Year : 2018

Integrating complex and diverse spatial datasets : applications to hydrogeophysics

Nussbaumer Raphaël

Nussbaumer Raphaël, 2018, Integrating complex and diverse spatial datasets : applications to hydrogeophysics

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive http://serval.unil.ch
Document URN : urn:nbn:ch:serval-BIB_F2878A4FBD5C3

UNIL | Université de Lausanne

Faculté des géosciences et de l'environnement
Institut des sciences de la Terre

Integrating complex and diverse spatial datasets: Applications to hydrogeophysics

Thèse de doctorat

Présentée à la
Faculté des géosciences et de l'environnement
Institut des sciences de la Terre
de l'Université de Lausanne
par

Raphaël Nussbaumer

Master of Science in Hydrology and Water Resources Management
Imperial College London

Jury
Prof. , Président du jury
Prof. Dr. Klaus Holliger, Directeur de thèse
Prof. Dr. Erwan Gloaguen, Co-Directeur de thèse
Prof. Dr. Grégoire Mariéthoz, Expert
Prof. Dr. Philippe Renard, Expert

Lausanne, 2018

UNIL | Université de Lausanne
Faculté des géosciences et de l'environnement
bâtiment Geopolis bureau 4631

IMPRIMATUR

Vu le rapport présenté par le jury d'examen, composé de

Président de la séance publique :	M. le Professeur Christian Kull
Président du colloque :	M. le Professeur Christian Kull
Directeur de thèse :	M. le Professeur Klaus Holliger
Co-directeur de thèse :	M. le Professeur Erwan Gloaguen
Expert interne :	M. le Professeur Grégoire Mariéthoz
Expert externe :	M. le Professeur Philippe Renard

Le Doyen de la Faculté des géosciences et de l'environnement autorise l'impression de la thèse de

Monsieur Raphaël NUSSBAUMER

Titulaire d'un
Master of Science (Hydrology & Water Resource Management)
de l'Imperial College London

intitulée

Integrating complex and diverse spatial datasets: applications to hydrogeophysics

Lausanne, le 16 novembre 2018

Pour le Doyen de la Faculté des géosciences et de l'environnement

Professeur Christian Kull

Contents

List of Figures

List of Tables

Résumé

Bien que vitales pour l'humanité, les eaux souterraines sont menacées à travers le monde par la détérioration anthropique. Leur gestion durable, cruciale pour préserver la ressource en eau, repose en partie sur des modèles numériques permettant de prévoir l'état de l'eau dans les aquifères, c'est-à-dire des roches contenant les eaux souterraines. Une étape clé dans la construction de tels modèles consiste à déterminer les paramètres hydrogéologiques décrivant le comportement de l'eau dans l'aquifère. Au cours des dernières années, l'application croissante de techniques géophysiques à la caractérisation des aquifères a permis une meilleure description de leurs paramètres hydrogéologiques. Cependant, l'intégration de telles données reste une tâche difficile à cause (1) des différences de couverture et de résolution entre les types de mesures et (2), de l'interaction complexe et non unique des paramètres en question. Dans ce contexte, la géostatistique fournit un cadre utile permettant l'intégration de ces données spatiales. Cette thèse adopte une perspective plus large, qui peut être résumée comme suit : «développer des méthodologies efficaces et précises pour l'intégration de données spatiales, variables en termes de couverture et de résolution, et liées par des relations complexes non-uniques et dépendant du site d'étude. Les contributions principales de ce travail de thèse peuvent être divisées en trois étapes, brièvement résumées ci-dessous.

L'étape initiale concerne la première partie de l'énoncé du problème et adopte une approche géostatistique générale et théorique. Elle consiste à améliorer l'efficacité et la précision d'une méthode géostatistique largement utilisée pour générer des champs gaussiens : la simulation gaussienne séquentielle (SGS). Le but de cet algorithme est de remplir une grille en visitant consécutivement chaque nœud et en échantillonnant une valeur dans une distribution conditionnelle locale. Dans un premier temps, nous examinons l'impact du type de chemin utilisé pour réaliser la simulation, c'est-à-dire la stratégie définissant l'ordre dans lequel les nœuds sont simulés. Nous montrons que les chemins dits de "dégroupement", c'est-à-dire

qui maximisent la distance entre les nœuds simulés consécutivement, conduisent à une meilleure reproduction de la structure spatiale dans les résultats de la simulation. Dans un deuxième temps, nous évaluons le gain en temps de calcul et les biais résultants de l'utilisation d'un chemin constant lors de plusieurs réalisations. Les résultats montrent que les biais résultant sont minimaux et facilement surpassés par un gain considérable en temps de calcul. Ceci permet d'augmenter la taille du voisinage utilisé lors de la simulation, et, au final, de réduire l'ampleur globale des biais dans les différentes réalisations.

La seconde étape consiste à développer une version améliorée de la simulation séquentielle bayésienne (SSB). Cette méthode de simulation permet d'intégrer une variable secondaire connue dans la simulation stochastique d'une variable primaire. Elle est fondée sur une simulation SGS à laquelle s'ajoute, pour chaque nœud simulé, l'intégration d'une variable secondaire co-localisée. Pour cela, la distribution conditionnelle issue de la simulation SGS est combinée avec une distribution provenant de la valeur connue de la variable secondaire. Notre proposition consiste à généraliser cette combinaison en attribuant un poids log-linéaire à chaque distribution. La nouveauté essentielle consiste alors à concevoir un schéma de pondération qui adapte la valeur des poids au cours de la simulation pour tenir compte de la variation de dépendance entre les deux sources d'information. Pour évaluer les gains obtenus par cette nouvelle approche, des tests sont effectués à partir d'une étude de cas hydrogéophysique consistant à simuler la conductivité hydraulique en utilisant comme source secondaire la tomographie en surface de résistivité électrique. Cette étude de cas montre que le schéma de pondération proposé améliore considérablement la reproduction de la structure spatiale tout en maintenant en accord les variables primaires et secondaires.

Enfin, la troisième étape consiste à développer une méthodologie capable d'augmenter la résolution des images tomographiques résultant d'inversions de données géophysiques soumises à des contraintes de lissage. L'idée clé est d'utiliser la matrice de résolution, calculée lors de l'inversion pour quantifier le lissage du tomogramme à travers un mapping linéaire. En utilisant le krigeage "zone-à-point", il est alors possible de simuler des réalisations à une échelle fine de la conductivité électrique contraintes au tomogramme par le mapping linéaire précédemment calculé. La méthode développée est capable de fournir plusieurs réalisations à un coût de calcul relativement faible. Ces réalisations reproduisent fidèlement la structure spatiale et la correspondance avec le tomogramme.

Abstract

Around the world, groundwater is vital for humankind, yet threatened by anthropogenic deterioration. Sustainable groundwater management is thus crucial and relies on numerical models for adequately forecasting groundwater conditions. A key step in the construction of such models is to determine hydrogeological parameters that describe the behaviour of the water in the aquifer (i.e. water-bearing rock). In recent years, the increasing application of geophysical techniques to characterize the aquifer has improved this endeavour. However, the integration of geophysical data with hydrogeological parameters remains a challenging task. This is due to fundamental differences in coverage and resolution as well as the complex and non-unique interrelation of the measured parameters. Geostatistics have proven to be a useful framework to integrate these spatial datasets. This thesis takes a broader perspective to address this topic, which can be summarized as: "developing computationally efficient and accurate methodologies for the integration of spatial datasets, which are variable in terms of coverage and resolution, and related through complex, site-dependent and/or non-unique relationship". The contribution presented in this thesis can be partitioned into three stages.

The initial stage is concerned with the first part of the problem statement taking a more general and theoretical geostatistical approach. More specifically, it aims to improve the efficiency and accuracy of Sequential Gaussian Simulation (SGS), which is a widely used geostatistical method employed to generate Gaussian fields. It populates a grid by consecutively visiting each node and sampling a value in a local conditional distribution. In the first project, we look at the impact of the type of simulation path, that is, the strategy defining the order in which the nodes are simulated. It is shown that declustering paths, which maximize the distance between consecutively simulated nodes, present the best reproduction of spatial structure. The second project assesses the computational gain and resulting biases of using a constant path for multiple realizations. Results show that these biases are minimal and

easily surpassed by the high computational gains, which in turn allow for increasing the neighbourhood size and thus reducing the overall magnitude of biases.

In a second stage, an improved version of Bayesian Sequential Simulation (BSS) is proposed. BSS integrates a known secondary variable in the stochastic simulation of a primary variable. The method is based on SGS with the addition that, for each simulated node, the conditional distribution is combined with a distribution coming from the known value of the collocated secondary variable. Our proposition is to generalize this combination by assigning a log-linear weight to each distribution. A key novelty of this work is to design a weighting scheme that adapts its values along the simulation to account for the variation of dependence between both sources of information. Tests are performed for a hydrogeophysical case study consisting of simulating hydraulic conductivity using surface-based electrical resistivity tomography as the secondary variable. This case study shows that the proposed weighting scheme considerably improves the realizations in terms of reproducing the spatial structure while maintaining a good agreement between primary and secondary variables.

In the third and final stage, we develop a methodology capable of downscaling tomographic images resulting from smoothness-constrained inversions of geophysical data. The key idea is to use the resolution matrix, computed during the inversion, to quantify the smoothing of the tomogram through a linear mapping. Using area-to-point kriging, it is then possible to simulate fine-scale realizations of electrical conductivity constrained to the tomogram through the previously computed linear mapping. The method developed is able to provide multiple realizations at a relatively low computational cost. These realizations accurately reproduce the spatial structure and the correspondence to the tomogram.

Chapter 1

Introduction

1.1 Preamble

1.1.1 Water paradoxes

Water is arguably the most vital molecule on our planet, and yet, it is a double-edged sword for our society today. Without pretence of drawing a thorough analysis, I outline a few apparent contradictions below:

- Many of its unique properties (low density of ice, high solvent power, and high heat capacity), resulting from the inherent polarity of the H_20 molecule, make life on Earth possible. Yet, throughout history, 224 conflicts have reportedly been triggered by water, hindering human life (Gleick, 2000).
- Water is present everywhere, covering 71% of our planet and yet, 884 million people still lack basic water drinking services UNICEF and WHO (2017).
- The importance of water for human economy (agriculture, fisheries, manufacturing industry, hydropower, recreational) is invaluable (e.g., EPA, 2013; van Ast et al., 2013) and yet, countries fail to take its sustainable management seriously. This was illustrated by the nomination of a climato-sceptic to the head of US Environmental Protection Agency who revoked the Clean Water Rule.

- Providing safe drinking water to 88% of the world population was the first Millennium Development Goal to be achieved in 2010 (WHO and UNICEF, 2012). Yet, this early success is hampered by a darker prediction: due to climate change and rising demand, it is expected that by 2025, two-thirds of the world population could be under water stress conditions (WWAP, 2012).
- In scientific literature, water is a hot topic with 6.3% of all records from the Web of Science platform recorded under the topic "water" (accessed on August 1, 2018). And yet, research is still unable to unravel its very origins on earth (Altwegg et al., 2015).

In light of these contradictions, water can be described as both vital for our civilisation and threatened by it. This complexity raises the need to better understand the presence of water and its movements, notably in the subsurface.

1.1.2 Hydrogeology: introduction and challenges

In this thesis, we focus on groundwater, which is studied in the field of hydrogeology. Understanding three key characteristics of groundwater is helpful in apprehending the reasons for its importance: a large storage volume per unit inflow, a relatively slow movement and a ubiquitous presence at the continental scale contrasted by a heterogeneous distribution at local to regional scales (Giordano, 2009; Gleeson et al., 2016).

Based on these properties, groundwater can be qualified as the buffer of the freshwater cycle. Indeed, it provides a reliable storage during the wettest parts of the year and a source of water during the driest parts of the year for both human use and natural ecosystems (e.g. baseflows of rivers). In addition, the long and slow migration of water through the soil matrix allows for a filtering of the water, making it a safe source of drinking water.

However, these characteristics also lead to challenges. The typically vast and heterogeneous distribution of groundwater together with the impossibility to observe it directly make it difficult to localize this resource, quantify its volume, and assess its quality. Furthermore, the remediation of contaminated groundwater is a laborious task.

The extent to which humans rely on groundwater resources and its usage varies greatly. As this cannot be discussed in depth here, readers are referred to Korzoun and Sokolov (1978) and Shiklomanov (2000) for information on the global distribution of water, Morris et al. (2003) and Zektser and Everett (2004) for a detailed description of groundwater use, Aquastat (FAO, 2016) for a large database on global water, and, WHYMAP (WHYMAP, 2002) for a worldwide hydrogeological mapping tool.

Groundwater is under tremendous threat worldwide. Below, a few threats are illustrated each with an example:

- **Land use change.** Plantations of eucalyptus in the Zululand coastal aquifers in South Africa have greatly reduced the water table because of their deep roots (Le Maitre et al., 1999; Albaugh et al., 2013). On the other hand, in the Wights and Lemon catchments in Western Australia, streamflow and groundwater has greatly increased following the destruction of the Jarrah forest for agricultural purposes (Ruprecht and Stoneman, 1993).
- **Excessive exploitation.** Global water demand has expanded drastically during the last century, leading to a depletion of groundwater throughout the globe (Vorosmarty, 2000; Wada et al., 2010). This depletion is particularly noticeable in certain areas of the US High Plains and is exacerbated by the pumping of fossil groundwater. As a result, Scanlon et al. (2012) predicts that these areas will be unable to support irrigation within the next 30 years.
- **Eutrophication.** Groundwater plays a considerable role in the eutrophication of surface water by transporting nitrate and phosphorus (Holman et al., 2008). Conversely, surface water also contaminates groundwater. In turn, this eutrophication is linked to the extinction of some the whitefish species (*Coregonus* spp.) in the Swiss pre-Alps lakes (Vonlanthen et al., 2012).
- **Pollution.** In the Katanga region in Eastern Democratic Republic of Congo effluents with high concentrations of heavy metals stemming from mines exploitations, contaminate both surface waters and groundwater, posing significant human health risks (Atibu et al., 2013). This is exacerbated by the conflict raging in this region and the fact that industries are generally small-scale, which makes it a difficult context for the design and enforcement of government regulations.

- **Salinity.** Groundwater in the coastal area of Bangladesh is threatened by saltwater incursion from the Bay of Bengal. This is indirectly caused by climate change through sea level rise and high tidal waves, as well as by agricultural withdrawal upstream leading to a small groundwater recharge of freshwater. This has been linked to poor health conditions for local inhabitants (Khan et al., 2011).
- **Sea level rise.** Famous for already having lost two islands (Warne, 2015), the Republic of Kiribati is facing land disappearance due to climate change (Webb and Kench, 2010). Yet, the greatest threat to human life on the islands might well come from the reduced availability of its groundwater. This groundwater takes the form of freshwater lenses floating on denser saltwater (Kuruppu and Liverman, 2011), which, due to sea level rise, are expected to decrease drastically or even disappear Metai (1997).

To address such challenges adequately and sustainably, local solutions need to be found through social, economic, technical, and political measures. All these rely on tools allowing an accurate quantification of underground water presence and movement for informed decision-making. In this context, hydrogeological modelling can provide fundamental insights.

1.1.3 Modelisation of hydrogeological systems

1.1.3.1 Why use a model?

Groundwater modelling provides a quantitative framework to help answer specific questions, when the alternative consisting of experiments and measurements (e.g., Darcy, 1856) is not feasible due to practical constraints.

For instance, going back to the challenges mentioned above, a good groundwater model could significantly contribute to answer the following questions:

- How would a new eucalyptus plantation affect the streamflow and groundwater level in the Jarrah Forest?
- What would be the maximum sustainable rate of withdrawal for farmers in the high plains of the US to avoid groundwater depletion?

- What would be the effect on phosphorus concentration in pre-Alps lakes if more vegetation were planted on the lakeshore?
- Which mining regulation would be most effective to minimize the health impacts resulting from heavy metal contamination in the groundwater of the Katanga region?
- What is the origin and the extent of salinization of the Bangladesh coastal groundwater?
- Where could we drill a well in Kiribati island to find freshwater? What would be a safe pumping rate to avoid salinization of these freshwater lenses?

These questions illustrate the first of the two purposes of a model, which is to forecast the behaviour of a system with a proposed action or inaction. Anderson et al. (2015) put forward a second purpose, namely to improve the understanding of the processes which govern the system. These models are referred to as interpretative models. An example of this second purpose is the work of Ireson et al. (2012). In order to better understand the challenging hydrogeological and chemical processes in the seasonally frozen soil of the northern latitudes, they used a modelling technique to provide important insights on the dynamic of water flow and salt movement.

Besides the realms of science and engineering, the benefits of using a model can be extended to other areas, such as legislation and law courts as they provide the technical tools and expertise to inform decision making. An example of the former is the implementation of the European Union Water Framework Directive requiring regular monitoring of groundwater, which can only be achieved with a numerical model (Hulme et al., 2002). An illustration of the latter can be found in A Civil Action (Zaillian, 1999; Harr, 1995), which recounts the true story of a court case treating of a groundwater pollution in Woburn, Massachusetts in 1986. Models of the groundwater and the spread of contamination presented by the expert played a key role in the trial (Bair and Metheny, 2011).

Before going into more detail, some limitations with regard to the role and possibilities of models need to be highlighted:

- It is impossible to model all processes, hence “all models are wrong, but some are useful” (Box and Draper, 1987)

- Model users need to pay attention not only to the known assumptions made by the model but also to the unknown unknowns, as famously put by Rumsfeld (2002), although in a unrelated context: "[...]as we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns – the ones we don't know we don't know. [...] it is the latter category that tend to be the difficult ones".
- Non-uniqueness in a model refers to the fact that different input parameters can provide the same outputs. This has to be accounted for in the interpretation of model results.
- Even sophisticated models will only be as good as the input data they are provided with, a problem familiarly described by computer scientists as: "garbage in, garbage out".

1.1.3.2 What is a model?

Scientific modelling is the generation of a purposeful, simplified and idealized representation of a particular phenomenon that is difficult to observe directly. This definition can be extended by exploring the procedure of building a model, commonly structured into the following steps,which are also illustrated in Figure 1.1 (e.g., Anderson et al., 2015; Essink, 2000; Baalousha, 2011).

1. Define the purpose of the model. This step has already been developed in section 1.1.3.1.
2. Conceptualize the model by identifying the relevant aspects of the system with regard to the purpose. This model becomes idealized as it omits certain parts of reality. "Whereas it is practically impossible to separately observe all phenomena connected with a regime of groundwater flow, a correct theory discloses every feature and draws attention to the most important properties of the flow even if they might be otherwise over- looked." (Tóth, 1963). Within this conceptualization step, it is possible to identify a physical model, which focuses on the representation of the physical processes of each aspect of the conceptual model. Simplifying assumptions of the physics are made based on the current knowledge of the system.
3. Describe the physical processes with mathematical formulae. Mathematical models can be characterized as analytical or numerical, deterministic or stochastic, lumped

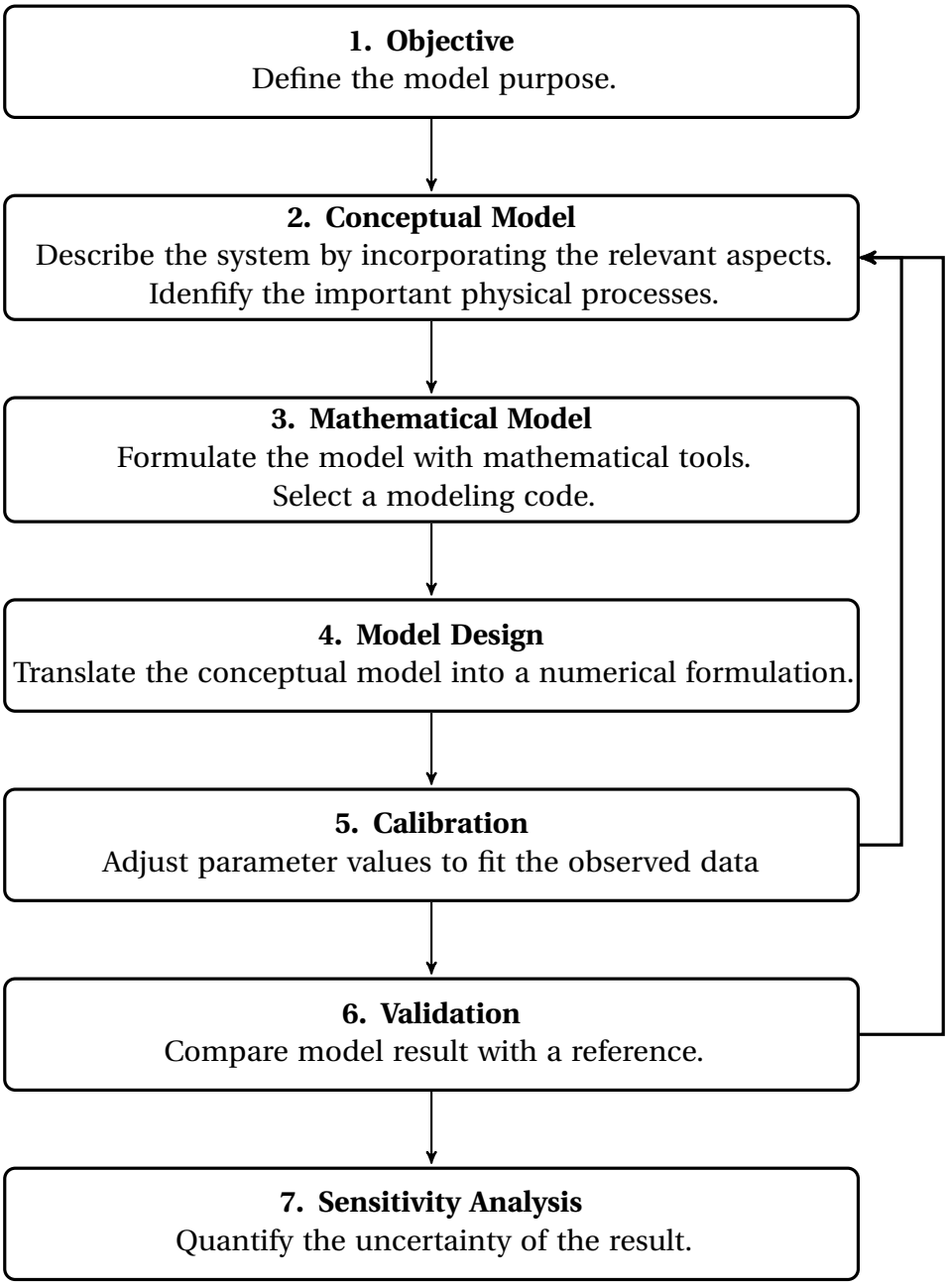


Figure 1.1 – General procedure of groundwater modeling. Adapted from Anderson et al. (2015), Essink (2000) and Baalousha (2011).

or distributed, black box or white box, physics-based or empirical, transient or steady state, linear or non-linear, 1, 2 or 3D.

4. Use numerical methods to implement the decisions made in the conceptual model. This includes designing the domain size (time and space), boundary and initial conditions, and hydrogeological parameters.
5. Adjust the hydrogeological parameters of the model to observed available data using a calibration method.
6. Check that the calibrated model is coherent with external source of information. Cross-validation is a commonly used technique if no external information are available.
7. Evaluate variability of the model response originating from the known input error, the uncertainty of parameters, or possible the internal stochasticity of the model.

In this thesis, we are particularly interested in providing accurate hydrogeological parameters in the model (steps 4 and 5 in Figure 1.1). This topic is the primary focus of next section.

1.1.4 Hydrogeological parameters

In order to better understand the importance of accurately estimating hydrogeological parameters, this section details the central equation of groundwater flow (mathematical model) and the nature and relevance of hydrogeological parameters. Figure 1.2 illustrates the link between steps 3-5 of Figure 1.1 The parameters as well as the initial and boundary conditions are implemented in the mathematical model in the model design step. These parameters are then calibrated so that for a given input, the output of the model corresponds to the observed data.

In hydrogeology, the most fundamental governing equation used to represent water movement is constructed by coupling the mass balance equation and Darcy's law (Darcy, 1856) into the 3D transient ground water flow for heterogeneous and anisotropic conditions Anderson et al. (e.g., 2015)

$$S_s \frac{\partial h}{\partial t} = -\nabla(-K\nabla h) + W. \tag{1.1}$$

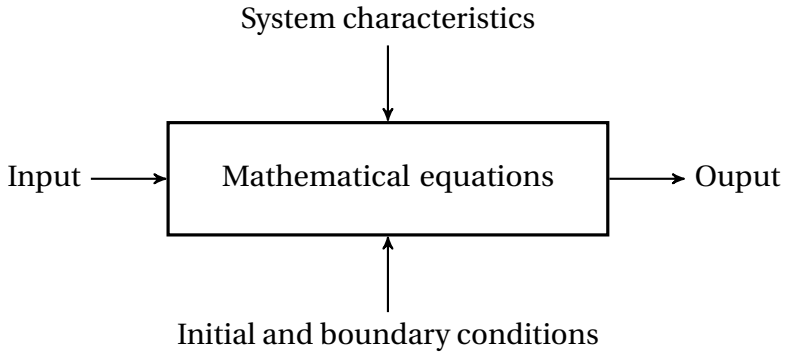


Figure 1.2 – General structure of the mathematical model in link with model design. Adapted from Essink (2000, p. 9)

This equation describes that the change of hydraulic head h over time t multiplied by the specific water storage S_s is equal to the negative of the gradient of the flux $q = -K\nabla h$ plus a sink/source term W. The capacity of the soil to transmit a fluid is quantified by the hydraulic conductivity K. When the purpose of the model is to quantify the transport of a solute, the popular advective-dispersive equation Anderson et al. (e.g., 2015) is used in addition

$$\phi \frac{\partial C}{\partial t} = -\nabla(-D\nabla C) - q\nabla C + R, \tag{1.2}$$

where the change over time of a concentration C multiplied by the porosity ϕ depends on the gradient of the diffusion flux given by Fick's law (Fick, 1855) $j = -D\nabla C$, the advection on the flow q, and a sink/source term R.

