mallants, O. Batelaan, M. Gedeon, M. Huysmans, and A. Dassargues (2012), Estimation of hydraulic conductivity and its uncertainty from grain-size data using glue and artificial neural networks, *Math. Geosci.*, *44*(6), 739–763.

Rosas, J., O. Lopez, T. M. Missimer, K. M. Coulibaly, A. H. A. Dehwah, K. Sesler, L. R. Lujan, and D. Mantilla (2014), Determination of hydraulic conductivity from grain-size distribution for different depositional environments, *Ground Water*, *52*(3), 399–413.

Schaap, M. G. (2013), Description, analysis, and interpretation of an infiltration experiment in a semiarid deep vadose zone, in *Advances in Hydrogeology*, pp. 159–183, Springer, N. Y.

Schaap, M. G., F. J. Leij, and M. T. van Genuchten (2001), Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions, *J. Hydrol.*, *251*(3–4), 163–176.

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

Tuli, A., K. Kosugi, and J. Hopmans (2001), Simultaneous scaling of soil water retention and unsaturated hydraulic conductivity functions assuming lognormal pore-size distribution, *Adv. Water Resour.*, *24*(6), 677–688.

van den Boogaart, K., J. J. Egozcue, and V. Pawlowsky-Glahn (2010), Bayes linear spaces, *SORT*, *34*(2), 201–222.

van den Boogaart, K. G., J. Egozcue, and V. Pawlowsky-Glahn (2014), Bayes Hilbert spaces, *Aust. N. Z. J. Stat.*, *56*, 171–194.

Vienken, T., and P. Dietrich (2011), Field evaluation of methods for determining hydraulic conductivity from grain size data, *J. Hydrol.*, *400*(1–2), 58–71.

Vogel, T., M. Cislerova, and J. Hopmans (1991), Porous media with linearly variable hydraulic properties, *Water Resour. Res.*, *27*(10), 2735–2741.

Vukovic, M., and A. Soro (1992), *Determination of Hydraulic Conductivity of Porous Media From Grain-Size Composition*, Water Resour. Publ., Littleton, Colo.