The most common way to solve these equations is through a numerical method, which approximates the differential equations, for example, with a finite difference or finite element approach.

The boundary conditions are included in the numerical model by modifying the equations at the nodes located on the edge of the domain according to the type of boundary conditions chosen (e.g. Dirichlet, Neumann, Cauchy). Initial conditions are simply imposed on the first step. The sink/source terms R or W are generally known and consequently added in the numerical model at the appropriate time and location. Finally, the main task left for the design model step is to provide estimate for the hydrogeological parameters K , S_s , D, and K . In the numerical method, these parameters are discretised on the domain, so that they

need to be estimated on a grid of same dimension and size than the mathematical model. Note that D and K are tensors, so that for each location, corresponding 3x3 matrices need to be estimated.

1.2 Motivation: Aquifer characterization

The overarching topic of this thesis is to develop methodologies capable of integrating spatial datasets of vastly differing coverage and resolution. This general problematic arose from its specific application to aquifer characterization. Below, I show how such methodologies are required for the combined integration of hydrogeological and geophysical datasets, which is the focus of chapters 4 and 5.

1.2.1 Aquifer characterization

As presented in the preamble, the output of hydrogeological models is highly dependent on accurate hydrogeological parameters. Estimating those parameters belongs to the general field of aquifer characterization, which is interested in evaluating all the parameters related to the aquifer, which is the soil/rock containing the groundwater. This broad field includes multiple facets, such as, for instance, describing the structure of rock formations, determining the location of bedrock, or the depth of water table.

The methodologies developed in this thesis could, in theory, be applied to several, if not all, spatially distributed parameters used in aquifer characterization. However, I focus below on hydraulic conductivity, as it is widely regarded as the most important, but also the most elusive hydrogeological parameter, and its estimation has been the primary focus of application in this thesis.

1.2.2 Hydraulic conductivity

Hydraulic conductivity is arguably one of the most, if not the most, essential and complex parameter in groundwater models as it defines the water flow, which then controls the infiltration within the vadose zone, groundwater movement, and the spread of contaminants.

In order to describe it properly, it is convenient to decompose hydraulic conductivity K into the contribution of the soil and of the fluid,

$$K = \kappa \frac{\rho g}{\mu}, \tag{1.3}$$

with the permeability κ, the constant of gravitation g, the fluid density ρ and viscosity μ.

At the macroscopic level, the permeability depends on various pore space properties such as pore size and shape, pore distribution, pore surface morphology and arrangement of pore (connectivity and tortuosity) (e.g., Schön, 2004).

At the mesoscopic level, permeability presents a complex and heterogeneous distribution. Firstly, its range of values varies considerably with the type of soil or rock material, ranging from 1 to $10^{-}13$ m/s (Freeze and Cherry, 1979). Therefore, hydraulic conductivity values present a complex spatial structure depending on the various and possibly overlapping geological units (e.g. layered, fractured, or trending). Secondly, within the same geological unit, its distribution is generally skewed and is commonly modelled with a log-normal distribution (Freeze and Cherry, 1979). Thirdly, hydraulic conductivity is directional, presenting anisotropy due to the sedimentation or orientation of fractures. Finally, in an unsaturated soil, the value of hydraulic conductivity is related non-linearly to water content (Van Genuchten, 1980) which increases its variability.

The consequences are numerous when it comes to the estimation of hydraulic conductivity: measurements are scale- and orientation-dependant, so that upscaling and downscaling are non-trivial tasks. Because the flow is mainly controlled by highly conductive zones, an accurate representation of the spatial distribution of extreme values (i.e. low and high hydraulic conductivity) is critical, particularly in terms of connectivity (e.g., Renard and

Allard, 2013). This has even more significant consequences for transport modelling (e.g., Poeter and Gaylord, 1990). Furthermore, the proper estimate of its range and distribution requires many samples because of its skewness.

This is why it is essential for aquifer characterization to provide high-resolution fields of hydraulic conductivity throughout the study domain that reproduce the spatial structure, both in terms of the small-scale texture and the larger-scale features. In addition, the uncertainty in data and methods used in aquifer characterization should be adequately reproduced in the hydrogeological parameters. This is typically performed by providing multiple hydraulic conductivity fields spanning over the range of possibilities.

Estimation of the hydraulic conductivity of a field is traditionally performed with either a sample in a laboratory or a field experiment. A list of these methods can be found in Hubbard and Rubin (2005, p. 8, Table 1.1) and their complete description in Butler (2005). A severe limitation to these simple technique is that they can only provide information near wells (Butler, 2005). The lack of hydrogeological methods for moderate to high resolution on a regional scale highlights is illustrated in Hubbard and Rubin (2005, p. 6, Figure 1.1). This figure also highlights the potential role of geophysics to fill this gap.

1.2.3 Hydrogeophysics

Geophysics studies physical processes and properties of the Earth. Historically, it started with the main mechanisms of the Earth (e.g. gravitational and magnetic fields, seismology, volcanism, rock formation), and then extended to exploration geophysics (e.g. mineral deposits and hydrocarbon reservoir). Applications to environmental sciences, such as groundwater, glacier or site remediation came later with the development of hydrogeophysics, defined as the "use of geophysical measurements for mapping subsurface features, estimating properties, and monitoring processes that are important to hydrogeological studies" (Hubbard and Rubin, 2005). For such applications, geophysical tools need to be adapted for hydrogeological situations of low temperature, pressure and less consolidated materials, which tend to be associated to a more variable environment. A list of geophysical techniques used

for hydrogeology can be found for instance in Hubbard and Linde (2010) and Hubbard and Rubin (2005).

These methods bear great potential for aquifer characterization (e.g., Hubbard and Rubin, 2005; Linde et al., 2006; Binley et al., 2015). Indeed geophysical techniques present several advantages complementary to hydrogeological estimation techniques. Multiple scales of investigation can be targeted with different methods, from laboratory- to regional-scale ($10^{-}4$ to 10^{5} m). As mentioned in the previous section, unlike hydrogeological methods, some geophysical techniques easily cover large surfaces (regional scale), providing spatially distributed values with moderate resolution. Lastly, these methods generally present the advantage of using of indirect and non-invasive techniques at a relatively low cost.

1.2.4 Integration of hydrogeophysical datasets

One of the main challenges of hydrogeophysics is the integration of datasets (Hubbard and Rubin, 2005; Linde et al., 2006). The reasons are the following:

- Datasets can vary in their nature (e.g. a qualitative opinion of a geologist expert with quantitative measurement) and present various levels of errors.
- Different datasets may present different coverage and resolutions. In addition, a measurement can be punctual or areal with various and possibly overlapping support scales.
- The relationships linking geophysical and hydrogeological properties are complex, including non-uniqueness, uncertainty, non-stationarity, poorly understood and underlying assumptions that are difficult to assess.
- Finally, the dataset size might pose computational problems.

A comparative list of methods for integrating hydrogeophysical datasets can be found in Linde et al. (2006) or Binley et al. (2015). These challenges form the origin for the goal of this thesis.

1.3 Problem statement

The general objective of this thesis can be summarized as:

"Developing computationally efficient and accurate methodologies for the integration of spatial datasets which are variable in terms of coverage and resolution, and present complex relationships (e.g. non-linearity, non-uniqueness and partially redundant information)."

1.4 Methodological background: Geostatistics

Solving the problems stated above requires the use of a set of mathematical tools formalized within geostatistics. The purpose of this section is to introduce these tools (1.4.1), motivate the need for their efficient implementation (1.4.2), and present the mathematical foundations necessary for a good understanding of this thesis (1.4.3).

1.4.1 Introduction

In its essence, geostatistics draws on the concepts and on the mathematical framework of classical statistics to describe and quantify a phenomenon spanning over space. Geostatistics is essentially enhancing conventional interpolation methods by providing a way to quantify, with uncertainty, a variable distant from measurement points. This information can be estimated or simulated. The difference is that estimation consists of finding the most likely value, while simulation is involved in finding a single possible/coherent value. Historically, the difference between the two has often been neglected; an enlightening example of the importance of simulation can be found in Lantuéjoul (2002, p.1).

Before discussing the use of geostatistics, it is important to understand two of its basic assumptions.

- Geostatistics assumes that physical processes can be regarded as a stochastic models, implying that the observed data are one of many possible realizations of an underlying random function. However, as stated by Chilès and Delfiner (1999, p.3) “Probabilities do not exist in Nature but only in our models”. The key here is that stochastic models are used not because observations are non-deterministic but because it is analytically useful to consider them as such.
- Geostatistics relies on the existence of some sort of spatial structure in the process under study. This is connected to the first Law of Geography (Tobler, 1970) "Everything is related to everything else, but near things are more related than distant things."

On a more practical side, the purpose of geostatistics can be organized in the following categories (Chilès and Delfiner, 1999), as illustrated by corresponding examples:

- **Survey optimization.** Where should an additional well be drilled to maximize the constraints on the location of a contaminant plume?
- **Integration of information.** How to combine the information of scattered high-resolution hydraulic conductivity values with low-resolution measurements of the electrical conductivity to characterize an aquifer?
- **Visualised spatial information.** How to present a spatio-temporal dataset of bird density measured by several weather radars?
- **Uncertainty and risk assessment.** What is the P10 and P90 (i.e. values such that 10%, and 90% respectively, of estimates exceed these values) of the recoverable petrol in a given reservoir?
- **Change of support.** How to downscale a coarse Digital Elevation Model (DEM) available over a large area to a fine resolution knowing the true resolution only at certain location?

The work of this thesis is primarily concerned with the integration of information but is also relies on methods concerned of change of support. Uncertainty and visualization are also treated to a lesser extent.

1.4.2 Improving algorithms

Moore's law (Moore, 1998), which stipulates a doubling of the number of components per integrated circuit every year, is widely understood and popularized as the explanation for the sustained massive growth of computational power. This famous law should, however, not be taken to mean that hardware is the unique and exclusive solution for computational challenges. Indeed, algorithms can have a much more significant impact on computational performance, as reported by Grötschel in Holdren et al. (2010) "a benchmark production planning model solved using linear programming would have taken 82 years to solve in 1988, using the computers and the linear programming algorithms of the day. Fifteen years later this same model could be solved in roughly 1 minute, an improvement by a factor of roughly 43 million. Of this, a factor of roughly 1,000 was due to increased processor speed, whereas a factor of roughly 43,000 was due to improvements in algorithms!" .

The limitation of hardware and the necessity of performant algorithms are even stronger in association with the concept of scalability in computational problems. Computational complexity classifies an algorithm by its scalability. It describes the relations between the amount of resources required to solve a problem (e.g. time, elementary operation or memory) and its input size (e.g. number of entries, size of a grid). With the current increase of datasets size, the linear improvement of hardware capability described in Moore's law falls short of providing sustainable solutions for solving algorithms of high order of complexity, such as quadratic problems (e.g. quicksort) or exponential problems (e.g. travelling salesman problem). Thus, when dealing with big data, algorithmic improvements measured in terms of complexity are of tremendous importance.

Big data is very much a challenge for the hydrogeology community. 3D discrete models of groundwater can easily reach several million nodes. Solving flow or transport-reaction models on such grids becomes very sensitive to the algorithmic efficiency of the solver. In the context of this thesis, the generation of hydrogeological parameters on such a grid is also a major challenge. This is why algorithms developed in geostatistics should have a computational complexity that scales well with the size of the problem.

In this thesis, the first two chapters consider the algorithmic efficiency of a geostatistical method named Sequential Gaussian Simulation. They both aim to provide guidelines for achieving optimal trade-offs between fast and accurate simulations.

1.4.3 Mathematical preliminaries

Several key concepts of geostatistics are briefly recalled below, as they will be used in the various chapters of the thesis. Readers are referred to standard textbooks for more detailed information (e.g., Journel and Huijbregts, 1978; Isaaks and Srivastava, 1989; Chilès and Delfiner, 1999; Goovaerts, 1997).

1.4.3.1 Random variable

In probability theory, a random variable (RV) denoted Z is a deterministic function mapping the set of possible outcomes of a random phenomenon to their values. A realization z of this random variable is the real value associated with the specific outcome ω of the random phenomenon,

$$\begin{array}{l} Z : \Omega \to \mathbb{R} \\ \quad \omega \mapsto z \end{array} . \tag{1.4}$$

For instance, the sum of two dices can be represented by a random variable, for which all possible combinations of the two dices are mapped to their sum. This formalism allows for defining rigorously the probability of this random function.

In geostatistics, it is convenient to associate several random variables with an index $\mathbf{u}$ representing, for instance, space or time. This can be defined by a random function such that

$$z(\mathbf{u}) = Z(\omega, \mathbf{u}) . \tag{1.5}$$

This extension of random variables is more generally known as a stochastic process for 1D indexing and a random field for multidimensional indexing. The term “regionalized variable” was introduced by Matheron (1963, 1965) for continuous space domains.

1.4.3.2 Stationarity

Stationarity is a mathematically convenient property of a random function, which ensures that all random variables $Z(\mathbf{u})$ are identically distributed. That is, for all $\mathbf{u}$, $Z(\mathbf{u})$ has the same probability distribution. This implies that the stochastic law of the process is invariant by translation.

Second-order stationarity relaxes the above concept of strict stationarity by requiring only the first two moments of Z to be stationary, that is, mean and variance.

Intrinsic stationarity is an even more flexible assumption, which imposes that the increment of Z be second-order stationary,

$$\begin{cases} \mathrm{E}\left[Z\left(\mathbf{u}+\mathbf{h}\right)-Z\left(\mathbf{u}\right)\right]=0 \\ \mathrm{E}[Z\left(\mathbf{u}+\mathbf{h}\right)-Z\left(\mathbf{u}\right)]^2=2\gamma\left(\mathbf{h}\right) \end{cases}, \tag{1.6}$$

where the variance of the increment $\gamma(\mathbf{h})$ is only dependent on the distance $\mathbf{h}$. This function is known as the semi-variogram and is discussed below.

1.4.3.3 Covariance function and variogram

The covariance function C_Z of the random function Z between two locations $\mathbf{u_1}$ and $\mathbf{u_2}$ is defined by

$$C_Z\left(\mathbf{u}_1,\mathbf{u}_2\right)=\mathrm{cov}\left(Z\left(\mathbf{u}_1\right),Z\left(\mathbf{u}_2\right)\right). \tag{1.7}$$

The assumption of second-order stationarity on Z simplifies the representation of the covariance function to a one-parameter function, known as covariogram

$$C_Z\left(\mathbf{h}\right)=\mathrm{cov}\left(Z\left(\mathbf{u}\right),Z\left(\mathbf{u}+\mathbf{h}\right)\right). \tag{1.8}$$

The correlogram normalizes the covariogram with respect to the value at lag 0,

$$\rho_Z(\mathbf{h}) = \frac{C_Z(\mathbf{h})}{C_Z(\mathbf{0})}. \tag{1.9}$$

Instead of looking at the covariance, one can use the variance of the difference of between the two locations $\mathbf{u_1}$ and $\mathbf{u_2}$ through the semi-variogram,

$$\gamma_Z(\mathbf{u}_1, \mathbf{u}_2) = \frac{1}{2}\mathrm{var}\left(Z(\mathbf{u}_1) - Z(\mathbf{u}_2)\right). \tag{1.10}$$

The variogram can be related to the covariance function for a random variable $Z(\mathbf{u})$ with intrinsic stationarity through

$$\gamma_Z(\mathbf{h}) = C_Z(\mathbf{0}) - C_Z(\mathbf{h}). \tag{1.11}$$

1.4.3.4 Ergodicity

A stationary random function is ergodic if its spatial average converges to its expected mean

$$\lim_{V\to\infty} \frac{1}{|V|}\int_V Z(\mathbf{u})\, d\mathbf{u} = \mathrm{E}\{Z(\mathbf{u})\}. \tag{1.12}$$

The same property exists for the covariance called second-order ergodicity. Ergodicity is necessary to be able to infer the means or covariance function of a stationary random function from a single realisation. Ergodic fluctuation is used to describe the fact that a discrepancy is expected between the empirical statistics of a realization and the theoretical model (Emery, 2004). The property of ergodicity is also relevant in the selection of domain size (Lantuéjoul, 1991).

1.4.3.5 Change of support

The change of support or regularization is used to combine variables with different support scales. This can be generalized with any average function $p(\mathbf{x})$, such that

$$Z_p(\mathbf{u}) = \int p(\mathbf{x})\, Z(\mathbf{u}+\mathbf{x})\, d\mathbf{x}, \tag{1.13}$$

or written with a convolution as $Z_p = Z * \breve{p}$, where $\breve{p}(\mathbf{u}) = p(-\mathbf{u})$. The covariance of the averaged process can be computed with

$$C_p = C_Z * \left(p * \breve{p}\right). \tag{1.14}$$

1.4.3.6 Gaussianity

A random function Z is Gaussian if all its finite dimensional distributions are multivariate Gaussians. That is, each linear combination of its components $\sum_i a_i Z(\mathbf{u}_i)$ is normally distributed. A Gaussian RF is fully described by its mean vector μ and covariance matrix $\mathbf{C_Z}$

$$Z(\mathbf{u}) \sim \mathcal{N}\left(\mu_{\mathbf{Z}}, \mathbf{C_Z}\right). \tag{1.15}$$

1.4.3.7 Kriging

The various kriging equations (simple, ordinary, universal) are elaborations of the regression of the following linear combination

$$Z^*(\mathbf{u_0}) - m(\mathbf{u_0}) = \sum_{\alpha=1}^{n_\alpha} \lambda_\alpha \left[Z(\mathbf{u}_\alpha) - m(\mathbf{u}_\alpha)\right]. \tag{1.16}$$

This equation expresses the unknown value Z^* at the location $\mathbf{u_0}$ as a linear combination of the known value Z at location $\mathbf{u}_\alpha$. $m_Z(\mathbf{u})$ is the expected value of Z in $\mathbf{u}$. The kriging weights λ_α quantify the relative importance of each point. Their values are inferred by solving a linear system of covariance, and, complementary constraints for ordinary and universal kriging.

Kriging has been extended to integrate the contribution of another variable Y with co-kriging,

$$Z^*(\mathbf{u_0}) - m_Z(\mathbf{u_0}) = \sum_{\alpha=1}^{n_\alpha} \lambda_\alpha \left[Z(\mathbf{u}_\alpha) - m_Z(\mathbf{u}_\alpha) \right] + \sum_{\beta=1}^{n_\beta} \lambda_\beta \left[Y\left(\mathbf{u}_\beta\right) - m_Y\left(\mathbf{u}_\beta\right) \right]. \tag{1.17}$$

1.4.3.8 Screening and relay effect

In kriging, the screening effect refers to a situation where the "nonzero kriging weights are concentrated on a subset of samples in the immediate vicinity of the estimated point" (Chilès and Delfiner, 1999, p. 202). As a result of this effect, it is common practice to exclude distant known values in the kriging estimation for computational reasons.

The relay effect describes the fact that a nonzero kriging weight can be assigned to a sample whose distance to the estimated point is equal to or greater than the range, because of the presence of an intermediate node.

1.5 Thesis structure

During the course of my PhD, I have tried, together with my supervisors and colleagues, to tackle some of the challenges of hydrogeophysical data integration. I started the PhD with the overall purpose to "develop methodologies that are capable of providing aquifer-scale models of the permeability distribution that allow for a faithful prediction of the pertinent flow and transport phenomena".

1.5.1 Chapters 2 and 3

In the first years, I concentrated on improving the algorithmic and computational aspects of BSS, which led me to study the more general sequential Gaussian simulation (SGS). I focused on the impact of the type of simulation path as well as the effect of using a constant

simulation path. These two works are presented in chapters 2 and 3 of the thesis and have been published as:

Nussbaumer, R., Mariethoz, G., Gloaguen, E., and Holliger, K. (2018a). Which Path to Choose in Sequential Gaussian Simulation. *Mathematical Geosciences*, 50(1):97–120. DOI:10.1007/s11004-017-9699-5.

Nussbaumer, R., Mariethoz, G., Gravey, M., Gloaguen, E., and Holliger, K. (2018b). Accelerating Sequential Gaussian Simulation with a constant path. *Computers & Geosciences*, 112(2018):121–132. DOI:10.1016/j.cageo.2017.12.006.

The corresponding code for these chapters is available on https://raphael-nussbaumer-phd.github.io/SGS/.

1.5.2 Chapter 4

Later, we came back to the initial task consisting of improving the downscaling of BSS. After observing a poor reproduction of the variance in BSS due to the assumption of conditional independence, we proposed to apply aggregation probability theory to BSS. This work is presented in chapter 4 and is under review in *Geophysical Journal International*.

The corresponding code for this chapter is available on https://github.com/Raphael-Nussbaumer-PhD/BSS

1.5.3 Chapter 5

While working on the problem of integrating low-resolution smooth tomograms with high-resolution scattered data, I developed a cokriging-based framework able to downscale the tomogram at the fine resolution while accounting exactly for the change of resolution and smoothing of tomogram. The resulting methodology is presented in chapter 5 and is under review in *Geophysical Journal International*.

The corresponding code for this chapter is available on https://raphael-nussbaumer-phd.github.io/A2PK/

Chapter 2

Which path to choose in Sequential Gaussian Simulation?

Raphaël Nussbaumer, Grégoire Mariethoz, Erwan Gloagen, Klaus Holliger

Published[1] in Mathematical Geosciences.

[1] Nussbaumer, R., Mariethoz, G., Gloaguen, E., & Holliger, K. (2017). Which Path to Choose in Sequential Gaussian Simulation. *Mathematical Geosciences*, **50**, 97–120. DOI:10.1007/s11004-017-9699-5

Abstract

Sequential Gaussian Simulation (SGS) is a commonly used geostatistical method for populating a grid with a Gaussian random field. The theoretical foundation of this method implies that all previously simulated nodes, referred to as neighbors, should be included in the kriging system of each newly simulated node. This would, however, require solving a large number of linear systems of increasing size as the simulation progresses, which, for computational reasons, is generally not feasible. Traditionally, this problem is addressed by limiting the number of neighbors to the ones closest to the simulated node. This does, however, result in artifacts in realizations. The simulation path, that is, the order in which nodes are visited, is known to influence the location and magnitude of these artifacts. So far, few rigorous studies linking the simulation path to the associated biases are available and, correspondingly, recommendations regarding the choice of the simulation path are largely based on empirical evidence. In this study, a comprehensive analysis of the influence of the path on the simulation errors is presented, based on which guidelines for choosing an optimal path were developed. The most common types of paths are systematically assessed based on the comparison of the simulation covariance matrices with the covariance of the underlying spatial model. Our analysis indicates that the optimal path is defined as the one minimizing the information lost by the omission of neighbors. Classification into clustering paths, that is, paths simulating consecutively close nodes, and declustering paths, that is, paths simulating consecutively distant nodes, was found to be an efficient way of determining path performance. Common examples of the latter are multi-grid, mid-point, and quasi-random paths, while the former include row-by-row and spiral paths. Indeed, clustering paths tend to inadequately approximate covariances at intermediate and large lag distances, because their neighborhood is only composed of nearby nodes. On the other hand, declustering paths minimize the correlation among nodes, thus ensuring that the neighbors are more diverse, and that only weakly correlated neighbors are omitted.

2.1 Introduction

Sequential Gaussian Simulation (SGS) is a popular technique to stochastically populate a grid with a Gaussian random field (Johnson, 1987; Journel, 1989; Isaaks, 1991; Deutsch and Journel, 1992; Gómez-Hernández and Cassiraga, 1994). In practice, SGS consists in visiting sequentially each node of the grid, computes the conditional probability distribution based on existing values using kriging, and assigns a value sampled from this distribution to the target node. This stochastic simulation technique has been applied in a wide range of disciplines, such as, for example, reservoir simulation (Verly, 1993), mining (Dimitrakopoulos et al., 2002; Zhao et al., 2007), hydrogeology (Lee et al., 2007), hydrology (Delbari et al., 2009), geophysics (Day-Lewis and Lane, 2004; Hansen et al., 2006; Abdu et al., 2008), soil science (Lin et al., 2001; Goovaerts, 2001), environmental science (Juang et al., 2004) and ecology (Mowrer, 1997).

A main drawback of this technique is its high computational cost originating from the computation of the kriging estimation of each node. The most widespread solution to this problem is the so-called limited neighborhood approach, that is, to keep only a limited number of conditioning nodes in the kriging estimation. However, the omission of information associated with this approach biases the kriging estimate, which in turn causes artifacts in the realizations. The simulation path, that is, the order in which nodes are simulated, has been recognized to influence the correctness of the simulation, however, these observations are largely based on empirical evidence (Tran, 1994; McLennan, 2002). A comprehensive study on how the sequential simulation path affects the occurrence and magnitude of bias is as of yet not available. This kind of information is, however, essential to allow for an educated choice of the path.

Here, this topic was addressed by presenting a comprehensive analysis of the connection between the simulation path and its impact on the deterioration of the realizations. Based on this, the most efficient path to preserve the desired simulation properties was investigated. Commonly used paths are thoroughly described, numerically assessed, and compared based

on the evaluation of the simulation covariance matrix (Emery and Peláez, 2011), thus allowing to draw general conclusions on the usefulness of each path.

The paper is organized as follows: Sec. 2.2 reviews the methodological background of SGS, limited neighborhood, and the simulation path; Sec. 2.3 presents the method used to quantify the bias in SGS; Sec. 2.4 attempts to provide the characteristics of an optimal path; Sec. 2.5 presents the analysis of various common paths; finally, Sec. 3.5 discusses limitations and constraints related to the choice of a given path.

2.2 Sequential Gaussian Simulation, Limited Neighborhood and Simulation Path

2.2.1 Background of Sequential Gaussian Simulation

SGS generates realizations $z^{(l)}(\vec{u})$ of a regionalized Gaussian random field $Z(\vec{u})$ at a discrete set of locations $\{\vec{u}_1,\ldots,\vec{u}_N\}$ by iteratively sampling a value for the field at each of these locations $z^{(l)}(\vec{u}_i)$. Based on the knowledge of the covariance C_Z, the estimation of $Z(\vec{u}_i)$ is performed by inferring the conditional mean and variance using kriging based on the previously simulated locations, often referred to as neighbors. With a zero-mean random field, SGS can be written as (Chilès and Delfiner, 1999)

$$z^{(l)}(\vec{u}_i) = \sum_{j=1}^{i-1} \lambda_j(\vec{u}_i) z^{(l)}(\vec{u}_j) + \sigma_E(\vec{u}_i) U_i, \quad \forall\, i = 1, \tag{2.1}$$

where λ_j are the kriging weights, σ_E^2 is the kriging variance error, and U_i is a standard Gaussian random variable. When the simulation starts with some known initial values $\{z(\vec{u}_1),\ldots,z(\vec{u}_{n_0})\}$, commonly referred to as hard data, Eq. (2.1) becomes

$$z^{(l)}(\vec{u}_i) = \sum_{k=1}^{n_0} \lambda_k(\vec{u}_i) z(\vec{u}_k) + \sum_{j=n_0+1}^{i-1} \lambda_j(\vec{u}_i) z^{(l)}(\vec{u}_j) + \sigma_E(\vec{u}_i) U_i, \quad \forall\, i = n_0+1,\ldots,N. \tag{2.2}$$

Since SGS is typically used on grids, our study is limited to gridded coordinate systems and, correspondingly, the terminology “node”, rather that “point” is employed. In this study, the term “estimation” was favor in the context of kriging as opposed to “prediction”, because of the temporal connotation of the latter (Chilès and Delfiner, 1999).

2.2.2 Computational Efficiency, Limited Neighborhood and Bias

A major drawback of SGS is that, in order to be mathematically rigorous, it needs to include all previously simulated nodes in the kriging estimation. As the simulation goes on, the number of previously simulated nodes grows, which in turn increases the size of the covariance matrix. Because the computational complexity of solving a kriging system of n nodes with the widely used LU, QR, or Cholesky solvers is $O(n^3)$ (Trefethen and Bau III, 1997), the simulation of a grid with N nodes has a complexity of $O(N^4)$ (Dimitrakopoulos and Luo, 2004; Srinivasan et al., 2008). The corresponding computational cost is excessive for typical applications.

To alleviate this problem, it is common practice to consider only a small number of representative neighbors in the estimation of each simulated node. This is generally referred to as the moving or limited neighborhood approach, as opposed to the unique or full neighborhood approach when all nodes are considered. Using a maximum of n neighbors reduces the overall computational complexity to $O(N^3 n)$ (Dimitrakopoulos and Luo, 2004). The retained neighbors are typically chosen based on their proximity to the estimated node, taking into account the correlation range and orientation. Indeed, due to the screening effect (Omre et al., 1993; Chilès and Delfiner, 1999), many of the distant nodes have a kriging weight close to zero and their absence implies only a small approximation.

Although inevitable in practice, the use of a limited neighborhood creates artifacts in the simulated fields (Meyer, 2004). By omitting certain nodes, not only the correlation of these nodes with the simulated node is neglected, but also the correlation between the omitted and retained nodes. This is reflected by a bias in the simulation covariance matrix and leads to errors in the simulation of the nodes. As SGS re-uses previously simulated nodes as

neighbors, the error is propagated and cumulated, thus creating significant artifacts in the final realizations.

In this paper, “error” is used as a generic term for any deviation from the true value, “artifact” is used for the resulting perceived error of a process, and “bias” is used for a systematic error, measured as the difference of the expected value of a variable to its true value. More specifically, in the context of this study, errors in SGS are due to biases in the sampling of each variable and result in tangible artifacts in the final realizations.

2.2.3 Simulation Path

The decomposition of the joint probability into conditional probabilities used in SGS does not assume any specific order in which the nodes are simulated. SGS with a unique neighborhood is therefore independent of the simulation path (Goovaerts, 1997) so that any type of path can be used without creating artifacts. However, with a limited neighborhood, the simulation path determines which nodes are available as neighbors. Hence, the simulation path influences indirectly the location and magnitude of the bias, resulting in the covariance matrix not being completely honored in the realizations. By extension, the path also defines which of these nodes are not taken into account in the neighborhood, and therefore which covariances are not correctly reproduced. Moreover, in determining which nodes are used for conditioning, the path can also influence the effect of cumulative biases, where nodes simulated with a bias are used as conditioning, thus propagating their error to the newly simulated nodes. Consequently, while the limited neighborhood is the origin of bias, the simulation path can be considered as the vector spreading the bias across the simulation.

In this paper, six types of paths commonly used in SGS (Fig. 2.1) were explored and whose characteristics are discussed in the following.

(i) The row-by-row path visits consecutive adjacent nodes (Daly, 2005). Its advantage is to ensure a nearly constant structure of neighboring nodes, thus allowing to re-use the same kriging weights and saving computational time (Deutsch and Journel, 1992; McLennan, 2002). However, because of its strong characteristic patterns this path is

qualified as regular and leads to artifacts in the choosen direction of the path (Deutsch and Journel, 1992; Gómez-Hernández and Cassiraga, 1994). McLennan (2002) does, however, note that the suspicions regarding the origin of the bias are not sufficiently well documented to be conclusive.

(ii) The random path entirely shuffles the order in which nodes are simulated. It is commonly recommended in the literature, because it does not seem to assume any specific structure and correspondingly, is viewed as minimizing the creation of artifacts (Deutsch and Journel, 1992; Gómez-Hernández and Journel, 1993; Goovaerts, 1997).

(iii) The spiral path sorts the nodes according to their distance to hard data. This naturally results in a path spiralling away from the hard data. McLennan (2002) explains the use of this path by the perception that simulating nodes near hard data would honor their influence more faithfully so that a spiral path would improve the representativeness of hard data and prevent artifacts.

(iv) A multi-grid path refers to a nested grid system where the simulation starts at the coarsest scale and moves to progressively finer scales by re-using the values simulated at the previous scale. At each grid scale, the simulation path can follow any of the types described above. In our study, a random path was used at each scale. From the early days of sequential simulation, it has been recognized that operating directly on a fine grid may not adequately reproduce large-scale features (Deutsch and Journel, 1992; Gómez-Hernández and Journel, 1993). Correspondingly, Gómez-Hernández and Journel (1993) proposed to use a multi-grid approach by first simulating nodes located on a coarser grid and the remaining nodes on subsequently refined grids. Tran (1994) and McLennan (2002) provide empirical tests showing the improvement of covariance reproduction when the domain is simulated by subsequently refined multiple grids. Conversely, Emery (2004) argues that multi-grid paths only delay the introduction of bias and that the number of pairs of nodes with erroneous covariance is not reduced.

(v) The mid-point path consecutively simulates the node that has the largest distance to its closest neighbor (Fournier et al., 1982; Barnsley et al., 1988; Emery and Peláez, 2011). For unconditional simulations, this path is similar to the multi-grid path if the first simulated node is located in one of the corners, with the exception of preferentially simulating the nodes located in the diagonal of the previous grid level, resulting in the

so-called diamond-square algorithm (Fournier et al., 1982). Emery and Peláez (2011) find a better reproduction of the second-order statistics for a mid-point path compared to the row-by-row or spiral paths. Omre et al. (1993) use a mid-node path to minimize simulation errors by allowing to use only the closest conditioning node in all directions. This will be discussed in more detail in the context of the screening effect (Sec. 2.4) and the Markov property (Sec. 2.5.1.1).

(vi) The quasi-random path is an alternative to the random path where the grid is visited more homogeneously (Chilès and Delfiner, 1999). It is rarely used and thus relatively unknown in SGS. Our implementation of this path relies on transforming a two-dimensional Halton sequence (Halton, 1960; Kocis and Whiten, 1997) into coordinates of the considered grid by using an acceptance/rejection method to avoid re-visiting nodes.

Amongst the paths described above, the row-by-row and spiral paths can be considered as deterministic since the same path will be used from one realization to another. The realizations produced are nevertheless not deterministic because the value of each simulated node is drawn using a different seed number. All other paths can be randomized, that is, a different specific path of the same path type is used for each realization. Note that here the term random path refers to the fact that the order in which the nodes are simulated is random while the term randomized path refers to each realization being generated with a different path. Randomizing the random, multi-grid and quasi-random path types is possible by changing the seed number, while the mid-point path requires the randomization of the selection of nodes which have an equal distance to their closest neighbors.

2.3 Quantification of Bias

As the underlying random field Z is assumed to be Gaussian, the evaluation of simulation can be based solely on reproducing its first- and second-order statistics defined by the mean and the covariance (Leuangthong et al., 2004). The former can be assessed based on the difference between the expected value of the simulation $\mathrm{E}[Z]$ and the kriging estimate. The

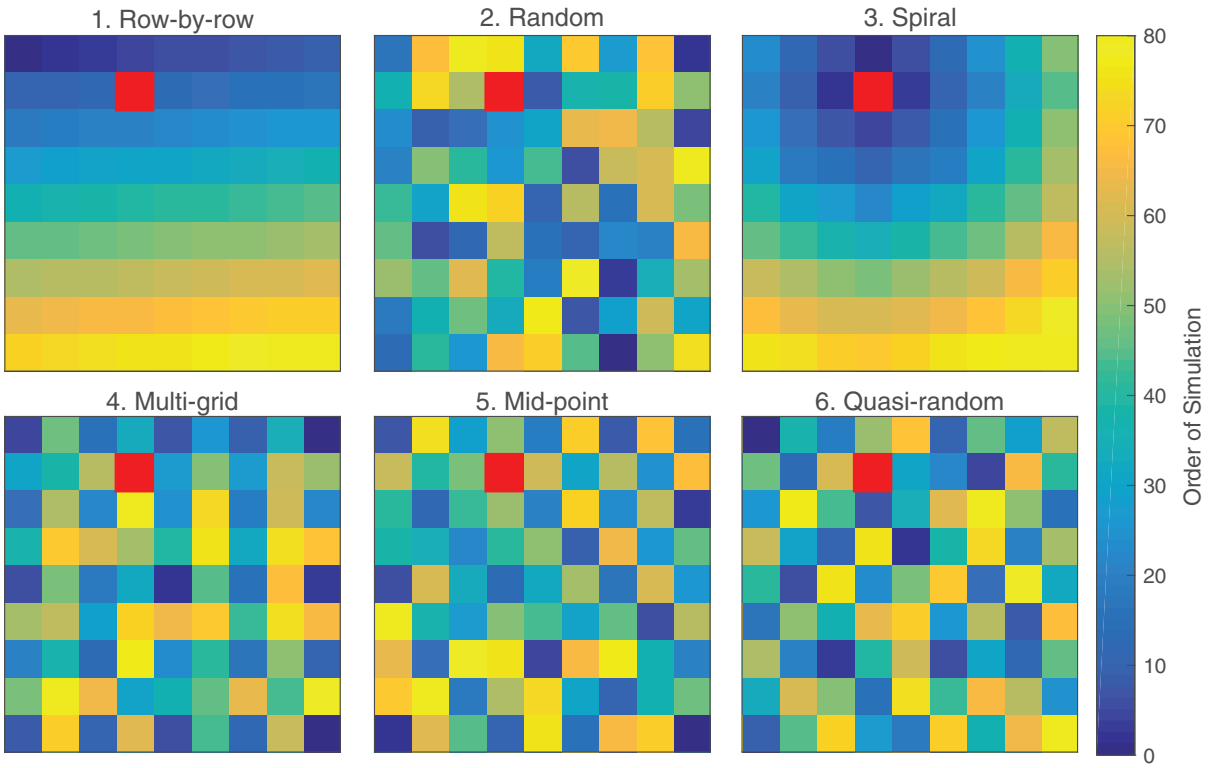


Figure 2.1 – Illustration of the 6 types of paths considered in this study on a 9x9 grid with one hard datum (red square). The color denotes the order in which the nodes are simulated. Nodes with lower values are simulated earlier than nodes with higher values

latter is fully defined by the discrepancy between the simulation covariance matrix $\vec{C}_{Z^{(l)}}$ and the model covariance matrix $\vec{C}_Z$.

Emery and Peláez (2011) present an elegant way to theoretically calculate the simulation covariance matrix from the kriging weights and the kriging errors variance. The strength of this approach is that it computes the exact covariance error for every single pair of nodes. In comparison, traditional covariance-function-based empirical assessment techniques (Emery, 2004; Leuangthong et al., 2004; Safikhani et al., 2017) struggle to account for the inherent fluctuation of the node statistics and perform an averaging of the covariance errors, which in turn results in a smoothing of the covariance.

2.3.1 Computation of the Simulation Covariance Matrix for Unconditional Simulations

Following Emery and Peláez (2011), SGS (Eq. (2.1)) can be re-written in matrix form to isolate $\vec{U}$ on the left-hand side and to combine the previously and currently simulated values on the right-hand side

$$\vec{U} = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ -\frac{\lambda_1^{n-1}}{\sigma_{n-1}} & \cdots & \frac{1}{\sigma_{n-1}} & 0 \\ -\frac{\lambda_1^{n}}{\sigma_{n}} & \cdots & -\frac{\lambda_{n-1}^{n}}{\sigma_{n}} & \frac{1}{\sigma_{n}} \end{bmatrix} \vec{z}^{(l)} = \vec{\Lambda}\vec{z}^{(l)}, \tag{2.3}$$

where $\lambda_j^i = \lambda_j(\vec{u}_i)$ and $\sigma_i = \sigma_E(\vec{u}_i)$. $\vec{\Lambda}$ is referred to as the lambda matrix and is built exclusively with kriging weights and variances.

If $E\left[\vec{z}^{(l)}\right] = 0$, the simulation covariance matrix can be computed based only on $\vec{\Lambda}$, as $\vec{U}$ has independent values, and, therefore, its covariance matrix is equal to the identity matrix

$$
\begin{aligned}
\vec{C}_{Z^{(l)}} &= E\left[\left(\vec{z}^{(l)} - E\left[\vec{z}^{(l)}\right]\right)\left(\vec{z}^{(l)} - E\left[\vec{z}^{(l)}\right]\right)^T\right] \\
&= E\left[\vec{z}^{(l)}\left(\vec{z}^{(l)}\right)^T\right] \\
&= E\left[\left(\vec{\Lambda}^{-1}\vec{U}\right)\left(\vec{\Lambda}^{-1}\vec{U}\right)^T\right] \\
&= \vec{\Lambda}^{-1}E\left[\vec{U}\vec{U}^T\right]\left(\vec{\Lambda}^{-1}\right)^T \\
&= \vec{\Lambda}^{-1}\left(\vec{\Lambda}^{-1}\right)^T.
\end{aligned}
\tag{2.4}
$$

In practice, $\vec{\Lambda}$ is constructed line-by-line according to the simulation path. This results in a sparse lower-triangular square matrix, because the computed weights correspond only to the previously simulated nodes.

2.3.2 Computation of the Simulation Covariance Matrix for Conditional Simulations

SGS with hard data (Eq. (2.2)) can be similarly rewritten by separating the weight of hard data $\vec{\Lambda}_0(n \times n_0)$ and the weight of the previously simulated data $\vec{\Lambda}\left((n-n_0)\times(n-n_0)\right)$

$$
\vec{U} = \begin{bmatrix} -\frac{\lambda_1^{n_0+1}}{\sigma_{n_0+1}} & \cdots & -\frac{\lambda_{n_0}^{n_0+1}}{\sigma_{n_0+1}} \\ \vdots & \ddots & \vdots \\ -\frac{\lambda_1^{n}}{\sigma_{n}} & \cdots & -\frac{\lambda_{n_0}^{n}}{\sigma_{n}} \end{bmatrix} \vec{z}_0 + \begin{bmatrix} \frac{1}{\sigma_{n_0+1}} & 0 & \cdots & & 0 \\ -\frac{\lambda_{n_0+1}^{n_0+2}}{\sigma_{n_0+2}} & \frac{1}{\sigma_{n_0+2}} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ -\frac{\lambda_{n_0+1}^{n-1}}{\sigma_{n-1}} & \cdots & -\frac{\lambda_{n-2}^{n-1}}{\sigma_{n-1}} & \frac{1}{\sigma_{n-1}} & 0 \\ -\frac{\lambda_{n_0+1}^{n}}{\sigma_{n}} & \cdots & & -\frac{\lambda_{n-1}^{n}}{\sigma_{n}} & \frac{1}{\sigma_{n}} \end{bmatrix} \vec{z}^{(l)}. \tag{2.5}
$$

The simulated values can then be isolated

$$\vec{z}^{(l)} = \vec{\Lambda}^{-1}\left(\vec{U} - \vec{\Lambda}_0\vec{z}_0\right), \tag{2.6}$$

and incorporated into the conditional covariance

$$\begin{aligned}
\vec{C}_{Z^{(l)}|Z_0} &= E\left[\left(\vec{z}^{(l)} - E[\vec{z}^{(l)} \mid \vec{z}_0]\right)\left(\vec{z}^{(l)} - E[\vec{z}^{(l)} \mid \vec{z}_0]\right)^T \mid \vec{z}_0\right] \\
&= E\left[\vec{z}^{(l)}\left(\vec{z}^{(l)}\right)^T \mid \vec{z}_0\right] - E\left[\vec{z}^{(l)} \mid \vec{z}_0\right] E\left[\left(\vec{z}^{(l)}\right)^T \mid \vec{z}_0\right] \\
&= \vec{\Lambda}^{-1}\left(E\left[\vec{U}\vec{U}^T + \left(\vec{\Lambda}_0\vec{z}_0\right)\left(\vec{\Lambda}_0\vec{z}_0\right)^T - 2\vec{U}\vec{\Lambda}_0\vec{z}_0 \mid \vec{z}_0\right]\right)\left(\vec{\Lambda}^{-1}\right)^T \\
&- E\left[\vec{\Lambda}^{-1}\vec{U} - \vec{\Lambda}^{-1}\vec{\Lambda}_0\vec{z}_0 \mid \vec{z}_0\right] E\left[\vec{U}^T\left(\vec{\Lambda}^{-1}\right)^T - \vec{z}_0^T\vec{\Lambda}_0^T\left(\vec{\Lambda}^{-1}\right)^T \mid \vec{z}_0\right] \\
&= \vec{\Lambda}^{-1}\left(\vec{I} + \vec{\Lambda}_0\vec{z}_0\vec{z}_0^T\vec{\Lambda}_0^T - 0\right)\left(\vec{\Lambda}^{-1}\right)^T - \vec{\Lambda}^{-1}\vec{\Lambda}_0\vec{z}_0\vec{z}_0^T\vec{\Lambda}_0^T\left(\vec{\Lambda}^{-1}\right)^T \\
&= \vec{\Lambda}^{-1}\left(\vec{\Lambda}^{-1}\right)^T,
\end{aligned} \tag{2.7}$$

which turns out to be identical to the unconditional case (Eq. (2.4)).

As opposed to the unconditional case, the expected value of the simulated data $\mathrm{E}\left[\vec{z}^{(l)}\right]$ is not always equal to zero in conditional simulations

$$E[\vec{z}^{(l)} \mid \vec{z}_0] = \vec{\Lambda}^{-1}E\left[\vec{U} - \vec{\Lambda}_0\vec{Z}_0 \mid \vec{z}_0\right] = -\vec{\Lambda}^{-1}\vec{\Lambda}_0\vec{z}_0. \tag{2.8}$$

2.3.3 Illustration

In order to demonstrate the exactness of the evaluation of the covariance matrix, 500 realizations of a one-dimensional grid of 257 nodes were simulated using a spherical covariance function with a range of 10. Figure 2.2 illustrates that the empirical covariance function

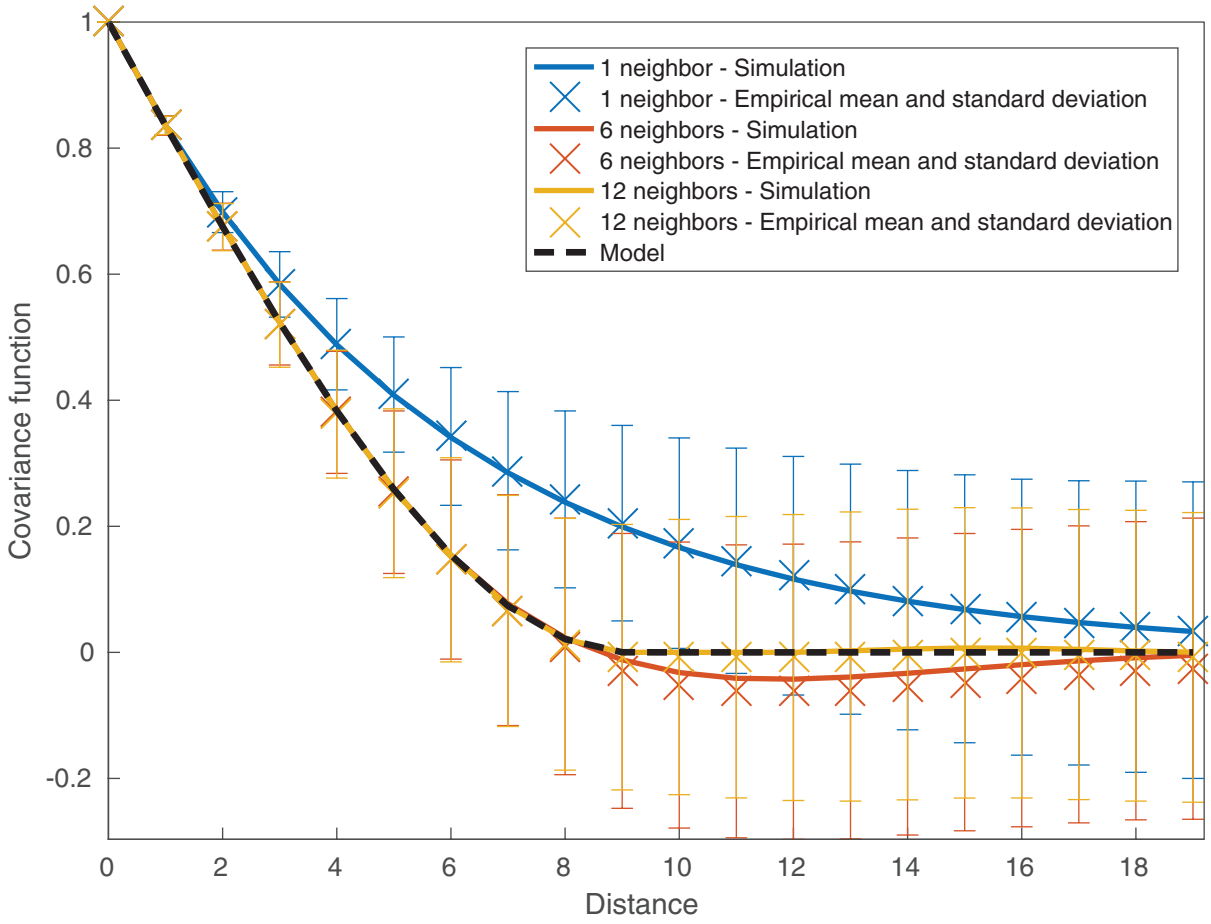


Figure 2.2 – Comparison of covariance function computed with the simulation covariance matrix (solid lines) and the empirical covariance function based on 500 realizations (crosses and error bars). The simulations were performed on a one-dimensional grid of 257 nodes using a spherical covariance function of range 10 and 3 different neighborhood sizes (1, 6, and 12 nodes). The dashed line denotes the model covariance function

agrees with the simulation covariance function. Both diverge from the model covariance function because of the use of a limited neighborhood. The calculation of the simulation covariance matrix thus proves to be accurate.

2.4 Characteristics of Optimal Paths

Based on the discussion on the origin of the bias in Sec. 2.2.2 and 2.2.3, the best path can be adequately characterized as the one minimizing the influence of the approximation resulting from the use of a limited neighborhood. However, this influence is complex and indirect,

making it difficult to narrow down a single best simulation path. Here, the analysis of the link between path and limited neighborhood was performed to infer some guidelines as to how to choose the optimal path.

Two levels of analysis are identified and described separately: the node level describing which is the optimal node to simulate in a given configuration of neighbors and, the simulation level characterizing which sequence of nodes minimizes the overall error.

In the following analysis, it is assumed that the neighborhood search strategy is operating by selecting a fixed number of nodes that are closest to the simulated node, regardless of the correlation range. The effects described in the following apply to most traditional covariance models, but may fail for some less common ones such as, for example, the pure nugget model.

At the level of node simulation, minimum bias is achieved by choosing to simulate the node whose conditional probability distribution is the least affected by the absence of neighbors due to the limited neighborhood. This means that the optimal node to simulate is the one whose neighborhood search excludes only unimportant nodes. The importance of a neighbor regarding the simulation of a node is measured by both its kriging weight and the modification of the kriging weights of other neighbors resulting from its removal from the neighborhood. Figures 2.3(a) through 2.3(e) illustrate five effects influencing the selection of the optimal node to simulate with a simple example. The simulation grids are one-dimensional for a-d and two-dimensional for e. For each example, two cases A and B simulating a different node are compared. Cases A are generally leading to smaller error than cases B. The simulated node is denoted as a red circle and is linked with red lines to its neighbors. The remaining black nodes are therefore neglected neighbors of the simulation. Dashed lines indicate that two neighbors are equidistant. The correlation range is 5 nodes and the neighborhood size is a unique node for a-c and two nodes for d-e.

(i) Decreasing distance effect: The importance of a node decreases as the distance to the simulated node increases. Therefore, the excluded nodes should be as far as possible from the simulated node (Fig. 2.3(a)).

(ii) Finite range effect: A node whose distance to both the simulated node and all neighboring nodes is larger or equal than the correlation range can be safely ignored. Therefore, such nodes can be excluded (Fig. 2.3(b)).

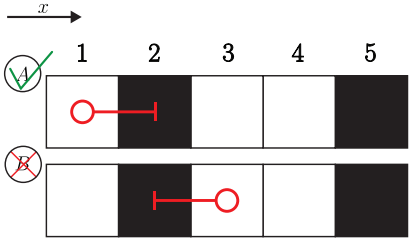
(iii) Relay effect (Rivoirard, 1984; Chilès and Delfiner, 1999): A node whose distance to the simulated node is equal to or greater than the range may not be removed without altering the simulation in the presence of an intermediate node correlated to both this node and the simulated node. This effect has various consequences depending on the neighborhood search strategy and the covariance model (Fig. 2.3(c)).

(iv) Screening effect (Rivoirard, 1984; Deutsch and Journel, 1992; Chilès and Delfiner, 1999): The information content of a node is reduced by the presence of an intermediate node. Therefore, it is better to exclude a node behind another one than to exclude an isolated node (Fig. 2.3(d)).

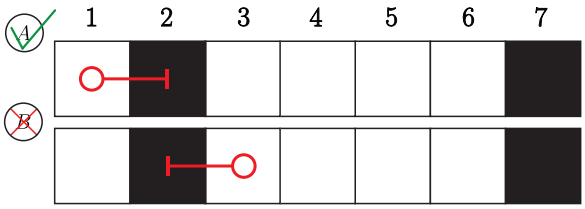
(v) Declustering effect: selecting two nearby nodes provides less information than selecting two isolated nodes. Therefore, the simulated node including neighbors from various directions is more favorable than clustered neighbors (Fig. 2.3(e)).

At the level of the global simulation, the optimal path is not necessarily the combination of all locally optimal nodes. Indeed, choosing the optimal node to simulate at any specific moment may lead to larger errors later in the simulation. Instead, the optimal path should take into account the fact that a simulated node subsequently becomes a conditioning node. In addition, the effect of cumulative bias can become significant at this level and, therefore, the optimal path should be built to avoid that a node with a large bias becomes a conditioning node. Moreover, the comparison of paths relies on a multi-dimensional error analysis because each path leads to different errors for each node. Therefore, the comparisons of different paths also depends on the way these errors are aggregated.

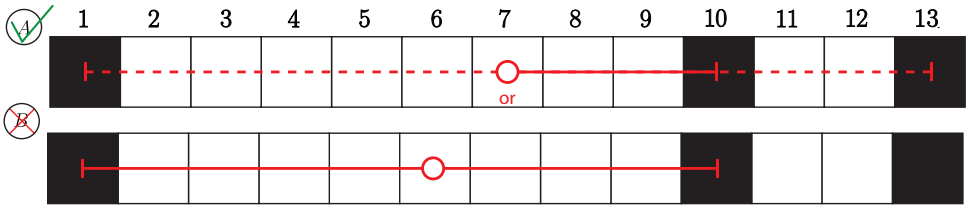
At both the node and the simulation levels, the neighborhood size and the shape of the covariance function model have a strong influence on the optimal path. For instance, a smaller neighborhood may result in excluding an influential node, thus changing the optimal node to choose. In addition, the boundary of the grid also influences the node selection, thus making the optimal path sensitive to the grid size.



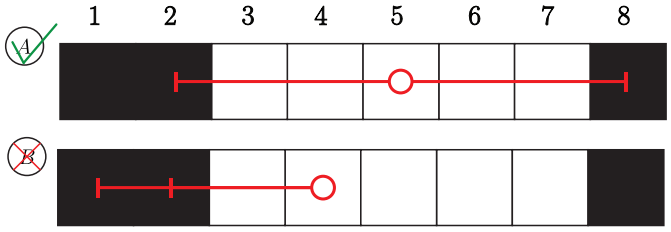
(a) Case A is optimal because the distance between the simulated node and the excluded neighbor is larger than for B: $d(x_1, x_5) > d(x_3, x_5)$



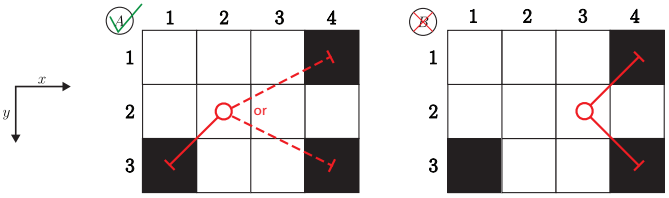
(b) In A, x_7 can be safely ignored because $\mathrm{d}(x_1, x_7) \geq r$ and $\mathrm{d}(x_2, x_7) \geq r$. This is not the case for B



(c) In B, although $\mathrm{d}(x_{13}, x_6) \geq r$, x_{13} should be included because of the relay effect with x_{10}: $\mathrm{d}(x_6, x_{10}) < r$ and $\mathrm{d}(x_{10}, x_{13}) < r$. Instead, with a traditional search neighborhood, x_1 would be included because it is closer to x_6 even though it has no influence on the simulated node $\mathrm{d}(x_1, x_6) \geq r$. In A, as both conditional nodes x_{13} and x_1 have the same distance to x_7, either of them could be selected as neighbors (dotted red line). Yet, including x_{13} would lead to an unbiased simulation while including x_1 does not



(d) For both A and B, the excluded node has the same distance to the simulated node $\mathrm{d}(x_5, x_1) = \mathrm{d}(x_5, x_8)$. However, because of the screening effect, the weight of x_1 in A is reduced by the presence of x_2 while in B the weight of x_8 is not reduced



(e) Because of the declustering effect, the weights of the two neighbors, (x_4, y_1) and (x_4, y_3) in B are lower than if only one of them was selected. In A, this effect is less pronounced as the two neighbors are further apart. Note that (x_4, y_1) brings more diverse information than (x_4, y_3) because it is further away from (x_1, y_3)

Figure 2.3 – Illustration of the 5 effects described above based on the comparaison of two cases A and B with a different simulated node denoted as red circles. Black squares correspond to conditioning nodes but only those linked to the simulated node with a red line are used in the simulation because of the limited neighborhood. The correlation range r is 5 nodes. The neighborhood size is one node in (a) and (b) and two nodes in (c), (d) and (e)

Despite the challenges of generalizing the design of an optimal path, the effects described above reveal one of its key characteristics: maximizing the distance between successively simulated nodes. Indeed, when the correlation between nodes is minimized, the relay, screening, and declustering effects are also minimized.

2.5 Assessement of Different Types of Paths

Based on the above description, two general families of paths can be identified: clustering and declustering paths. Row-by-row and spiral paths can be classified as clustering paths as they simulate consecutively the nearest nodes thus creating clusters. Conversely, the multi-grid, mid-node, and quasi-random paths tend to spread-out the simulated nodes and hence are referred to as declustering paths.

For the following analysis, six commonly used covariance functions have been selected to represent a diversity of structures: exponential, Gaussian, spherical, hyperbolic, k-Bessel, and cardinal sine. Their ranges were normalized to have similar integral values (Lantuéjoul, 2002). For the two-dimensional simulations, the neighborhood sizes (4, 8, 12, 20, 32, 52, 108) were chosen such that the neighbors are homogeneously distributed around the simulated node.

2.5.1 Clustering Paths

2.5.1.1 Row-by-row Path

For the sake of simplicity, the row-by-row path is first assessed on a one-dimensional grid. Thus, the neighborhoods are exclusively composed by of previously simulated nodes. This structure ensures a perfect propagation of the information of previously simulated nodes. Yet, because the neighborhood is only occupied by nodes whose distance to the simulated node is equal to or less than the neighborhood size, larger lag distances are never considered.

Consequently, the reproduction of larger lag distances of the simulation covariance matrix can only be approximated based on covariances of shorter lags.

Let us consider a theoretical example with four evenly spaced nodes simulated with a row-by-row path and a neighborhood size of one node, that is, using only the previously simulated node. The variance is normalized to 1 and the covariance of lag h is denoted by $c_h = 1 - \gamma(h)$. Thus, $\vec{\Lambda}$ becomes

$$\vec{\Lambda} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -\frac{c_1}{1-c_1^2} & \frac{1}{1-c_1^2} & 0 & 0 \\ 0 & -\frac{c_1}{1-c_1^2} & \frac{1}{1-c_1^2} & 0 \\ 0 & 0 & -\frac{c_1}{1-c_1^2} & \frac{1}{1-c_1^2} \end{pmatrix}, \tag{2.9}$$

and, using Eq. (2.4), the simulation covariance matrix can be derived as

$$\vec{C} = \begin{pmatrix} 1 & c_1 & c_1^2 & c_1^3 \\ c_1 & 1 & c_1 & c_1^2 \\ c_1^2 & c_1 & 1 & c_1 \\ c_1^3 & c_1^2 & c_1 & 1 \end{pmatrix}^2. \tag{2.10}$$

In this example, the simulation exactly reproduces the model covariance for lag distances smaller than the neighborhood size, that is, c_1, but approximates the values at larger lags by a power-law-type extrapolation of the covariance at lag 1, such as, for example, $c_2 \approx c_1^2$. The overall error of the simulation is the mismatch measured by $|c_1^h - c_h|$. This error will be different for each covariance function, depending on how well the covariance function can be approximated with this power law.

When the neighborhood size increases, the covariance of the simulation becomes more complicated to compute, but it always corresponds to a polynomial expression of the covariance at shorter lags. For example, with two neighbors, the covariance at lag 3 is approximated by

$$c_3 = \frac{c_1 + c_1 c_2^2 - 2c_1 c_2}{\left(c_1^2 - 1\right)^2}. \tag{2.11}$$

Using the symbolic toolbox in MATLAB, the mathematical expressions of a row-by-row simulation with a grid of 12 nodes were computed using neighborhood sizes between 1 and 6 nodes. Because the row-by-row path ensures that the neighborhood has the same configuration, the kriging weights are the same and the simulation covariance matrix has equal diagonal elements, that is, the covariance of all pairs of nodes with the same lag distance are identical (e.g., Eq. (2.10)). Because of this property, no averaging is required in the computation of the covariance function (Fig. 2.5).

In all cases, the covariances of lags shorter than the neighborhood size are perfectly reproduced. However, each covariance function behaves differently for longer-range lags. The exponential covariance function exhibits no error because of its Markov property, which implies that a simulated node depends only on its immediate neighbors (Omre et al., 1993; Chilès and Delfiner, 1999). The previously considered theoretical example confirms this property as the error $\|c_1^h - c_h\|$ vanishes for an exponential covariance function with $\left(e^{-1}\right)^h = e^{-h}$. This property has also been described as a perfect one-dimensional screening effect, where the first encountered node shields the influence of those behind it (Chilès and Delfiner, 1999). Autoregressive processes generalized this effect to covariance models composed of damped exponentials and damped sine waves (Box et al., 2008; Chilès and Delfiner, 1999). The unstable behavior associated with Gaussian and cardinal sine covariance functions can be explained by the difficulty of fitting a polynomial to their functions because of the presence of an inflection point. For the spherical, k-Bessel and hyperbolic covariance functions, the simulation covariance follows an exponential-type function fitted to the first n correct lag distance covariances.

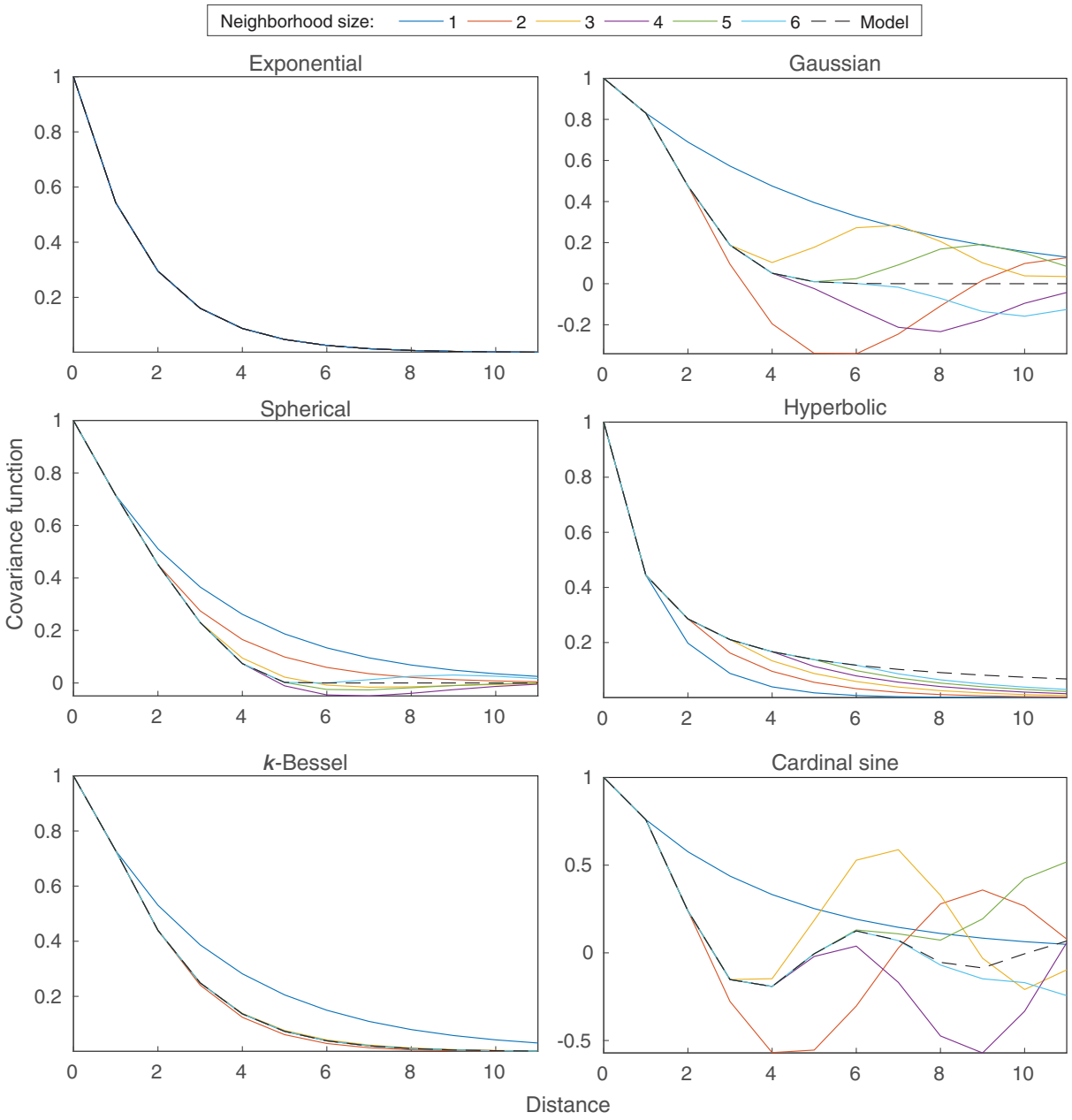


Figure 2.5 – Simulation covariance functions for a one-dimensional grid of 12 nodes with a row-by-row path for 6 different covariance functions and neighborhood sizes. The dashed line denotes the model covariance function

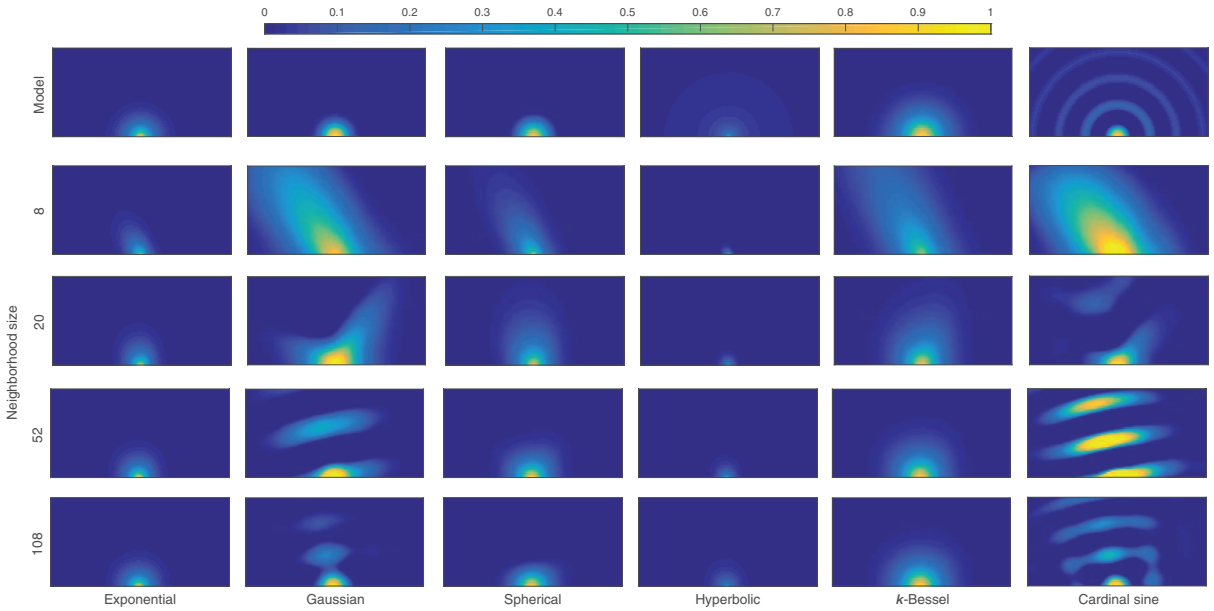


Figure 2.6 – Two-dimensional covariance functions for a grid of 64x64 nodes using a row-by-row path for different number of neighbors and covariance functions. Only the upper half of the covariance function map is displayed as the bottom half is a symmetric image of the upper half. The first row shows the model covariance function

The extension of these results to two dimensions is not trivial, because anisotropic effects occur. Figure 2.6 shows the two-dimensional covariance functions for a 64x64 grid with varying neighborhood size and covariance function. Because of the unilateral order of simulation, the bottom half of the two-dimensional covariance functions is a symmetric image of the upper half and thus not shown.

In two dimensions, errors appear in the diagonal of the covariance function due to the direction of the path. For a Gaussian covariance function, the row-by-row path creates very large sinusoidal artifacts similar to the one-dimensional case. For the spherical covariance function, the sign of the error is similar to the one-dimensional case with an overestimation of the covariance for small neighborhoods and the largest errors in the vicinity of the range. For the exponential covariance function, the perfect screening effect described in one dimension does not hold in two dimensions and, consequently, errors arise. Chilès and Delfiner (1999) showed that a covariance function that can be factorized along its components, such as, for instance, the k-Bessel covariance function has a perfect two-dimensional screening effect. However, in order to generate a perfect simulation, all neighborhoods have to form a closed contour around the simulated node. The neighboring search strategy to build such a closed

contour is complex and, hence, has not been implemented here. This explains the small but persistent errors encountered with the k-Bessel covariance function. Overall, exponential and k-Bessel covariance functions remain the most suitable models for the row-by-row path. The reader is referred to Boulanger (1990) for the extension of autoregressive processes in two dimensions.

2.5.1.2 Spiral Path

A common misconception associated with this path is its assumption that the most important nodes are the ones closest to the hard data. Thus, the rule that the most important area has to be simulated first (Gómez-Hernández and Journel, 1993) causes the path to spiral away from hard data.

This path ensures a perfect reproduction of the influence of hard data on the close-by nodes because the hard data will always be included in their neighborhoods. However, as soon as there are too many nodes in the neighborhood, the hard data will be among the first to be eliminated in the kriging system, and its influence will be approximated by intermediates nodes similarly to the row-by-row path. As a result, the spiral path is particularly problematic, because it minimizes the presence of hard data in the neighborhood of simulated nodes.

To illustrate this, a 65x65 grid containing a single hard datum with a value of 1 centered in the middle of the grid was simulated using a spiral path. The expected value is symmetric around the hard datum and thus a one-dimensional cross-section of the expected value is sufficient to adequately illustrate the resulting artifacts. Figure 2.7 shows the expected values $\mathrm{E}[\vec{z}^{(l)} \mid \vec{z}_0]$ along a one-dimensional section starting in the middle of the grid, such that the nodes are simulated from left to right. The expected values of the first simulated nodes match the kriging estimation perfectly, but once the hard datum is no longer included in the neighborhood, the expected value of the simulated nodes shows that the information of the hard datum is propagated by a polynomial that differs from the kriging estimator. These errors exhibit a similar structure as the covariance function errors generated by the row-by-row path. Indeed, the field of expected values of a simulation with a single hard datum with a value of one is exactly equal to the covariance function of the distance to the hard datum. Similar to the row-by-row path, certain covariance functions are well approximated

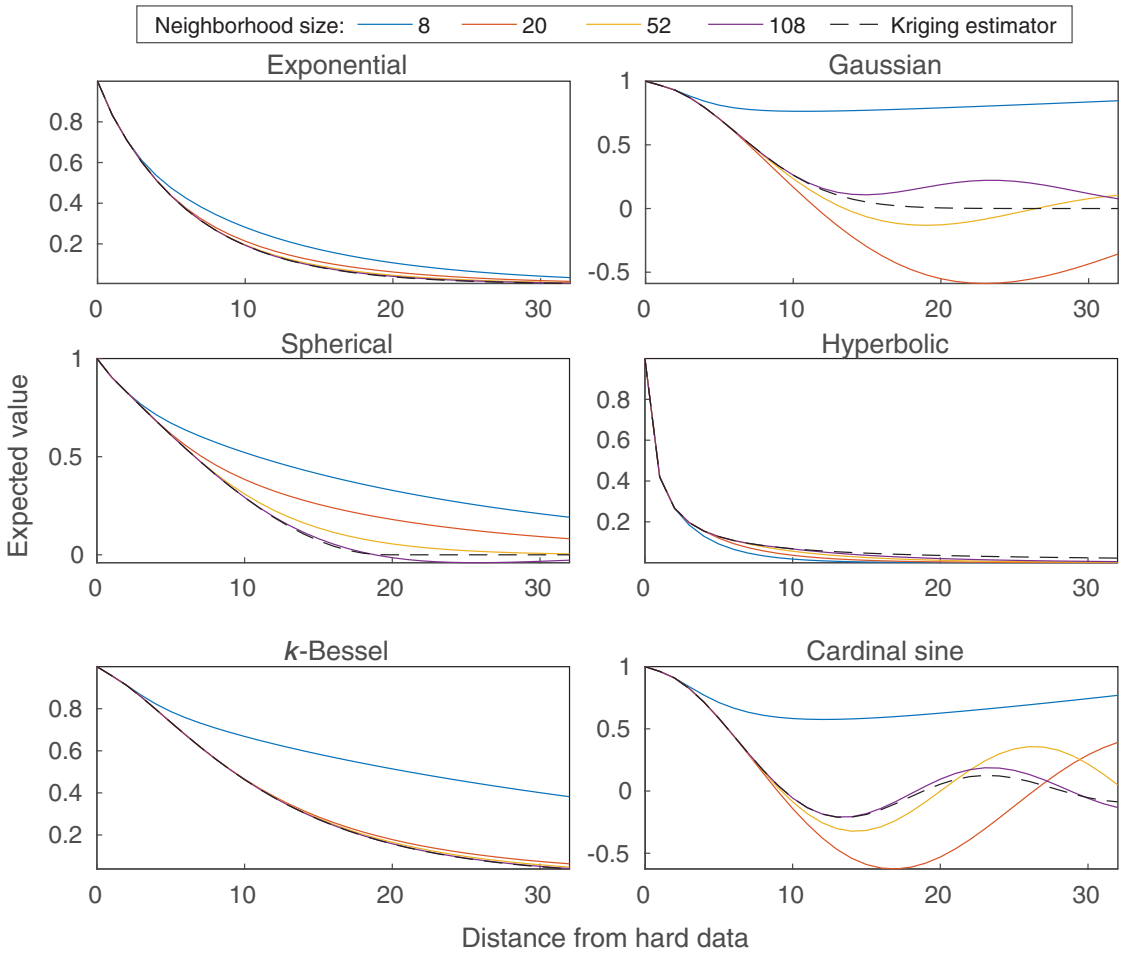


Figure 2.7 – Cross-section of the expected values starting from the center of the 65x65 grid where a single hard datum with value of 1 is present. The black dotted line is the kriging estimation and represents the target for the expected value of the simulation

by this extrapolation (exponential, hyperbolic, k-Bessel), while others are poorly reproduced (Gaussian, spherical, cardinal sine).

When more conditioning data are available, the spiral path simulates the nodes in clusters around each hard datum. Figure 2.8 shows a section of the expected value of simulations with four hard data with different values and positions (black crosses). The random path is used as a benchmark.

This example demonstrates the problem of cluster merging associated with the spiral path. At first, each cluster evolves independently, propagating the information of each hard datum according to the covariance function approximation. As the clusters are growing, at a certain moment in the simulation, the neighborhood of one cluster will comprise a node of another

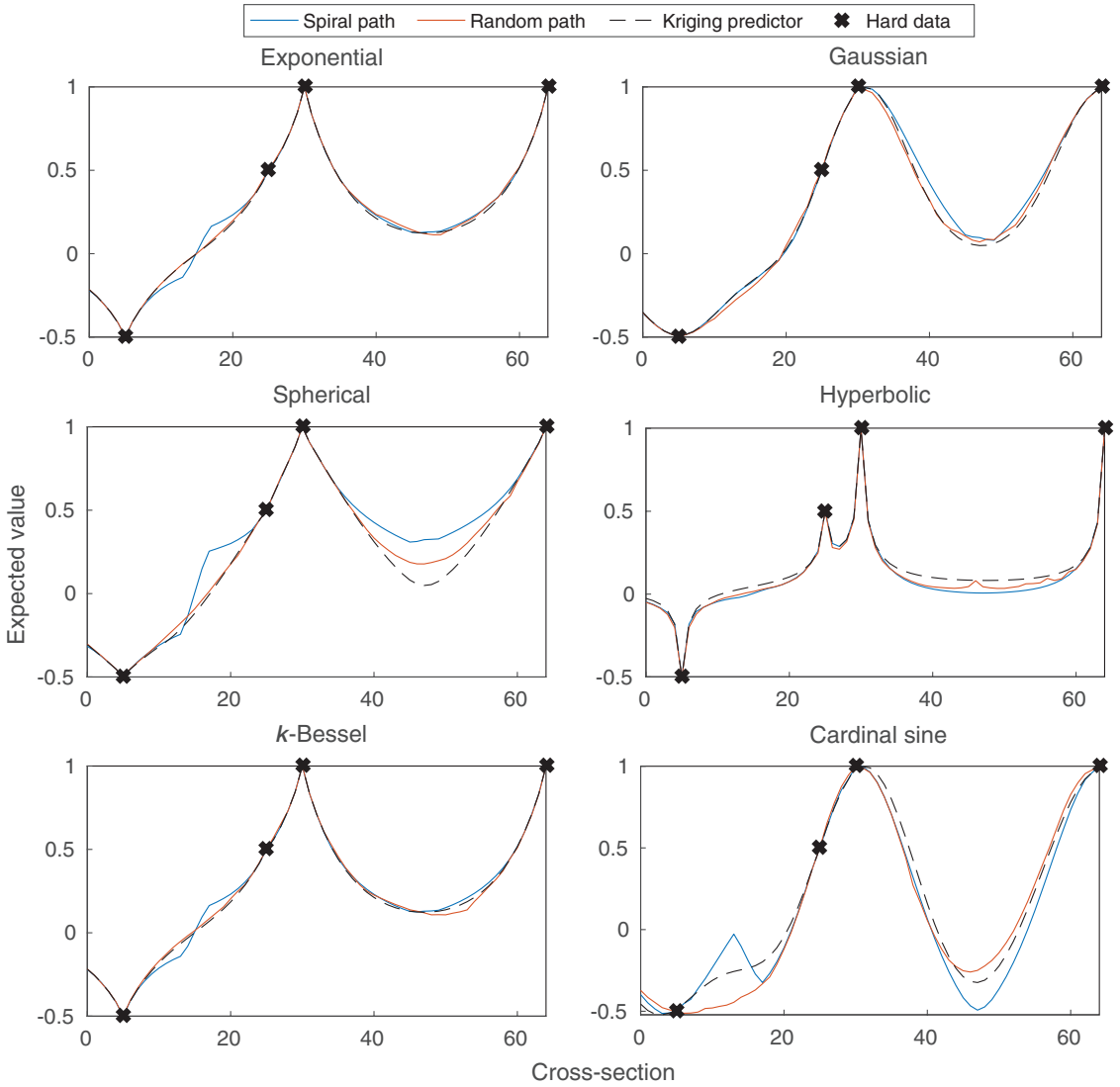


Figure 2.8 – Comparison of one-dimensional sections of the expected value of a simulation on a 65x65 grid with four hard data positioned along this section for spiral and random paths and different types of covariance functions

cluster, which results in the merging of the two clusters. However, as this happens late in the simulation and because each cluster evolved independently, this merging is abrupt and hence creates large discontinuities in the realizations. Conversely, the random path leads to a better reproduction of the expected value.

2.5.2 Declustering Paths

Unlike clustering paths, declustering paths favor the simulation of nodes that are far from their neighbors and thus avoid the creation of clusters. Since the three declustering paths, multi-grid, mid-point, quasi-random, are similar, they will be analyzed together.

Declustering paths allow for having neighborhoods composed of more spread-out nodes, which has two positive effects with regard to minimization of the error. First, inside the limited neighborhood, each selected node is representative of a different area of the grid, thus providing diverse and moderately correlated information. Secondly, because the correlations among the nodes inside and outside the neighborhood are weak, the removal of the nodes outside the neighborhood has little influence on the simulated node.

The error at the level of the node simulation can be measured with the relative screen effect approximation loss or RSEA (Dimitrakopoulos and Luo, 2004), which corresponds to the normalized mean-square error difference of the simulated value

$$RSEA(\vec{u}_i) = \frac{1}{\mathrm{Var}\{Z^{(l)}(\vec{u}_i)\}} \frac{1}{2} \mathrm{E}\left\{ \left[Z^{(l)}(\vec{u}_i) - Z(\vec{u}_i) \right]^2 \right\} = 1 - \sqrt{\frac{\mathrm{Var}\{Z(\vec{u}_i)\}}{\mathrm{Var}\{Z^{(l)}(\vec{u}_i)\}}}, \tag{2.12}$$

where $\mathrm{Var}\{Z(\vec{u}_i)\}$ and $\mathrm{Var}\{Z^{(l)}(\vec{u}_i)\}$ denote the kriging variance error σ_E^2 for a full and limited neighborhood, respectively. Readers are refered the corresponding paper for the demonstration.

Figure 2.9 compares the RSEA according to the simulation order of three path types for each of the six covariance functions considered in this study. For each path type and covariance model, 48 realizations were performed with a different randomized path. The percentiles P10, P50 (median) and P90 RSEA values for the nodes simulated at a given order of simulation are shown on Fig. 2.9. For all covariance functions, the very first nodes are simulated without any error because all neighbors are included. Then, all models behave differently in response to their various covariance function characteristics. The median value suggests that quasi-random and multi-grid paths perform only slightly better than the random path. However, the P90 shows that quasi-random and especially multi-grid paths minimize large RSEA values.

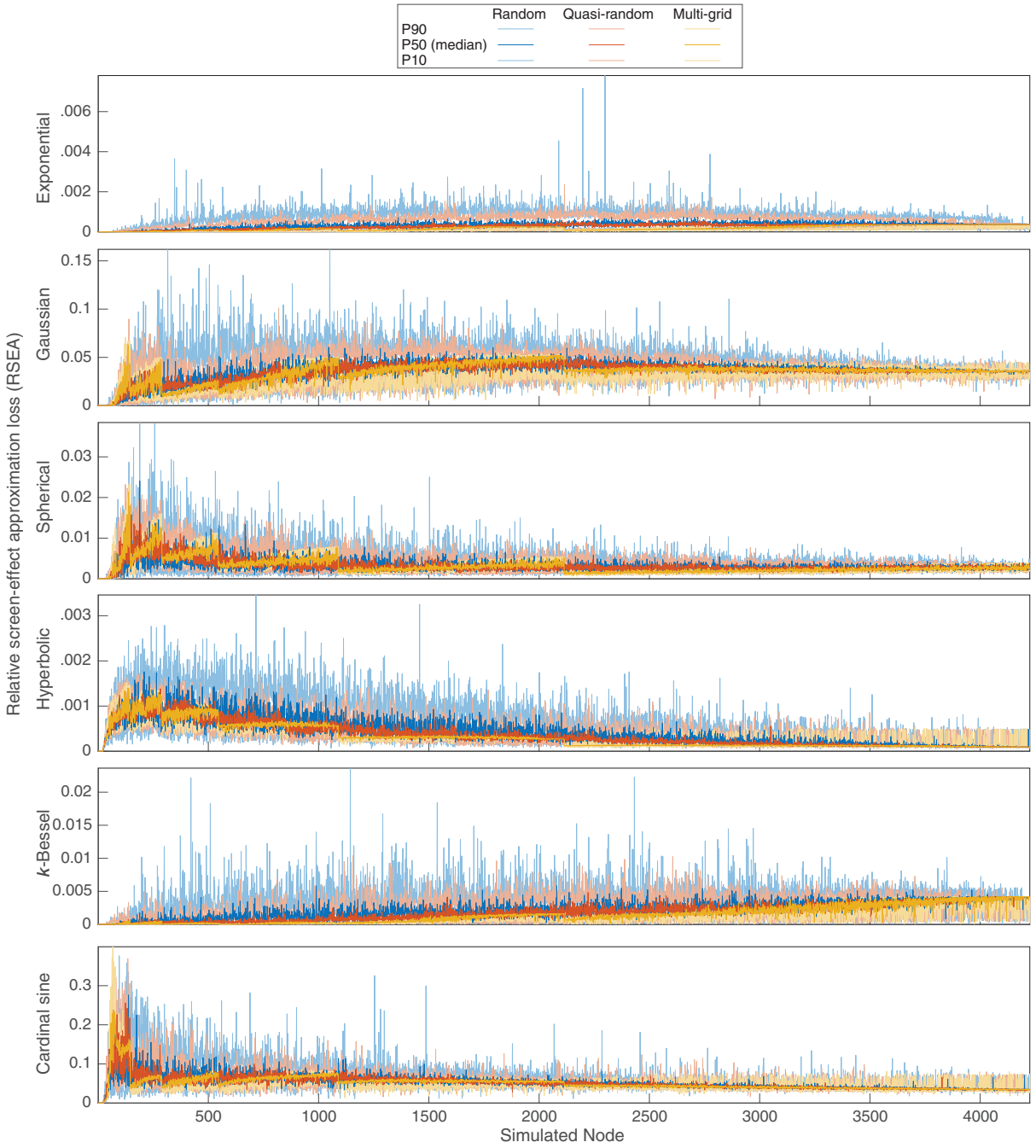


Figure 2.9 – P10, P50 and P90 RSEA (Eq. (2.12)) for a random (blue), a quasi-random (red), and multi-grid (yellow) paths for various types of covariance models as a function of the simulation order. The results are based on 48 realizations, each of which has been carried out with a different randomized path

Indeed, the main advantage of these paths is to avoid neighborhood configurations leading to large errors.

The error over the whole simulation is assessed by the mismatch between the simulation covariance matrix and the model covariance matrix. Following an approach suggested by Emery and Peláez (2011), the error matrix can be aggregated through the standardized Frobenius norm, which corresponds to a normalized root mean square error of the form

$$\eta = \frac{\|\vec{C}_{Z^{(l)}} - \vec{C}_Z\|}{\|\vec{C}_Z\|}, \tag{2.13}$$

where $\|\cdot\|$ denotes the Frobenius norm, also referred to as $L_{2,2}$.

Figure 2.10 compares the standardized Frobenius norm of simulations using random and declustering paths for different types of covariance functions and neighborhood sizes. Varying grid sizes and correlation ranges were also tested, but because theses parameters were not found to have a significant influence, only simulations for a 64x64 grid with a correlation range of 15 nodes are shown. The standardized Frobenius norm value is computed for 48 realizations for each path type and for each neighborhood size. Please note that each realization uses a different randomized path, either of the multi-grid, quasi-random or random path type. Figure 2.10 shows the averages of the 48 realizations as well as the their standard deviations as errorbars. These results demonstrate that the variation of the standardized Frobenius norm among realizations is much smaller compared to simulations with different path types. For all covariance functions, a significant error reduction is observed for quasi-random and multi-grid paths compared to random path. In a logarithmic scale, this error reduction remains constant with an increased neighborhood size. Figure 2.10 also illustrates the major influence of increasing the neighborhood size for reducing the error.

Figure 2.11 shows the covariance function errors for simulations with 20 neighbors and a range of 20. Because the covariance matrix contains several pairs of nodes with the same lag distance, the distribution of covariance errors is displayed for each lag distance. Each covariance function produces a different error structure, but these structures are not dependent on the path type. The neighborhood size has an important influence on the shape of these error structures. Declustering paths are associated with a major error reduction for lags between half to twice the range. This result confirms that multi-grid paths, and declustering paths in general, improve the covariance function reproduction for large lag distances.

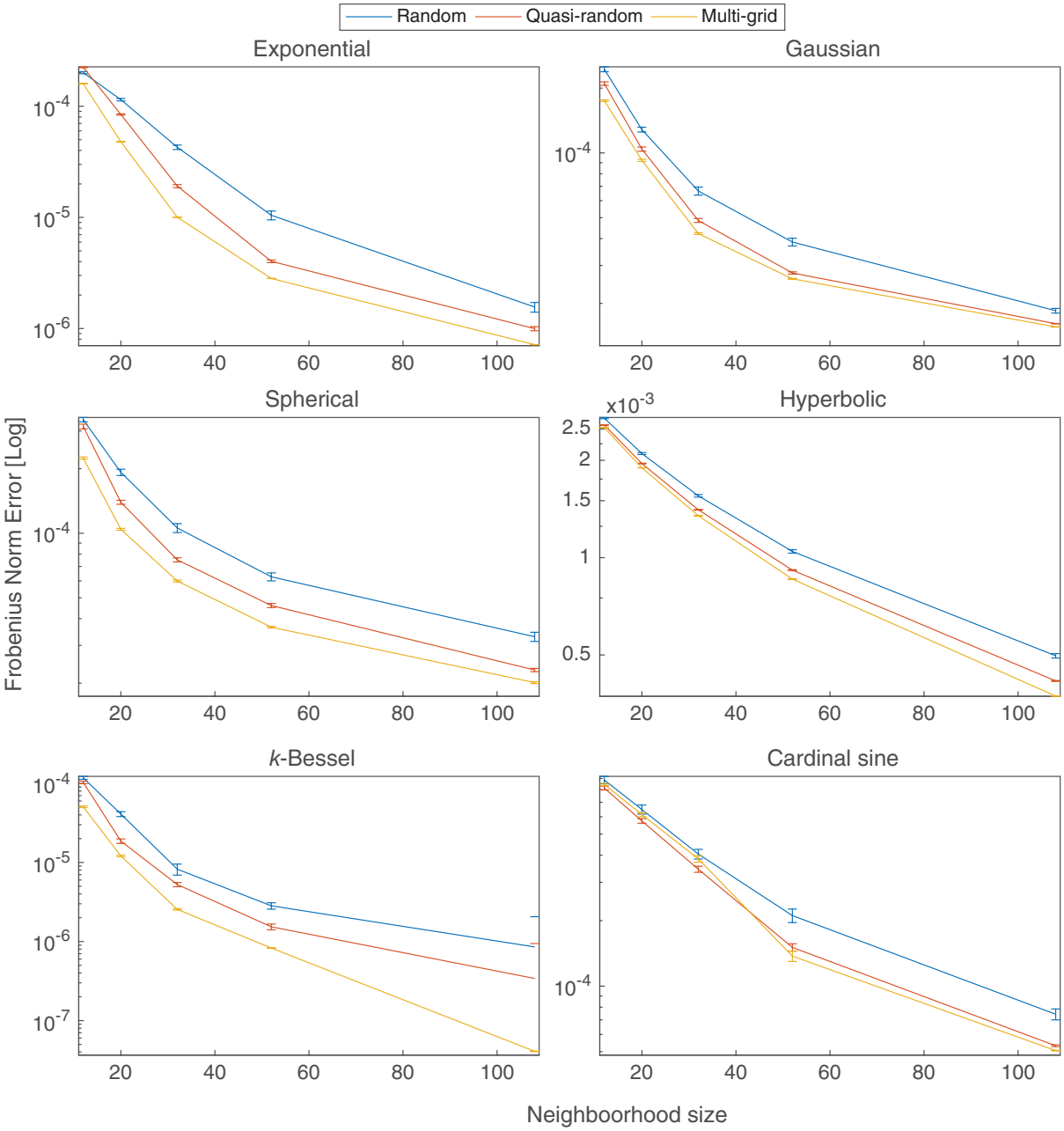


Figure 2.10 – Average Standardized Frobenius norm for simulations using random (blue), quasi-random (red) and multi-grid (yellow) paths for different covariance functions and neighborhood sizes. 48 realizations with a different randomized path were computed for each type of path and neighborhood size. The corresponding standard deviations of the standardized Frobenius norm error are displayed as errorbars

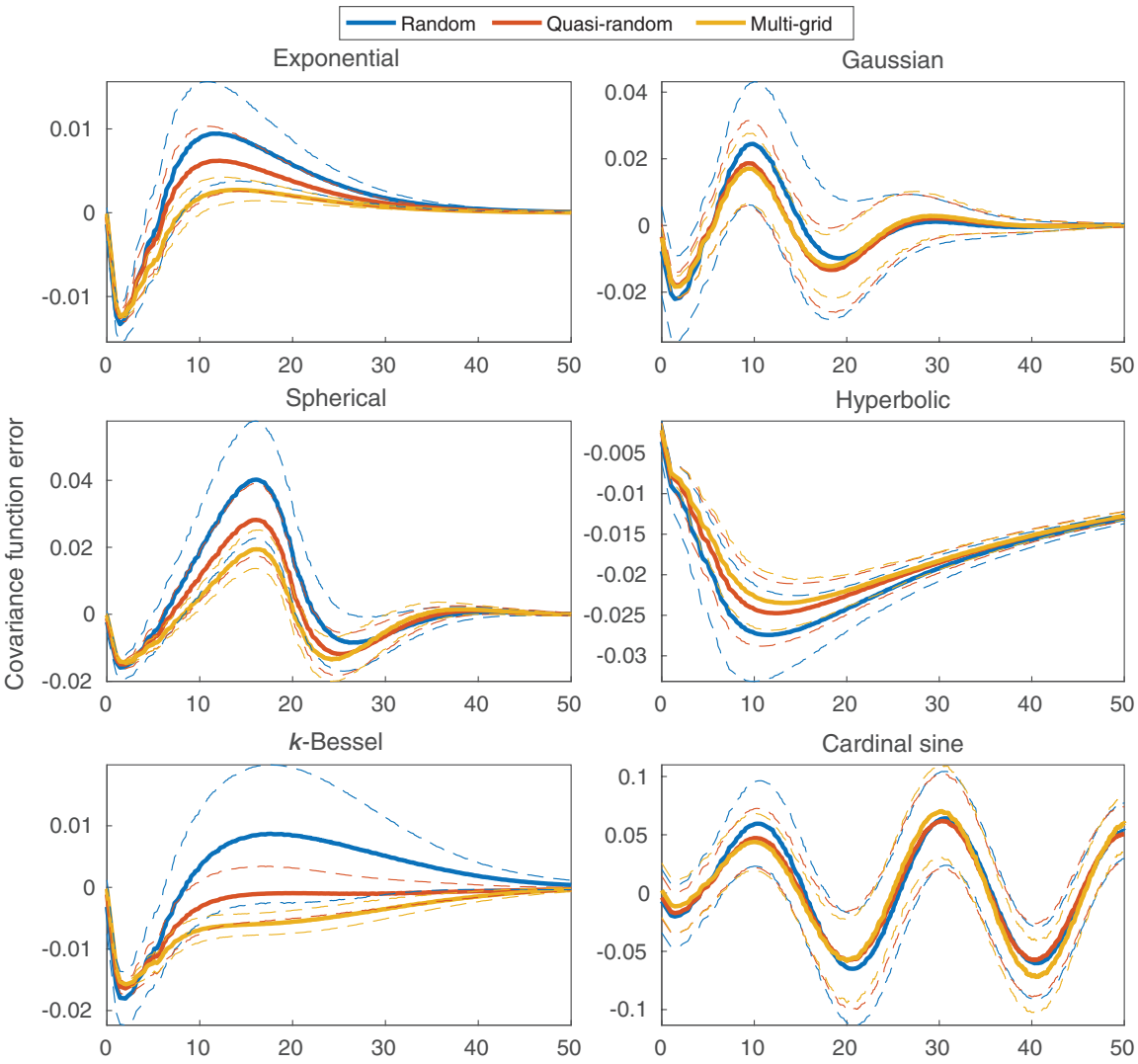


Figure 2.11 – Covariance function error for simulations using different types of paths and covariance functions. Solid lines denote the mean values and dotted lines the one standard deviation interval resulting from the aggregation of the covariance matrix with equal lag distance

2.6 Discussion

2.6.1 Conditional Simulations

Conditional simulations can be viewed as simulations where the paths start with the known nodes being perfectly simulated. When only few hard data are available, a multi-grid path is optimal as the simulation of the remaining nodes will be only mildly affected by the hard data. However, if many hard data are present, their positions will affect the optimal path at the simulation level and hence a multi-grid path might no longer be the best choice. Yet, declustering of the nodes remains the overall approach to follow, and in such a context, the mid-node path is a good alternative. It differs from the multi-grid approach, in that it iteratively selects the node that has the largest distance to its closest neighbor, thus taking into account the positions of the hard data.

It is common practice to attribute high confidence to hard data and request that their influence is correctly propagated throughout the simulation. Although this objectif is hard to address through the choice of the type of path, the neighborhood search method can be adapted to achieve it. Indeed, as demonstrated by Emery and Peláez (2011), if all hard data are imposed in the neighborhoods of all simulated nodes, the expected value of the simulated nodes is perfectly reproduced over the simulated domain. This kind of approach can be implemented with a two-part search strategy, which sorts and selects hard data separately from previously simulated nodes (Deutsch and Journel, 1992).

2.6.2 Search Method

Throughout this study, the selection of nodes was based only on their distance to the simulated node relative to correlation range and orientation. The multiple and interacting effects described in Sec. 2.4 suggest that distance is not always a sufficient criterion for determining the optimal neighborhood, as for example with the relay effect illustrated in Fig. 2.3(e). Generally, this situation only arises when more diverse information comes from a distant node.

Fortunately, this scenario can be avoided through the use of declustering paths. Yet, in the presence of clustered hard data, this effect cannot be ignored and the search method should be suitably adapted. The octant/quadrant search method (Isaaks and Srivastava, 1989) has been designed to balance the directional sampling of the neighborhood by separately searching each quadrant of the domain. This is a useful approach in the presence of clusters of hard data, while for unconditional simulations with a declustering path our simulations showed significant errors. This is consistent with the findings of Goovaerts (1997) and Deutsch and Journel (1992) who report that octant/quadrant search tends to create discontinuities in the absence of clusters, because of the drastic change of neighborhood from one node to the next.

2.7 Conclusion

In this study, a systematic analysis was presented to explore the link between the simulation path and artifacts in SGS realizations. The common practice of limiting the number of neighbors in the kriging estimation results in biases. In this respect, the role of the simulation path is to determine which nodes are included or omitted in the neighborhood and, thus, it ultimately decides which pairs of nodes will be incorrectly correlated and how large the associated errors will be . Exact simulation covariance matrices are computed to rigorously assess the relationship between biases in the simulations and parameters like grid size, neighborhood size, covariance function and range, and, most importantly, the different types of paths. This assessment of different simulation paths results in a classification into clustering paths, which simulate consecutively nearby nodes, and declustering paths, which maximize the distance between consecutively simulated nodes.

Clustering paths, such as the row-by-row and spiral paths, perfectly reproduce the covariances at short lag distances, but result in large, and often inadequate, approximations for larger lag covariances. This is due to the inherent structure of these paths, which results in the kriging neighborhood being exclusively constituted by nearby nodes. For the same reason, the spiral path poorly propagates the influence of hard data. Conversely, declustering paths,

such as multi-grid, mid-point or quasi-random paths, maintain a balanced and diverse neighborhood while minimizing the correlation between nodes within the neighborhood and beyond, thus resulting in minimum biases. Their overall path structure also minimizes cumulative bias by simulating less correlated nodes first, thus avoiding the re-use of nodes associated with large biases.

Chapter 3

Accelerating Sequential Gaussian Simulation with a Constant Path

---

Raphaël Nussbaumer, Grégoire Mariethoz, Mathieu Gravey, Erwan Gloagen, Klaus Holliger

Published[1] in *Computers & Geosciences.*

[1] Nussbaumer, R., Mariethoz, G., Gravey, M., Gloaguen, E., & Holliger, K. (2018). Accelerating Sequential Gaussian Simulation with a constant path. *Computers & Geosciences*, **113**, 121–132. DOI:10.1016/j.cageo.2017.12.006

Abstract

Sequential Gaussian Simulation (SGS) is a stochastic simulation technique commonly employed for generating realizations of Gaussian random fields. Arguably, the main limitation of this technique is the high computational cost associated with determining the kriging weights. This problem is compounded by the fact that often many realizations are required to allow for an adequate uncertainty assessment. A seemingly simple way to address this problem is to keep the same simulation path for all realizations. This results in identical neighbourhood configurations and hence the kriging weights only need to be determined once and can then be re-used in all subsequent realizations. This approach is generally not recommended because it is expected to result in correlation between the realizations. Here, we challenge this common preconception and make the case for the use of a constant path approach in SGS by systematically evaluating the associated benefits and limitations. We present a detailed implementation, particularly regarding parallelization and memory requirements. Extensive numerical tests demonstrate that using a constant path allows for substantial computational gains with very limited loss of simulation accuracy. This is especially the case for a constant multi-grid path. The computational savings can be used to increase the neighbourhood size, thus allowing for a better reproduction of the spatial statistics. The outcome of this study is a recommendation for an optimal implementation of SGS that maximises accurate reproduction of the covariance structure as well as computational efficiency.

3.1 Introduction

Sequential Gaussian Simulation (SGS) is a popular method for generating stochastic values on a grid under the constraints of a statistical model and, possibly, some initially known values, herein referred to as hard data. SGS has been extensively used by practitioners because of its intuitive theoretical basis, its simple numerical implementation, and its high flexibility (e.g., Gómez-Hernández and Journel, 1993; Pebesma and Wesseling, 1998). Arguably, the major drawback of SGS is its computational cost. The exact estimation of kriging relies on taken into account all conditioning nodes, which results in large linear systems that need to be solved. For a square matrix of size n, common linear solvers have a computational complexity of $O(n^3)$ (Trefethen and Bau III, 1997), which means that the computational effort is proportional to the cube of the matrix dimension. Therefore, the sequential simulation of a grid with N nodes represents an $O(N^4)$-type problem (Dimitrakopoulos and Luo, 2004).

Various attempts have been undertaken to reduce the associated computational cost. The most widespread approach is the so-called limited or moving neighbourhood, that is, the approximation of the kriging estimate by using only a limited number of conditioning points referred to as the neighbours (e.g., Isaaks and Srivastava, 1989; Deutsch and Journel, 1992; Goovaerts, 1997). This reduces the computational complexity of SGS to $O(k^3N)$, where k denotes the number of neighbours. This approach is rooted in the observation that neighbours which are located far away from the simulated point receive small or even vanishing weights. This effect originates from the rapid decrease of correlation with distance inherent to most covariance functions and is enhanced by the presence of intermediate neighbours screening the influence of those behind (e.g., Chilès and Delfiner, 1999). However, the ommission of neighbours has shown to bias the simulation covariance matrix (Emery and Peláez, 2011; Nussbaumer et al., 2018a), which in turn results in artifacts in the realizations (e.g., Meyer, 2004). Recent works on reducing such detrimental effects, while limiting the neighbourhood size and optimizing the computational efficiency, include those of Gribov and Krivoruchko (2004), Rivoirard and Romary (2011) and Dimitrakopoulos and Luo (2004).

An alternative to reducing the size of the kriging covariance matrix is to approximate it. Barry and Kelley Pace (1997) formulate covariance-based kriging, which leads to the inversion of sparse symmetric matrices. Sparse matrix solvers improve considerably the computational performance, but this approach is limited to simulations based on covariance functions with a finite range. Furrer et al. (2006); Memarsadeghi and Mount (2007) further increase the sparsity of the matrix by tapering the covariance for large lag-distances. Related approaches comprise the approximate iterative method (Billings et al., 2002), the low rank approximation (Kammann and Wand, 2003), the Sherman-Morrison-Woodbury formula (Sakata et al., 2004), and fast summation methods (Memarsadeghi et al., 2008; Srinivasan et al., 2008).

Another approach is to only consider simulations whose covariance function is from a limited set of easily solvable covariance models. Omre et al. (1993) proposes the screening sequential simulation, which provides exact simulations for covariance models with the Markov property, such as, for example, the exponential model in 1D. Hartman and Hössjer (2008) approximate the simulated Gaussian field with a set of Gaussian Markov random fields (Rue and Tjelmeland, 2002), which can be simulated exactly and efficiently. Finally, Cressie and Johannesson (2008) consider covariance models composed of a fixed number of basic non-stationary functions. This technique is also referred to as fixed-rank kriging. A related approach is the predictive processes method (e.g., Banerjee et al., 2008).

A more general technique to cope with the high computational costs of SGS is parallelization, which reduces the computational time by splitting the work among several cores (Vargas et al., 2007; Mariethoz, 2010; Nunes and Almeida, 2010; Rasera et al., 2015). It is important to note that parallelization is not reducing the computational burden, but merely spreads it over several cores, and hence is just a useful complement to the other techniques.

The approach explored in this study aims at decreasing the overall computational cost by taking advantage of the large number of realizations typically needed in geostatistical applications. Indeed, an uncertainty assessment can only be performed with an ensemble of realizations spanning the variability of outcomes. When the simulation path, that is, the order in which the nodes are simulated, is kept identical among multiple realizations, the neighbourhood configurations of each simulated node are also identical throughout these

realizations. Because the kriging weights are computed solely with the relative distances between nodes, a constant neighbourhood configuration produces the same kriging weights. Therefore, these weights only need to be computed once and then can be re-used for all realizations. This reduces the computational effort of each additional realization to simple matrix multiplications.

While some works outline the advantages of using a constant path (e.g., Verly, 1993), the overwhelming majority still discourages its use, because of the risk to draw correlated realizations, and rather advocate a randomized path to explore the solution space more homogenously (e.g., Deutsch and Journel, 1992; Goovaerts, 1997). Conversely, Cáceres et al. (2010); Boisvert and Deutsch (2011) reported that using a constant path in SGS does not result in a significant reduction of the space of uncertainty for neither first- nor second- order statistics, while allowing for compelling reductions in computational cost. However, both studies are based on empirical evidence and hence the generic validity of their findings remains to be verified.

In the present work, we seek to provide a thorough understanding of the implications of changing the simulation path in order to assess the constant path method. The paper is organized as follows. We begin by presenting a methodological description of randomized path simulations (section 3.2), followed by the implementation of a constant path method (section 3.3) and the quantification of the associated computational gains (section 3.4). Finally, we discuss some limitations of the covariance matrix evaluation (section 3.5).

3.2 Theory of Randomized Paths Simulations

In order to understand the implications of generating stochastic realizations based on the same simulation path, the links between the random function (RF) Z, the realizations z, and the path p_i need to be explored in some detail.

3.2.1 Definition of a Random Function

In probability theory, a random variable (RV) denoted X is a deterministic function mapping the set of possible outcomes Ω of a random phenomenon to their values, usually a real number $\mathbb{R}$,

$$\begin{aligned} X &: \Omega \to \mathbb{R} \\ &\omega \mapsto x. \end{aligned} \tag{3.1}$$

In the definition of a RV, Ω has to be a probability space, which implies that each possible outcome ω has a well-defined probability. Thus, the probability $P(X \leq x_T)$ is defined by the set of events $\{\omega \in \Omega : X(\omega) \leq x_T\}$.

For instance, a RV describing the sum of two rolled dice n_1 and n_2 is defined as the function mapping every possible outcome (n_1, n_2) to the measure $n_1 + n_2$

$$X(\{n_1, n_2\}) = n_1 + n_2. \tag{3.2}$$

With this formalism, the probability of the sum of two dice being 5 is defined as

$$P(X = 5) := P(\{n_1, n_2 \in \{1,2,3,4,5,6\} : n_1 + n_2 = 5\}) = P(\{1,4\},\{2,3\},\{3,2\},\{4,1\}) = 4/6^2 = 1/9. \tag{3.3}$$

A realization $x^{(l)}$ is the value observed from a RV X given a specific outcome of the random phenomenon, also called random variate, ω_l

$$x^{(l)} = X(\omega_l). \tag{3.4}$$

3.2.2 Sequential Gaussian Simulation

SGS is an algorithm whose purpose is to produce realizations $z^{(l)}(\boldsymbol{u})$ of a *regionalized multi-Gaussian random function (RF)* $Z(\boldsymbol{u})$.

1. A RF is a collection of indexed RV. If the indexation is multi-dimensional, the collection is usually referred to as random field instead.
2. A RF is called regionalized (Matheron, 1965) if it is distributed in a continuous space domain $D \subset \mathbb{R}^n$,

$$Z = \{Z(\boldsymbol{u}), \boldsymbol{u} \in D\}, \tag{3.5}$$

   where $\boldsymbol{u}$ represents a space coordinate vector.
3. A RF is multi-Gaussian if any finite collection of its components has a multi-variate normal distribution. While this constraint is restrictive, it allows for the RF to be fully determined by its first- and second-order moments, that is, the mean $\boldsymbol{\mu}_Z$ and the covariance matrix $\boldsymbol{C}_Z$

$$Z \sim \mathcal{N}(\boldsymbol{\mu}_Z, \boldsymbol{C}_Z). \tag{3.6}$$

SGS takes advantage of this multi-Gaussian property to produce realizations of Z. It iteratively visits each node of the grid, computes the kriging estimate and variance error σ_E based on previously simulated nodes and samples a value from the corresponding conditional probability distribution. A newly simulated node thus becomes a conditioning node for the next one to be simulated. Mathematically, this can be summarized as

$$Z(\boldsymbol{u}_i) = \sum_{j=1}^{i-1} \lambda_j(\boldsymbol{u}_i) Z(\boldsymbol{u}_j) + \sigma_E(\boldsymbol{u}_i) U(\boldsymbol{u}_i), \quad \forall i = 1, \ldots, n, \tag{3.7}$$

where U is a standard Gaussian vector used for randomly sampling the conditional distribution and λ_j are the kriging weights which define the influence of the conditioning nodes.

In the framework of probability theory, the underlying random phenomenon of $Z(\boldsymbol{u})$ is the standard normal variable U (equation 3.7). Indeed, given a specific vector U_l, SGS always produces the same realization $z^{(l)}$

$$z^{(l)} = Z(U_l). \tag{3.8}$$

3.2.3 Algorithm-Driven Random Function (ADRF)

For computational reasons, SGS is commonly used with a limited neighbourhood. As a result, the realizations are altered and the actual simulated RF deviates from the original RF Z. In such a case, SGS is sensitive to both the simulation path and the neighbourhood search strategy.

As it is a common situation in geostatistics to have algorithms deviating from the original RF, Boucher (2007) introduced the formalism of an algorithm-driven random function (ADRF): an RF defined by an algorithm which is parametrized by a random variate, or seed number, and a set of parameters. From now on, the term simulated RF will be used to refer to the ADRF simulated by the SGS algorithm. For simplicity, the neighbourhood search strategy is assumed to be defined as the n closest neighbours of the simulated node and is therefore constant for a neighbourhood configuration.

In this context, SGS is a technique which produces realizations of the simulated RF Z_{p_i} defined by

$$z_{p_i}^{(l)} = Z(U_l; p_i) = Z_{p_i}(U_l), \tag{3.9}$$

where U_l is the random variate of the underlying random process and p_i the simulation path used as a parameter in the simulated RF.

3.2.4 Covariance Matrix for Error Quantification

The errors due to the limited neighbourhood in Z_{p_i} can be fully characterized through the mismatch $\boldsymbol{\varepsilon}$ between the covariance matrix of the simulated RF $\boldsymbol{C}_{Z_{p_i}}$ and the covariance matrix of the target RV $\boldsymbol{C}_Z$ (Emery and Peláez, 2011), hereafter, referred to as the simulated covariance matrix and the model covariance matrix, respectively,

$$\boldsymbol{\varepsilon} = \boldsymbol{C}_{Z_{p_i}} - \boldsymbol{C}_Z \tag{3.10}$$

The model covariance matrix is computed from the covariance function of the spatial model

$$\boldsymbol{C}_Z(\alpha, \beta) = C(\boldsymbol{u}_\alpha - \boldsymbol{u}_\beta). \tag{3.11}$$

The simulation covariance matrix $\boldsymbol{C}_{Z_{p_i}}$ can be theoretically evaluated based on the kriging weights and the variance errors. Indeed, Emery and Peláez (2011) showed that equation 3.7 can be reformulated to extract U

$$U = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ -\frac{\lambda_1^{n-1}}{\sigma_{n-1}} & \cdots & \frac{1}{\sigma_{n-1}} & 0 \\ -\frac{\lambda_1^{n}}{\sigma_{n}} & \cdots & -\frac{\lambda_{n-1}^{n}}{\sigma_{n}} & \frac{1}{\sigma_{n}} \end{bmatrix} Z_{p_i} = \Lambda Z_{p_i}, \tag{3.12}$$

where Λ is constructed during the simulation by storing each kriging weight and variance error. The simulation covariance matrix can be expressed exclusively with Λ

$$
\begin{aligned}
\boldsymbol{C}_{Z_{p_i}} &= E\left[\left(Z_p - E\left[Z_{p_i}\right]\right)\left(Z_p - E\left[Z_{p_i}\right]\right)^T\right] \\
&= E\left[Z_{p_i} {Z_{p_i}}^T\right] \\
&= E\left[\left(\Lambda^{-1}U\right)\left(\Lambda^{-1}U\right)^T\right] \\
&= \Lambda^{-1} E\left[UU^T\right](\Lambda^{-1})^T \\
&= \Lambda^{-1}(\Lambda^{-1})^T.
\end{aligned} \tag{3.13}
$$

A key advantage of this evaluation of a simulated RF is that it does not rely on realization statistics, but purely on the actual simulated RF equations. The random variate is taken out of the equation, leaving only the deterministic part of the algorithm.

3.2.5 Simulated Random Function with a Randomized Path

The purpose of this study is to compare the error of a simulated RF with and without a randomized path, that is, the simulation path is changed for each realization or kept constant. When the simulation path p_i is randomly sampled among a set of paths P, it can be viewed as a random variate rather than a parameter. In this case, a newly simulated RF Z_P can be defined

$$
z_P^{(l,i)} = Z_P(U_l, p_i), \quad p_i \in P, \tag{3.14}
$$

where both U_l and p_i are random variates. Yet, it is important to note that as each realization $z_P^{(l,i)}$ of Z_P is sampled with a unique path p_i and, therefore, it is also a realization of the simulated RF Z_{p_i} where the path p_i is viewed as a parameter rather than a random variate

$$
Z_P(U_l, p_i) = z_P^{(l,i)} = z_{p_i}^{(l)} = Z_{p_i}(U_l). \tag{3.15}
$$

The simulation covariance matrix $\boldsymbol{C}_{Z_P}$ can be computed based on $\boldsymbol{C}_{Z_{p_i}}$ using the discrete version of the covariance due to the finite number of realizations. When n_k realizations $\{z_P^{(l,k)}, k = 1, \ldots, n_k\}$ are performed, the empirical covariance at two locations $\boldsymbol{u}_\alpha$ and $\boldsymbol{u}_\beta$ of the grid is given by

$$\begin{aligned} \boldsymbol{C}_{Z_P}(\alpha, \beta) &= \mathrm{Cov}\left[Z_P(\boldsymbol{u}_\alpha), Z_P(\boldsymbol{u}_\beta)\right] \\ &= \frac{1}{n_k} \sum_k \{z_P^{(l,k)}(\boldsymbol{u}_\alpha) - E[Z_P(\boldsymbol{u}_\alpha)]\}\{z_P^{(l,k)}(\boldsymbol{u}_\beta) - E[Z_P(\boldsymbol{u}_\beta])\}. \end{aligned} \tag{3.16}$$

Emery and Peláez (2011) showed that the expected value of each simulated RF Z_{p_i} is exactly reproduced for unconditional simulations and for conditional simulations provided that all hard data are present in every kriging system. We thus have

$$E[Z_P(\boldsymbol{u})] = E[Z_{p_i}(\boldsymbol{u})] = E[Z(\boldsymbol{u})], \quad \forall\, p_i \in P. \tag{3.17}$$

Furthermore, as outlined by equation 3.15, each realization $z_P^{(k,l)}$ has only one path associated with it and can be replaced by $z_{p_k}^{(l)}$.

Combining these two aspects in equation 3.16 simplifies the covariance matrix of the simulated RF Z_P to an arithmetic average of the covariance matrices of the simulated RFs Z_{p_k} where $p_k \in P$

$$\begin{aligned} \boldsymbol{C}_{Z_P}(\alpha, \beta) &= \frac{1}{n_k} \sum_k \{z_{p_k}^{(k)}(\boldsymbol{u}_\alpha) - E[Z_{p_k}(\boldsymbol{u}_\alpha)]\}\{z_{p_k}^{(k)}(\boldsymbol{u}_\beta) - E[Z_{p_k}(\boldsymbol{u}_\beta])\} \\ &= \frac{1}{n_k} \sum_k \boldsymbol{C}_{Z_{p_k}}(\alpha, \beta). \end{aligned} \tag{3.18}$$

Let $\varepsilon^{p_i}_{\alpha,\beta}$ denote the covariance error of the simulated RF Z_{p_i} between a pair of points $\boldsymbol{u}_\alpha$, $\boldsymbol{u}_\beta$. Using equation 3.10, we then have

$$\varepsilon^{p_i}_{\alpha,\beta} = \boldsymbol{C}_{Z_{p_i}}(\alpha,\beta) - \boldsymbol{C}_Z(\boldsymbol{u}_\alpha,\boldsymbol{u}_\beta), \tag{3.19}$$

and hence, using equation 3.18,

$$\varepsilon^{P}_{\alpha,\beta} = \frac{1}{n_P}\sum_{i}^{n_P} \boldsymbol{C}_{Z_{p_i}}(\alpha,\beta) - \boldsymbol{C}_Z(\boldsymbol{u}_\alpha,\boldsymbol{u}_\beta) = \frac{1}{n_P}\sum_{i}^{n_P} \varepsilon^{p_i}_{\alpha,\beta}. \tag{3.20}$$

Therefore, the errors of a simulated RF with a randomized path Z_P are equal to the average of the errors of all simulated RFs Z_{p_i}, such that $p_i \in P$. This is exact for an infinite number n_P of paths in P and is reasonably well approximated if a large number of paths is used. This average operator has two main effects: (1) the overall expected value of the errors are identical but (2) their variances are reduced proportionally to the number of paths used.

3.3 Numerical Implementation and Computational Savings

3.3.1 Pseudo-Code

The simplest modification of the classical SGS (Algorithm 1) allowing for a constant path consists in moving the loop over the realizations into the loop over the grid nodes, following

the computation of the weights (Algorithm 2). This simple permutation of loops illustrates why the simulation path P has to remain the same for all realizations.

Input: Number of realizations m, grid size n, covariance structure $covar$, neighbourhood size k

Output: Realizations $R[m,n]$

$R[m,n]$ ▷ `Initialize empty realizations`

for $j \leftarrow 1$ **to** m **do**

  $P[n] \leftarrow$ `GeneratePath(`n`, seed)` ▷ `Randomly generates path`

  for $i \leftarrow 1$ **to** n **do**

    $id_i \leftarrow P(i)$ ▷ `Get index of simulated node`

    $\boldsymbol{id}_k \leftarrow$ `Findneighbours(`$id_i, P(1 \text{ to } i-1), k$`)`

    $\boldsymbol{\lambda}, \sigma_E \leftarrow$ `ComputeWeight(`$id_i, \boldsymbol{id}_k, covar$`)`

    $R(j, id_i) \leftarrow$ `SimulateValue(`$\boldsymbol{\lambda}, \sigma_E, R(j, \boldsymbol{id}_k)$`, seed)`

  end

end

Algorithm 1: Traditional SGS

Input: Number of realizations m, grid size n, covariance structure $covar$, neighbourhood size k

Output: Realizations $R[m,n]$

$R[m,n]$

$P[n] \leftarrow$ `GeneratePath(`n`, seed)` ▷ `Generates a single path`

for $i \leftarrow 1$ **to** n **do**

  $id_i \leftarrow P(i)$

  $\boldsymbol{id}_k \leftarrow$ `Findneighbours(`$id_i, P(1 \text{ to } i-1), k$`)`

  $\boldsymbol{\lambda}, \sigma_E \leftarrow$ `ComputeWeight(`$id_i, \boldsymbol{id}_k, covar$`)`

  for $j \leftarrow 1$ **to** m **do**

    $R(j, id_i) \leftarrow$ `SimulateValue(`$\boldsymbol{\lambda}, \sigma_E, R(j, \boldsymbol{id}_k)$`, seed)`

  end

end

Algorithm 2: SGS with constant path

3.3.2 Parallelization

Parallelization is easily implemented on the traditional SGS by sending each realization to a different core, which corresponds to parallelizing the loop of line 2 in Algorithm 1. In the new implementation, the realization loop can also be parallelized (line 7 in Algorithm 2). Yet, a more efficient implementation is described in Algorithm 3, where the weights are computed and stored prior to the simulation. An important advantage of separating the computation of the weights and the simulation of the nodes is the ability to compute the weights in parallel. In practice, this is implemented by searching for neighbours with a lower path index than that of the simulated node, instead of searching for all existing values in the realization. Note that this parallelization strategy is not as easily implemented in the traditional SGS algorithm because the simulation of the node is in the same loop and requires the value of the neighbours. Finally, a third option is to parallelize the whole algorithm so that each core computes a subset of realizations, each with a different constant path. This hybrid solution between randomized and constant path does not combine the computational benefits of

parallelization and constant path, but improves realization accuracy as is later discussed in the paper.

Input: Number of realizations m, grid size n, covariance structure $covar$, neighbourhood size k

Output: Realization $R[m,n]$

```
1  Λ[n,k]                                  ▷ Initialize the kriging weights array
2  Σ_E[n]                                  ▷ Initialize the kriging variance error array
3  ID_k[n,k]                               ▷ Initialize neighbours indices array
4  P[n] ← GeneratePath(n, seed)
5  parfor i ← 1 to n do
6      id_i ← P(i)
7      ID_k(i) ← Findneighbours(id_i, P(1 to i − 1), k)
8      Λ(i), Σ_E(i) ← ComputeWeight(id_i, ID_k(i), covar)
9  end
10 R[m,n]
11 parfor j ← 1 to m do
12     for i ← 1 to n do
13         id_i ← P(i)
14         R(j, id_i) ← SimulateValue(Λ(i), Σ_E(i), R(j, ID_k(i)), seed)
15     end
16 end
```

Algorithm 3: Parallelized SGS with constant path

3.3.3 Memory Requirements

Compared to the traditional approach (Algorithm 1), the serial implementation (Algorithm 2) does not require additional memory. However, the parallel implementation (Algorithm 3) has to store the neighbours indices $ID_k[n,k]$, the kriging weights $\Lambda[n,k]$, and the kriging variance errors $\Sigma_E[n]$, which results in a memory increase of $4nk + 8nk + 8n = (12k + 8)n$ bytes. In comparison, storing m realizations requires $8nm$ bytes. Thus, reducing m by $1.5k$

compensates the memory increase, which is not a problem for typical applications where $m >> k$.

In simulations with large grids and/or large numbers of realizations, memory can still become an important issue. In the traditional implementation (Algorithm 1), this was handled by writing each realization or group of realizations on the disk when completed. In the parallelized implementation (Algorithm 3), the realizations can similarly be written on the disk in the realization loop (line 11). If the neighbourhood size k is large, storing the weights Λ and the neighbours indexes ID_k can become a challenge. The solution for such situations is to write them during the first loop (line 5) and read them in the realization loop (line 11) but this can increase the computational time. Note that, in most typical situations, using a constant path remains more efficient because the computation of the weights is more expensive than writing and reading a file. Moreover, simulations with such settings are usually performed on high-performance computers, which has large amounts of memory available and thus are able to store at least the weights and indices. Despite not storing the weights and indices, the serial implementation (Algorithm 2) cannot be efficiently used for simulations with a large grid and/or large numbers of realizations.

3.3.4 Computational Time

With the constant path approach, the only operation left to be repeated for each realization is to iterate through all the nodes and simulate a value, which corresponds to a simple vector multiplication (line 14 in Algorithm 3). As such, with an ideal implementation, the gain in computational time corresponds to the effort of finding the neighbours (line 7 in Algorithm 3), and computing the kriging weights and variance errors (line 8 in Algorithm 3) for each additional realization $j = \{2, \ldots, m\}$. in the following, these two computational efforts are described in more detail to better understand the conditions and, the extent of the benefits of using a constant path.

First, the kriging neighbours are found by posing the optimization problem known as k-Nearest Neighbour (k-NN). The optimal solution to this problem varies with the simulation parameters: (1) grid size, (2) neighbourhood size, (3) number and location of hard data, and

(4) covariance model range. Because of this diversity of settings, the choice of an appropriate neighbourhood search strategy is of upmost importance as it often is the bottleneck with regard to computational time in SGS. The most common strategies in SGS are listed below. The "exhaustive search" sorts all n available nodes by their distance to the simulated node and takes the first k nodes. The complexity of the well-known quicksort algorithm is $O(n\log(n))$ (Hoare, 1962). A considerable improvement of this strategy consists of only sorting the first k nodes, thus reducing the complexity to $O(n+k\log(k))$ with partial quicksort (Martínez, 2004). This strategy is optimal for simulations of a small grid with a large neighbourhood size and presents the advantage of being able to treat equally data at any location, such as, for instance, hard data not located in the grid. The "spiral search" (Deutsch and Journel, 1992) visits each node of the grid by order of proximity to the simulated node, skips the empty nodes, and stops once k simulated nodes have been found. This method is efficient for large grids with a small neighbourhood, particularly when combined with a multi-grid path. Hard data have to be moved to the closest node, at least temporarily. Another trick is to define an exploration distance threshold to limit the search to the nodes within a certain distance to the simulated one. The "superblock search" (Journel and Huijbregts, 1978; Deutsch and Journel, 1992) and "tree-based search" (Hassanpour and Leuangthong, 2006; Manchuk and Deutsch, 2012) provide alternative solutions, which can be more efficient, especially for the simulation of irregular points or when dealing with hard data within a two-part search method (Deutsch and Journel, 1992).

Secondly, the time to compute the kriging weights and variance errors is mainly determined by solving kriging systems, one for each simulated node. This has an overall computational complexity of $O(nk^3)$ (Dimitrakopoulos and Luo, 2004), making the neighbourhood size k the main parameter influencing the computational time. Another common solution to reduce the computational cost in regular grids is to use a covariance lookup table, which pre-computes and stores the covariance of all pairs of nodes within the search ellipsoid. Spiral search is particularly well-suited for this.

The total computational savings provided by the constant path are the sum of these two computational costs multiplied by the number of realizations required, such that implementing the constant path approach is only rewarding when several realizations are needed.

In addition, the constant path allows for a more flexible choice of parameters because the computational cost associated with these parameters is only paid once. This means that the selection of a sub-optimal neighbourhood strategy does not increase the computational cost as much as it would in the traditional approach. Since this cost arises only once, this provides the opportunity of increasing the neighbourhood size to reduce the simulation error without excessively increasing computational burden.

The computational benefit of using a constant path is numerically assessed with the speed-up, that is, the ratio of computational time of the traditional SGS T^{trad} over the constant path SGS T^{cst}. It can be expressed as a function of the number of realizations m based on the simulation of a single realization,

$$S = \frac{T^{trad}}{T^{cst}} = \frac{mT^{trad}(1)}{T^{cst}(1) + (m-1)T^{cst}_{real}(1)}, \tag{3.21}$$

where $T^{cst}_{real}(1)$ refers to the time spent in the realizations loop for a single realization (lines 11-16 in Algorithm 3). Figure 3.1 shows the speed-up for unconditional simulations performed with a spherical covariance function of range of 20, using a spiral search and a multi-grid path, and without covariance lookup table. The constant path SGS is simulated with the parallel implementation (Algorithm 3) but running on a single core. The processor used is a AMD Opteron (2300Mhz). These results show that the speed-up increases with the neighbourhood size and scales well with grid size. In this particular case, around 3'000, 5'000 and 6'000 realizations can be performed with a constant path for the same duration than 100 realizations without constant path for neighbourhood sizes of 20, 52 and 108 nodes, respectively. This case study reveals that, even if only 5 realizations are needed, it is computationally more efficient to generate them with a constant path and a neighborhood of 108 nodes than without constant path and a neighborhood of 20 nodes.

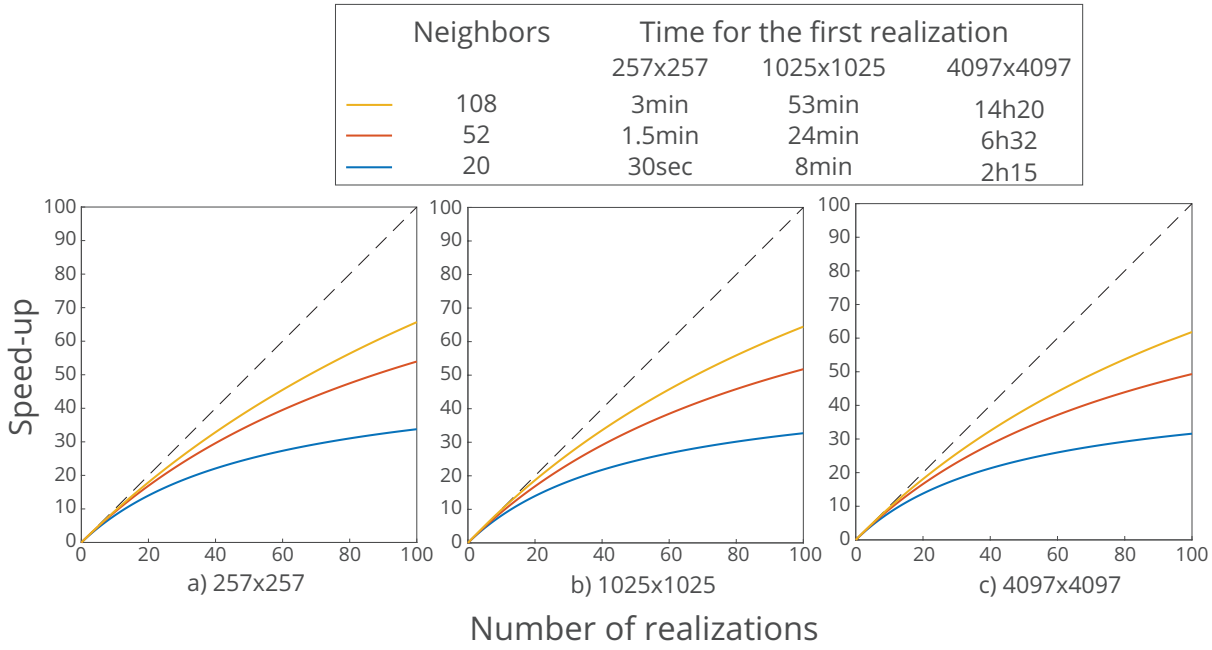


Figure 3.1 – Speed-up as a function of the number of realizations performed for 3 differents grid sizes and neighbourhood sizes. The grid sizes were selected such that the multi-grid path is optimal and the neighbourhood size such that the resulting full neighbourhood is symmetric. The computational time for the first realization is also shown in the legend.

3.4 Simulation Errors

In the following, six covariance functions (exponential, Gaussian, spherical, hyperbolic, k-Bessel, and cardinal sine) are considered because of their variability close to the origin and near the range as well as is their popularity. A correlation range of 20 nodes normalized by their integral value is used. Because of the large cost of computing the exact covariance matrix, a small grid of 65x65 nodes is used in this section.

3.4.1 Covariance Errors

In order to compare large covariance matrices, Emery and Peláez (2011) proposed to aggregate the covariance errors with the standardized Frobenius norm (SFN)

$$\eta^{p_i} = \frac{\|\boldsymbol{C}_{Z_{p_i}} - \boldsymbol{C}_Z\|}{\|\boldsymbol{C}_Z\|}, \tag{3.22}$$

where $\|\cdot\|$ denotes the $L_{2,2}$ or Frobenius norm. Limitations associated this metric are discussed in section 3.5.3.

With this single value, it is possible to compare the error of simulated RFs using different numbers of paths n_P as well as different covariance function types. The SFNs of 512 simulated RFs with a constant path Z_{p_i} were computed. Then, using equation 3.20, the simulated RFs with a randomized path Z_P, with $n_P = 2, \ldots, 128$ were computed. Figure 3.2 shows the corresponding results in the form of boxplots. Simulations were performed with a neighbourhood size of 20 nodes and a fully randomized path.

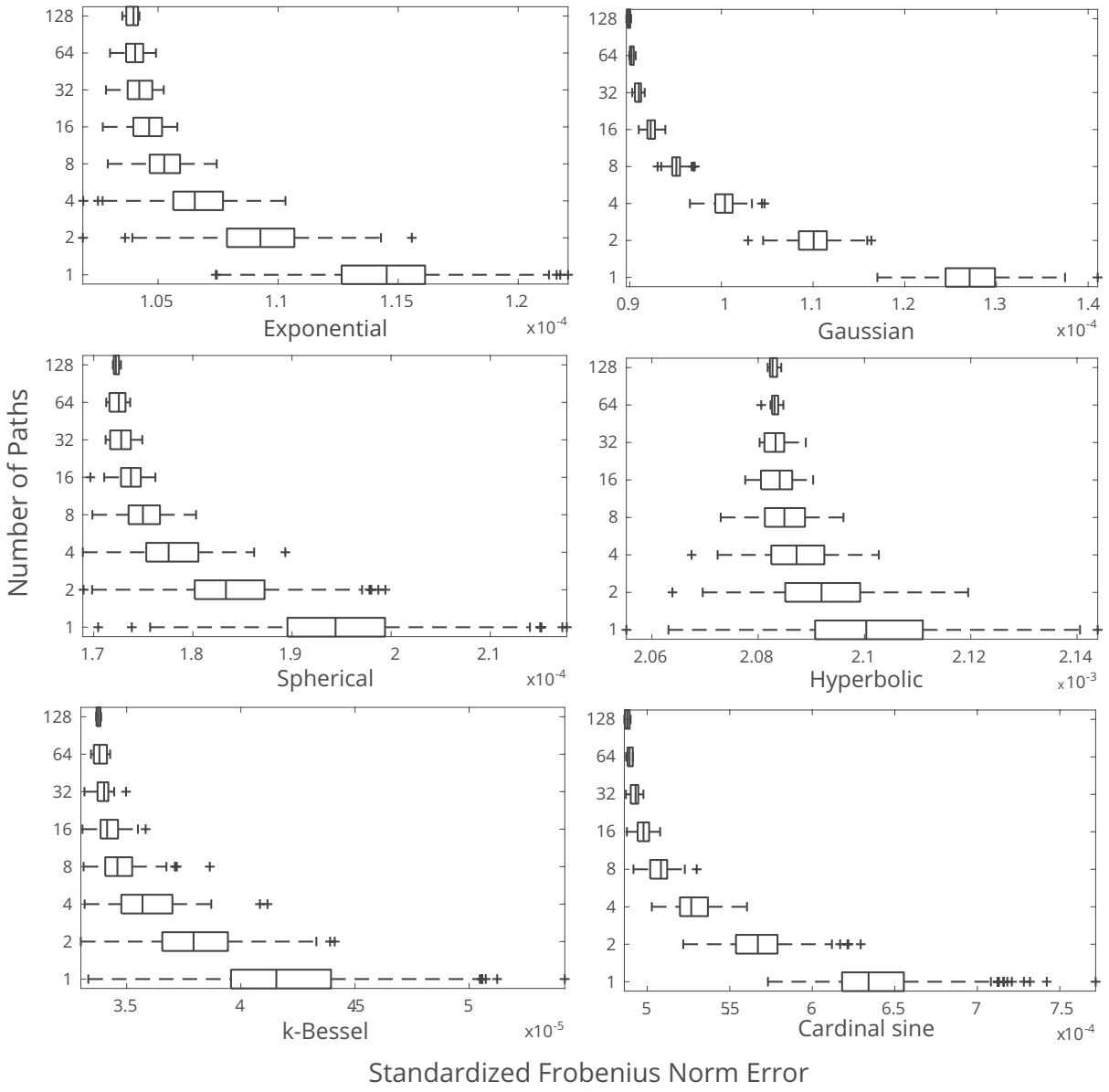


Figure 3.2 – Boxplots of the SFN of simulated RFs Z_P for different numbers of simulation paths $n_P = 1, \ldots, 128$. 512 simulated RFs Z_{p_i} were computed and different numbers of them combined to construct the simulated RFs with a fully randomized path Z_P according to equation 3.20.

For all covariance function types, a decrease of the mean and variance of the SFN is observed for simulated RFs using larger numbers of paths. The reduction of the variance is an expected result of the averaging described in equation 3.20. But the decrease of the mean error is due to the aggregation of errors in the SFN. This will be further explored in the discussion part. It can also be noted that the value of the SFN rapidly converges and, in this specific case,

with more than approximately 16 different paths no significant error reduction is observed. The magnitude of the error reduction is strongly linked to the shape of covariance function where, for instance, the Gaussian or cardinal sine present larger reductions than hyperbolic or exponential covariance functions.

3.4.2 Sensitivity to Neighbourhood and Path Type

To provide a perspective on these error reductions obtained by varying the path, we compare them to those obtained for an increase in neighbourhood size and the use of different types of paths.

Firstly, the influence of neighbourhood size is analyzed. Figure 3.3 shows the SFN for simulated RFs for a variable number of paths and several neighbourhood sizes. The number of neighbours used in the kriging has a stronger influence on the error than varying the path. Only the simulated RF for a spherical covariance function is shown because the results for all other covariance function types considered in this study are qualitatively similar. These results reinforce the suggestion made in section 3.3.4 to take advantage of the computational savings associated with the use of a constant path to increase the neighbourhood size. Indeed, a much larger reduction of covariance errors is achieved by increasing the neighborhood size than by using different randomized path.

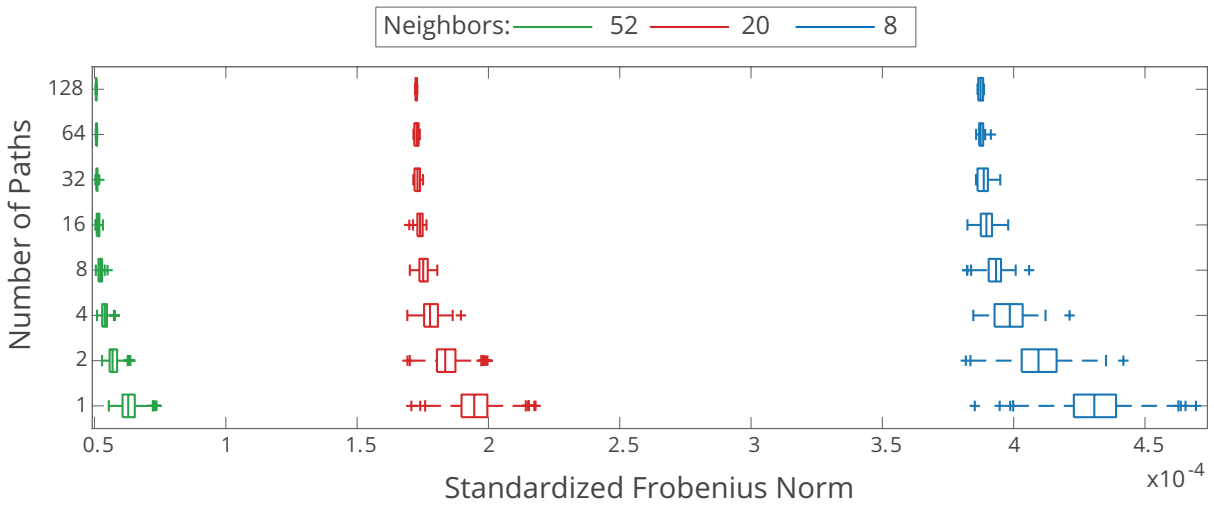


Figure 3.3 – Boxplot of the SFN for several simulated RFs for different numbers of simulation paths and different neighbourhood sizes for a spherical covariance function.

Secondly, two different path types are compared: a random path which visits each cell of the grid equiprobably and a multi-grid path using a series of nested grids to guide the order of the simulation. More specifically, the first grid simulates the four corners of the final grid, then each subsequent grid doubles the previous grid resolution, re-using the previously simulated nodes and simulating the empty nodes randomly. Note that both random and multi-grid paths can be randomized. Figure 3.4 compares the SFN of simulated RFs with random and multi-grid paths for different numbers of randomized paths.

As already shown in Nussbaumer et al. (2018a), Figure 3.4 illustrates that a multi-grid path generally improves the reproduction of the covariance matrix. However, it also shows that using multiple randomized multi-grid paths only reduces slightly both the magnitude of the SFN and its variability. Note that the cardinal sine covariance function presents an interesting execption where increasing the number of randomized paths results in greater SFN reduction with a random path than with a multi-grid path. This can be explained by the fact that when few neighbors are available the reproduction of the covariance function can follow a non-linear behaviour due to the cyclical nature of this covariance function. Figure 10 in Nussbaumer et al. (2018a) presents the covariance function errors for the cardinal sine covariance function and shows that all types of paths produce similar errors. With this particular covariance function, the neighbordhood size is the most important parameter, as illustrated by Figure 9 in Nussbaumer et al. (2018a).

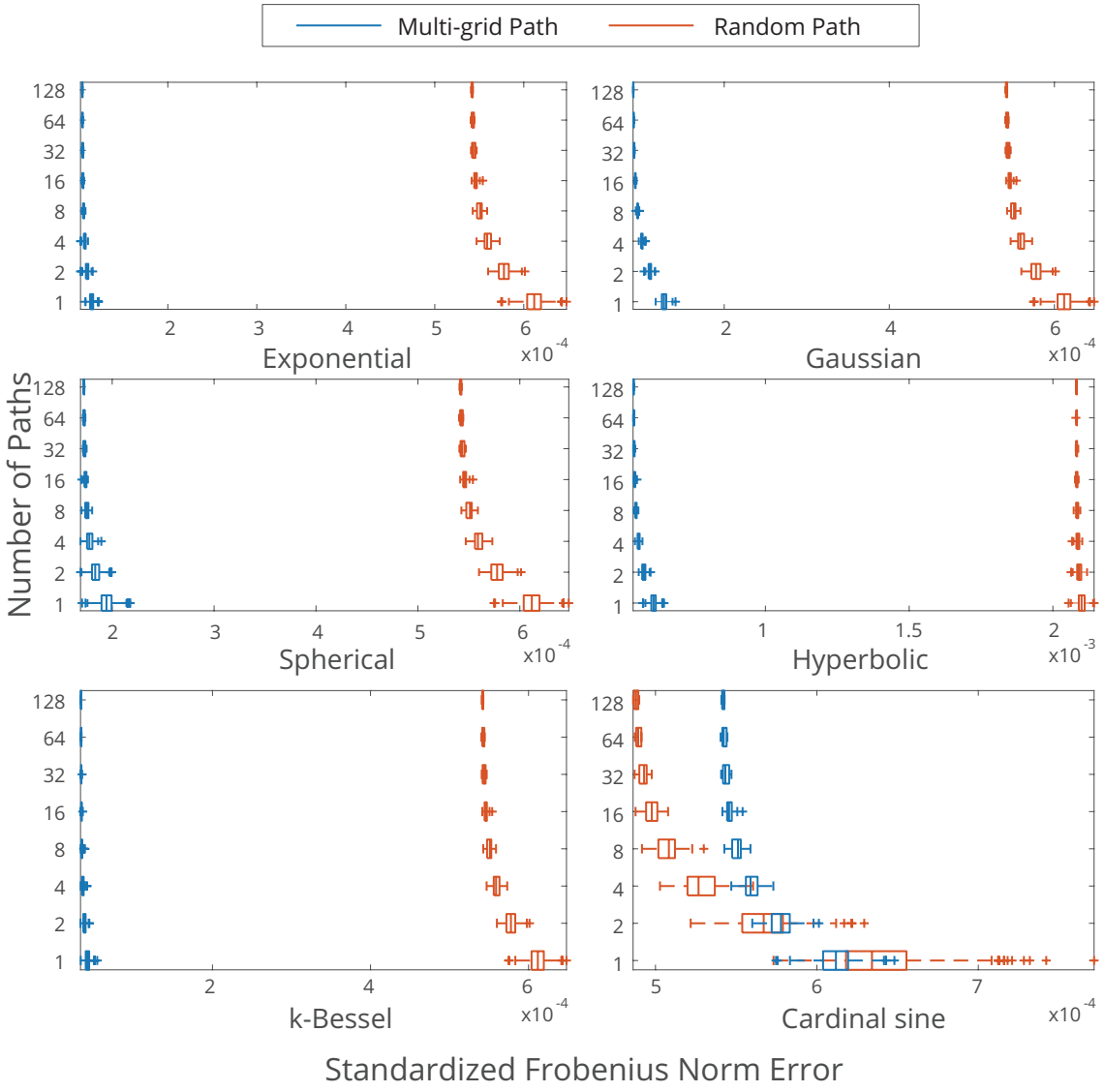


Figure 3.4 – Boxplot of the SFN for several simulated RFs using different numbers of simulation paths for multi-grid or random simulation paths.

A compelling asset of the multi-grid path is it allows for switching from either randomized to constant path or vice versa at each grid level. This feature is attractive because a randomized path at the initial coarse levels maximizes the variability of the poorly constrained nodes. Conversely, a constant path at the subsequent finer levels allows for important computational gains while not compromising the variability, because these nodes are anyway well constrained by the coarser grid. Figure 3.5 shows the SFNs and corresponding speed-ups

for simulations starting with a randomized path and switching to constant path at differents multi-grid levels. The simulations were performed on a 129x129 grid with a spherical covariance function and 20 neighbours.

Switching to a constant path during levels 1 to 4 results in similar SFN values than with a fully constant path for any number of paths. This means that changing the order of simulated nodes belonging to these levels does not result in any measurable improvement in the reproduction of the final covariance matrix. In this case study, switching path at level 5 results in a relatively limited error reduction when several randomized path are used. Switching path at the following levels has little or no impact on the reproduction of the covariance matrix, which can be explained by the grid spacing of these levels being smaller than the covariance function range. Indeed, during the simulation of those levels, the simulation of nodes is well constrained by the previous multi-grids. In all cases, the magnitude of the errors is small, even when using fully constant path, as shown in Figure 3.3. As a general rule, multi-grid levels with a grid spacing much lower than the covariance range can be simulated with a constant path without affecting the SFN values. The corresponding speed-up indicates that using a constant path on the last levels of the multi-grid provides the largest computational improvement due to the large number of nodes.

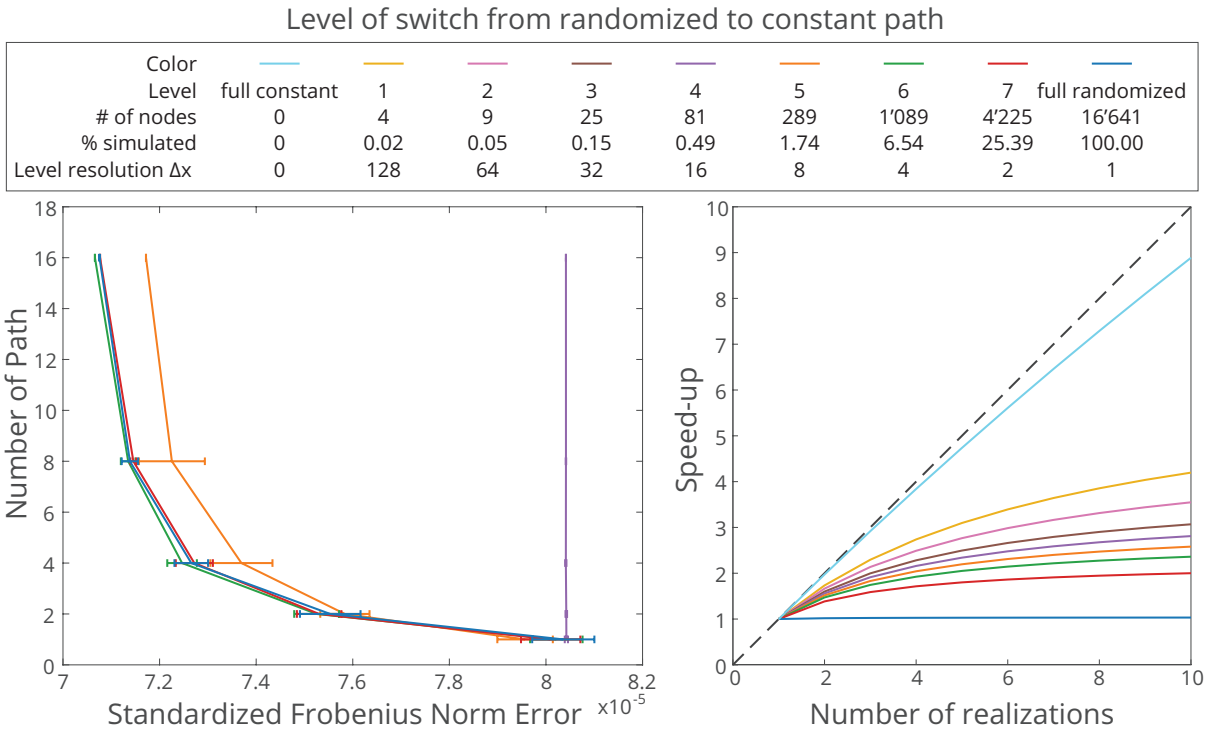


Figure 3.5 – (Left) Mean and standard deviation of the SFN using up to 16 different randomized multi-grid paths for simulations where the constant path approach is switched on at different grid level. Note that the curves for levels 1-4 and for the fully constant path are superimposed. (Right) Corresponding simulation speed-up of simulation.

3.4.3 Conditional Simulation

For conditional simulations, the constant path strategy can be implemented in the same way as for the unconditional case by computing once and storing the kriging weights associated with the hard data for all simulated nodes and then re-using them for each additional realization.

Conditional simulations can be considered as a special case of unconditional simulations, for which the first nodes of the path are the hard data with a constant outcome and whose sampling outcome is constant. Emery and Peláez (2011) show that the assessment of a conditional simulation is based on the reproduction of the covariance matrix of the simulated nodes and the reproduction of the expected field. They demonstrate that the covariance matrix of the simulated nodes is similar to the one of the unconditional case (Equation 11 in

Emery and Peláez (2011)). The exact reproduction of the expected value, which corresponds to matching the kriging predictor, is only achieved when all hard data are used in every kriging system (Appendix 1 in Emery and Peláez (2011)). In this case, using a constant or randomized path has the same benefits and consequences as it has for unconditional simulations where a constant path is enforced on the first nodes equivalent to the hard data.

However, if too many hard data are present, not all can be retained in every kriging system. In this case, the expected field does not match the kriging predictor, and, in turn, Equation 3.17 does not hold. The consequences on the effects of using a constant path are difficult to assess in general. However, if hard data are included more generously with a two part search strategy, it can be expected that the effect of the missing hard data is limited. Consequently, from a qualitative point view, the overall results of this study remain valid.

Another effect of including hard data is that the presence of numerous and scattered hard data constrains the simulation, such that, randomizing the path results in a smaller reduction of errors. An equivalent effect can be observed in Figure 3.5, where simulations switching from a randomized to constant path for the last levels of the multi-grid produce similar errors reductions.

3.5 Discussion

3.5.1 Unlucky Path

An unlucky path is a particular path which leads to especially bad covariance matrix reproduction. These paths can be discussed by comparing the tails of the errors distribution in Figure 3.2,3.3 and 3.4. With the reduction of variance demonstrated in equation 3.20, maximal SFN values in Figure 3.2 also decrease with the largest number of simulation paths. Based on Figures 3.3 and 3.4, increasing the neighborhood size or using a multi-grid approach reduces even further the occurrence of unlucky paths as the variance of their resulting covariance errors is smaller.

3.5.2 Empirical Covariance

In this section, the assessment of error based on aggregating multiple realizations errors with empirical inter-realization statistics is discussed.

Firstly, even if the covariance error of several realizations is small, it does not mean that each realization has a smaller error. Indeed, each realization could contain large errors, but averaging these errors may lead to small inter-realization errors. This is an important consideration for our comparison because each realization of a simulation RF with a randomized path is produced with a single path and, therefore, with the same potential errors as a realization of a simulation RF with a constant path. This means that although a simulation RF with a randomized path can improve the overall inter-realizations statistics, each individual realization still contains a similar amount of errors.

Secondly, when covariance errors are empirically computed, the number of realizations is limited and the ergodic fluctuations should be taken into account. In theory, for equation 3.18 to be correct, an infinite number of realizations of each constant path simulation is required. As the path changes for each realization, discrepancies between the simulation covariance and the model covariance are expected (Matheron, 1989; Emery, 2004).

3.5.3 Aggregation of the Covariance Matrix Errors

While the SFN provides a reliable and simple measure of error (equation 3.22), there are also some limitations that need to be highlighted. Essentially, the SFN can be viewed as a normalized root mean square error (RMSE) measure

$$\eta^{p_i} = \frac{\sum_{\alpha}^{N}\sum_{\beta}^{N}\left(\varepsilon_{\alpha,\beta}^{p_i}\right)^2}{\sum_{\alpha}^{N}\sum_{\beta}^{N}\left(C_Z(\boldsymbol{u}_\alpha - \boldsymbol{u}_\beta)\right)^2}, \tag{3.23}$$

which has the well-known effect of enhancing the influence of larger errors. Thus, a few pairs of nodes with large covariance errors are predominant over many small errors. While this is a commonly accepted procedure, a thorough investigation of the effect of varying the path requires analysis of the exact distribution of the covariance errors. Figure 3.6 displays the distribution of all values of $\varepsilon_{\alpha,\beta}$ (equation 3.19) of 20 simulated RFs with a constant path and the simulated RF with a randomized path. The latter is constructed by combining the previous 20 simulated RFs according to equation 3.20.

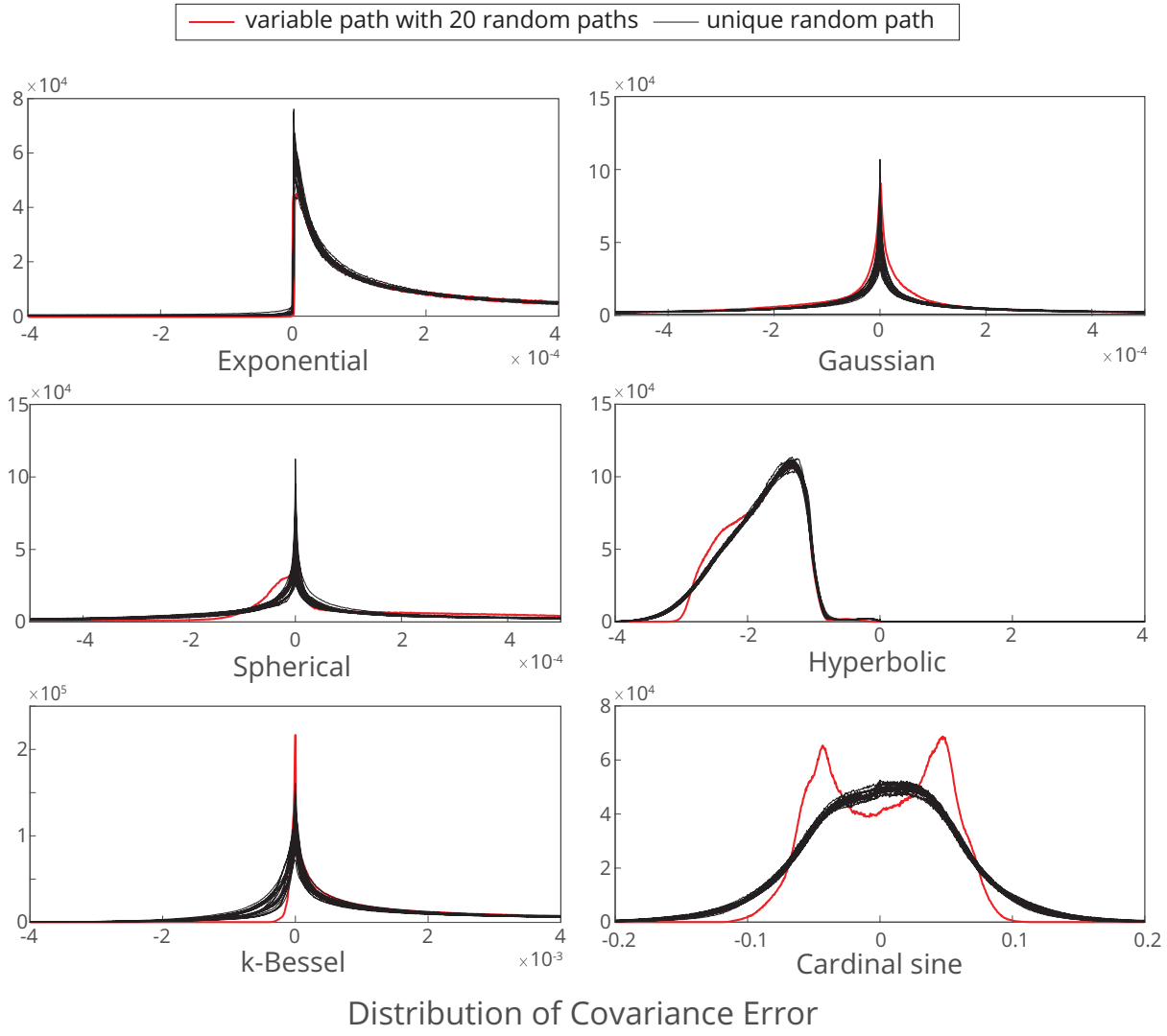


Figure 3.6 – Distributions of all values $\varepsilon_{\alpha,\beta}$ (equation 3.19) of 20 simulated RFs with a constant path (black lines) and of a simulated RF with a randomized path (red lines).

Each covariance function type produces different distributions, but in each of them, the 20 simulated RFs with a constant path generate similar distributions. In comparison, the simulated RF with a randomized path has generally a fewer number of covariance pairs with large and small errors. However, the number of intermediate errors is increased. This can be explained by the averaging of the covariance error for each pair of nodes in equation 3.20. Indeed, averaging tends to reduce extreme values and to increase the medium-range values. This means that, although extreme errors are avoided, the number of pairs of nodes with no errors is also reduced. Therefore, if instead of looking at the root mean square error, the number of correct pairs of nodes were computed, a simulated RF with a constant path would produce a more favourable result.

3.6 Conclusions

This study analysed in detail the benefits and limitations of using a constant simulation path in SGS. Compared to the traditional implementation of SGS, the use of a constant path provides computational gains of several orders-of-magnitude. While randomizing the path among realizations slightly reduces the covariance errors, this reduction is only significant for a limited number of randomized paths. Most importantly, our analyses demonstrated that, for the majority of our simulations, these error reductions were relatively small in comparison to the effects of increasing the neighbourhood size or, to a lesser extent, using a multi-grid path. The optimal approach with regard to computational efficiency and statistical fidelity is to switch from a randomized path to a constant path in the course of a multi-grid path. The timing of this switch should correspond to a grid spacing smaller than the covariance range. However, as this solution can be rather tedious to implement and given the relatively limited gains it provides in terms of error reduction, our recommendation is to use a fully constant path and to employ the associated computational gains to increase the neighborhood size. The constant path approach can readily be extended to other sequential simulation methods as long as some computationally expensive operations are the same for all realizations. In order to be common, such operations have to be independent of the data values, such as, for instance, the computation of the kriging weights or the neighborhood search. This includes

methods such as Sequential Cosimuation, Sequential Indicator Simulation (Isaaks, 1984) or Direct Sequential Simulation (Billings et al., 2002).

Chapter 4

Hydrogeophysical data integration through Bayesian Sequential Simulation with log-linear pooling

---

Raphaël Nussbaumer, Grégoire Mariethoz, Erwan Gloagen, Klaus Holliger

Under review in *Geophysical Journal International.*

Abstract

Bayesian sequential simulation (BSS) uses a secondary variable to guide the stochastic simulation of a primary variable. A key implicit assumption of BSS is the independence of the collocated secondary variable and the neighbouring points of the primary variable conditional to the simulated point. This assumption is, however, only valid for specific bivariate covariance models and generally not acceptable in geophysical applications. In the past, coping with this prerequisite has required to perform the simulation process in multiple steps, which resulted an algorithmically complex and computationally very expensive procedure. Here, we present an alternative concept, in which the aggregation of the primary and secondary variable is driven by a log-linear pooling approach. This allows accounting for the inherent dependence between primary and secondary variables using suitable aggregation weights. The dynamic adaptation of these weights for optimally exploiting the inherent information redundancy of the primary and secondary variable is an important and novel aspect of this study. The proposed method is tested and verified with regard to a realistic synthetic case study consisting of the integration of surface-based electrical resistivity (secondary variable) tomography (ERT) data acquired over a heterogeneous alluvial aquifer and local in situ measurements of the hydraulic conductivity (primary variable). We find (i) that, compared to traditional BSS, the use of constant aggregation weights greatly enhances the reproduction of spatial statistics in the realizations while maintaining a faithful correspondence with the geophysical data and (ii) that significant additional improvements can be achieved through dynamic adaptations of the weights in the course of the simulation process.

4.1 Introduction

Adequate knowledge of the hydraulic conductivity distribution within an aquifer is essential for the development of effective groundwater management and remediation strategies. Indeed, the hydraulic conductivity is critical for hydrological modelling as its spatial heterogeneity determines the flow and transport characteristics of the studied subsurface regions. The hydraulic conductivity can be measured at different scales: core analyses and slug tests provide information with regard to the small-scale heterogeneity, while well tests estimate averages over large volumes. The corresponding gaps in terms of resolution and coverage can be bridged by specifically targeted geophysical measurements (e.g., Rubin and Hubbard, 2005). Arguably, the most effective way to address this problem is through the quantitative integration of geophysical and hydraulic data (e.g., Hyndman and Gorelick, 1996) which, until recently, tended to be limited to small-scale studies involving high-resolution geophysical data (e.g., Dafflon et al., 2009). Here, we address the common and pertinent scenario of integrating highly resolved, but spatially very sparse in situ measurements of the hydraulic conductivity with spatially extensive, but poorly resolved tomographic images of a vaguely related rock physical property.

Geostatistics offers a toolbox of methods to stochastically populate a grid with a variable while respecting given spatial constraints as well as available observed data. Over the years, many geostatistical simulation methods have been developed, such as sequential Gaussian simulation (SGS) (Journel, 1989; Deutsch and Journel, 1992), turning bands (Journel, 1974), truncated pluri-Gaussian simulations (Mariethoz et al., 2009a), simulated annealing (Deutsch and Wen, 2000), and multiple-point geostatistics (Guardiano and Srivastava, 1993). Such methods are particularly adapted to the simulation of hydraulic conductivity. However, the addition of geophysical data becomes a challenging endeavour for these rigorous methods. Traditionally, geophysical data, considered as secondary variables in geostatistcal simulations, tend to be available over the entire domain considered, albeit often at a much coarser resolution than the mesh size used for the stochastic simulation of the primary variable. Such situations are typically addressed through collocated cosimulation (Xu et al., 1992; Chilès and Delfiner, 1999), although the inherent assumption regarding a linear relationship

between primary and secondary variables is often deemed as being unrealistic in practice (Gómez-Hernández and Wen, 1998; Zinn and Harvey, 2003).

Bayesian sequential simulation (BSS) (Doyen and Boer, 1996) provides a simple and reliable alternative to collocated cosimulation. The method is based on SGS, with the addition that the conditional distribution estimated by kriging is updated by the joint distribution between primary and secondary variables in a Bayesian framework based on the value of the collocated secondary variable. Arguably, a key asset of BSS is its fundamental flexibility regarding the relationship between the primary and secondary variable. While, as outlined above, standard cosimulation methods require both variables to be linearly related and to exhibit a homoscedastic behaviour (i.e., the correlation of both variables is constant on the range of values considered), BSS can handle virtually any relationship and does not require explicit knowledge of the corresponding cross-variogram. The method was initially applied to the simulation of lithoclasses based on seismic impedance measurements (Doyen and Boer, 1996). Later, Dubreuil-Boisclair et al. (2011); Ruggeri et al. (2013, 2014) adapted it to simulation of the fine-scale hydraulic conductivity structure guided by information provided by low-resolution surface-based electrical resistivity tomography (ERT) measurements. To our knowledge, this is the first approach of this kind that allowed hydrogeophysical data integration to be extended from the predominantly local to the sub-regional scale.

While BSS has been recognized as a useful and flexible tool (Ezzedine et al., 1999; Doligez et al., 2015; Chen et al., 2001), it also presents significant and, as of yet, unresolved challenges. In particular, the adequate reproduction of the variance and the fine-scale structure has proved to be been a difficult task. The reason for this is that Bayesian updating relies on the assumption of conditional independence between 1) the primary variable neighbours of the simulated location and 2) the collocated secondary variable. This assumption is referred to as conditional independence because the assumption of independence is conditional to the primary variable value at the simulated location. This assumption requires specific bivariate covariance models and is not valid in the general context of sequential simulation (Journel, 2002; Mariethoz et al., 2009b; Allard, 2018). In the context of the hydrogeophysical applications mentioned above, this assumption is adequate because regularization constraints applied in the course of the inversion of the geophysical measurements result in

the corresponding tomographic images to be overly smooth. To overcome biases and artefacts caused by this assumption of independence, Ruggeri et al. (2013, 2014) added several algorithmically complex and computationally expensive steps to BSS in order to ensure an adequate reproduction of the underlying statistics. The first step is to downscale the smooth electrical conductivity structure inferred through ERT to the desired resolution. A second step performs gradual deformation (Hu et al., 2001) to make sure that the downscaled electrical conductivity structure matches the original coarse-scale measurements while retaining the correctness of the overall fine-scale structure. Finally, a third step simulates the hydraulic conductivity based on the downscaled fine-scale electrical conductivity field. An additional challenge is that the resolution of the tomogram is varying with depth because of the surface-based electrode configuration used. Previous works (Ruggeri et al., 2013, 2014) dealt with this by adjusting the variance of the sampling of the joint probability density function (pdf) according to the sensitivity of the tomogram as computed by the inversion.

The scope of this study is to overcome these limitations through a novel strategy, which allows performing BSS-based hydrogeophysical data integration in one single step. In our approach, the redundancy of information between the primary and secondary variables is taken into account in the context of the probability aggregation framework (Genest and Zidek, 1986; Allard et al., 2012), which has already been used for various applications in the geosciences, such as the interpolation of satellite images (Mariethoz et al., 2009b), the reconstruction of 3D volumes based on 2D sections (Comunian et al., 2012), and the inclusion of auxiliary information in multiple-point geostatistical simulations (Hoffimann et al., 2017). Probability aggregation involves the definition of weights that account for redundancy between the various sources of information. In this paper, we test and compare different weighting strategies to account for the redundancy of information inherent to BSS with regard to a pertinent synthetic hydrogeophysical dataset at the sub-regional scale (Ruggeri et al., 2013).

The paper is structured as follows. We first provide a basic outline of the traditional BSS method (section 4.2). We then describe the origin of the assumption of independence between the primary and secondary variables inherent to BSS and propose a solution to overcome this limitation using probability aggregation (section 4.3). Finally, the newly proposed algorithm is tested and verified with regard to a pertinent synthetic case study (section 4.4).

4.2 Traditional BSS

The input data for BSS typically consist of a few highly resolved, but sparse measurements of the primary variable X, also referred to as hard data, and spatially exhaustive, but poorly resolved estimates of the secondary variable Z. No a priori relationship between the primary and secondary variables is assumed and, therefore, any joint distribution $P(X, Z)$ can be used. In previous works, this joint distribution was inferred from the hard data and corresponding collocated values of the secondary variable (Dubreuil-Boisclair et al., 2011; Ruggeri et al., 2013, 2014). The iterative procedure followed by BSS for each simulated node is the following (Figure 4.1):

1. An unpopulated cellX_i is randomly selected on the grid of the primary variable.
2. Using kriging, the conditional distribution $P(X_i \mid X_{<i})$ is computed based on the neighbouring cells, denoted as $X_{<i}$, which may include hard data and previously simulated cells. Note that, if X is not Gaussian, a normal-score transform is applied to the neighbours values prior to kriging and then the conditional distribution $P(X_i \mid X_{<i})$ is back-transformed to the non-Gaussian space.
3. The conditional distribution $P(X_i \mid Z_i)$is extracted from the joint distribution $P(X, Z)$ based on the known collocated secondary variable Z_i.
4. The two distributions $P(X_i \mid X_{<i})$ and $P(X_i \mid Z_i)$are combined using Bayesian updating.
5. A value for X_i is sampled from the resulting distribution.
6. The grid is updated with the new value.

4.3 Accounting for information redundancy

4.3.1 Independence in Bayesian updating

In the following, we show that, because of its sequential nature, BSS inherently relies on the assumption of conditional independence between the secondary variable and the previously

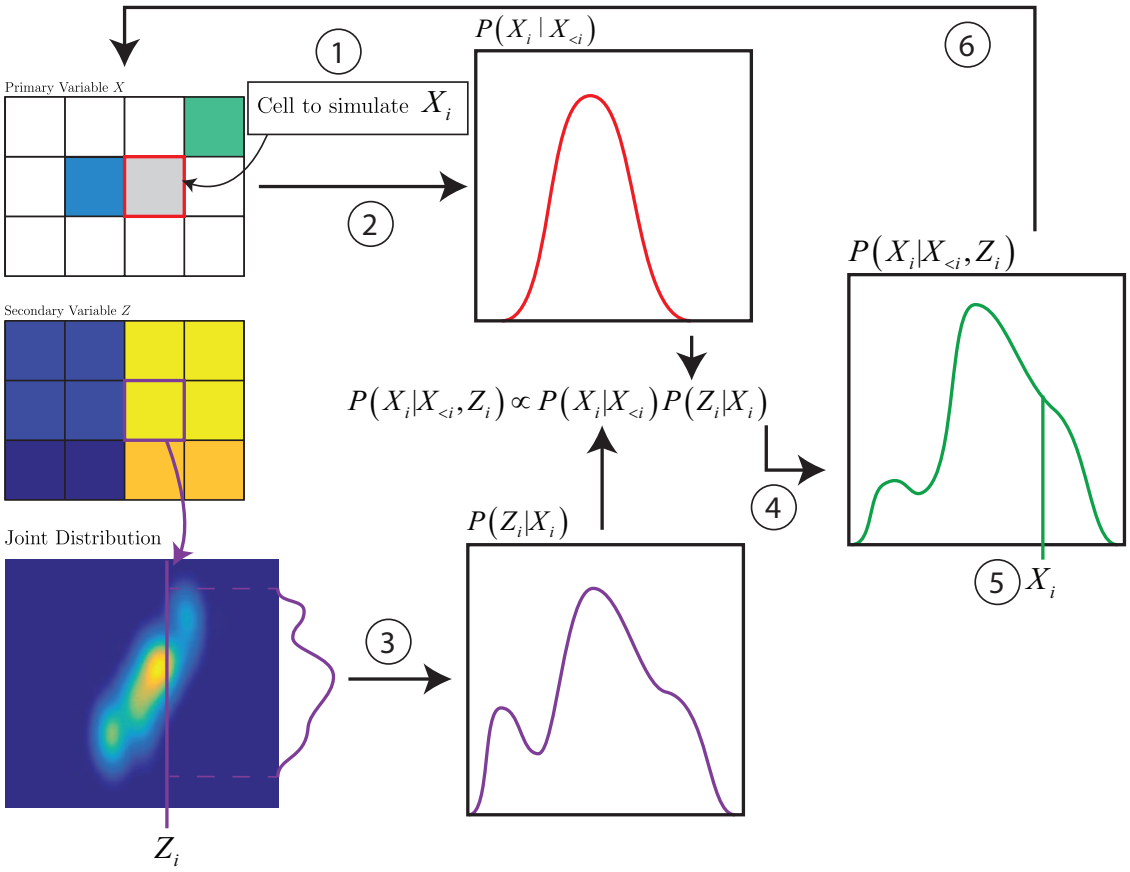


Figure 4.1 – Schematic illustration of the classical BSS approach: (1) Selection of the unknown cell to simulate X_i, (2) kriging estimate of the measured and previously simulated values of the primary variable $P(X_i \mid X_{<i})$ (3) estimation of marginal distribution from the joint probability distribution of the primary X_i and secondary Z_i variables, (4) determination of the posterior distribution, (5) random sampling of the posterior distribution, and (6) assigning the sampled value to the grid.

simulated values of the primary variable. Let us consider the conditional probability of an unknown event A based on several known events $\{D_1, \ldots, D_n\}$

$$P(A|D_1, \ldots, D_n) = \frac{P(A, D_1, \ldots, D_n)}{P(D_1, \ldots, D_n)}. \tag{4.1}$$

Substituting the numerator of Eq. (1) by the chain rule of probability

$$P(A_1, \ldots A_n) = \prod_{k=1}^{n} P(A_k|A_1, \ldots A_{k-1}), \tag{4.2}$$

leads to

$$P(A|D_1, \ldots, D_n) = P(A) \frac{\prod_{j=1}^{n} P\left(D_j|A, D_1, \ldots, D_{j-1}\right)}{P(D_1, \ldots, D_n)}. \tag{4.3}$$

In the context of sequential simulation, A represents the simulated variable X_i and the conditional data D is composed of the previously simulated nodes $X_{<i} = \{X_1, \ldots, X_{i-1}\}$

$$P(X_i|X_{<i}) = \frac{P(X_i) \prod_{j=1}^{i-1} P\left(X_j|X_i, X_{<j}\right)}{P(X_{<i})}. \tag{4.4}$$

Yet, in BSS, D also includes the collocated secondary variable Z_i, which leads to

$$P(X_i|X_{<i}, Z_i) = \frac{P(X_i) \prod_{j=1}^{i-1} P\left(X_j|X_i, X_{<j}\right)}{P(X_{<i}, Z_i)} P(Z_i \mid X_i, X_{<i}). \tag{4.5}$$

Substitution of the numerator of Eq. (5) with the numerator of Eq. (4) leads to

$$\begin{aligned} P(X_i|X_{<i}, Z_i) &= \frac{P(X_i|X_{<i})\, P(X_{<i})}{P(X_{<i}, Z_i)} P(Z_i \mid X_i, X_{<i}) \\ &\propto P(X_i|X_{<i})\; P(Z_i|X_i, X_{<i}) \end{aligned} \tag{4.6}$$

This equation states that the conditional probability $P(X_i|X_{<i})$ can be updated to incorporate a new source of information Z_i. Based on Eq.(6), BSS assumes that the value of the secondary variable Z_i and the previously simulated points of the primary variable $X_{<i}$ are independent

conditional to the simulated point X_i

$$P(Z_i|X_i, X_{<i}) = P(Z_i|X_i), \tag{4.7}$$

in order for the Bayesian updating to result in

$$P(X_i|X_{<i}, Z_i) \propto P(X_i|X_{<i})\, P(Z_i|X_i). \tag{4.8}$$

The conditional independence of equation (7) implies that all the information related to Z_i included in $X_{<i}$ is also included in X_i which relates to the so-called screening effect (Chilès and Delfiner, 1999). The resulting equation (8) can be referred to as a separable model because the primary and secondary variables are combined. Separability is defined as the ability to factorize the covariance of a variable along each dimension. For the considered bivariate case, this yields

$$\operatorname{cov}(X(\mathbf{u}), Z(\mathbf{u}+\mathbf{h})) = \frac{\operatorname{cov}(X(\mathbf{u}), X(\mathbf{u}+\mathbf{h})) \cdot \operatorname{cov}(X(\mathbf{u}), Z(\mathbf{u}))}{\operatorname{var}(X(\mathbf{u}))}. \tag{4.9}$$

Models following this form are known as Markov-type models and are equivalent to the model assumed when cokriging is simplified to collocated cokriging (Chilès and Delfiner, 1999, p. 305). An example of such a model is the proportional model (Matheron, 1965).

If this condition is not valid, the assumption of conditional independence becomes problematic in the course of the sequential simulation process because the neighbours $X_{<i}$ have already been influenced during their simulation by their corresponding collocated secondary variable $Z_{<i}$. The iterative procedure of BSS makes it especially sensitive to this assumption as any bias introduced is amplified during the simulation. Moreover, in the typical case where the secondary variable is significantly smoother than the primary variable, the spatial correlation of Z_i is significant and, consequently, the assumption of separability is even more difficult to justify. In a setting that will be presented in section 4.4.1, Figure 4.3.1 illustrates the conditional probability distribution function used in traditional BSS for three different nodes of the grid.

The first scenario (Figure 4.3.1a) corresponds to a node simulated at the beginning of the simulation path, which, therefore, is poorly informed by its neighbours. Moreover, this specific node is located at the bottom of the domain, where the secondary variable is not very informative. Consequently, both $P(X_i \mid Z_i)$and $P(X_i \mid X_{<i})$ produce estimates that are close to the marginal distribution. However, traditional BSS combines them as independent and generates a distribution with an artificially smaller variance than the marginal.

The second scenario (Figure 4.3.1b) also corresponds to a node simulated at the beginning of the process. However, this node is located near the top of the domain where the secondary variable is informative. In this case, kriging is poorly constrained, resulting in a distribution close to the marginal. In such a situation, traditional B.SS fails to neglect the kriging estimate.

The node corresponding to the third scenario (Figure 4.3.1c) is among the last ones to be simulated. Correspondingly, kriging provides a very well constrained estimation since the neighbours are close to the simulated node. In contrast, the secondary variable provides an estimation similar to the marginal distribution because the location of the node near the bottom of the domain makes the secondary variable poorly informative. In this case, traditional BSS is not able to determine that the secondary information has already been included in the neighbours and should not be included again.

4.3.2 Incorporating log-linear pooling into BSS

Probability aggregation provides a general framework to combine different estimations of an event A, each based on different data $\{D_1, \ldots, D_n\}$ with unknown dependences. A pooling operator F is defined to approximate the conditional distribution $P(A \mid D_1, \ldots, D_n)$ based on the individual conditional probabilities $P(A \mid D_i)$

$$P(A, D_1, \ldots, D_n) \approx F(P(A), P(A \mid D_1), \ldots, P(A \mid D_n)). \tag{4.10}$$

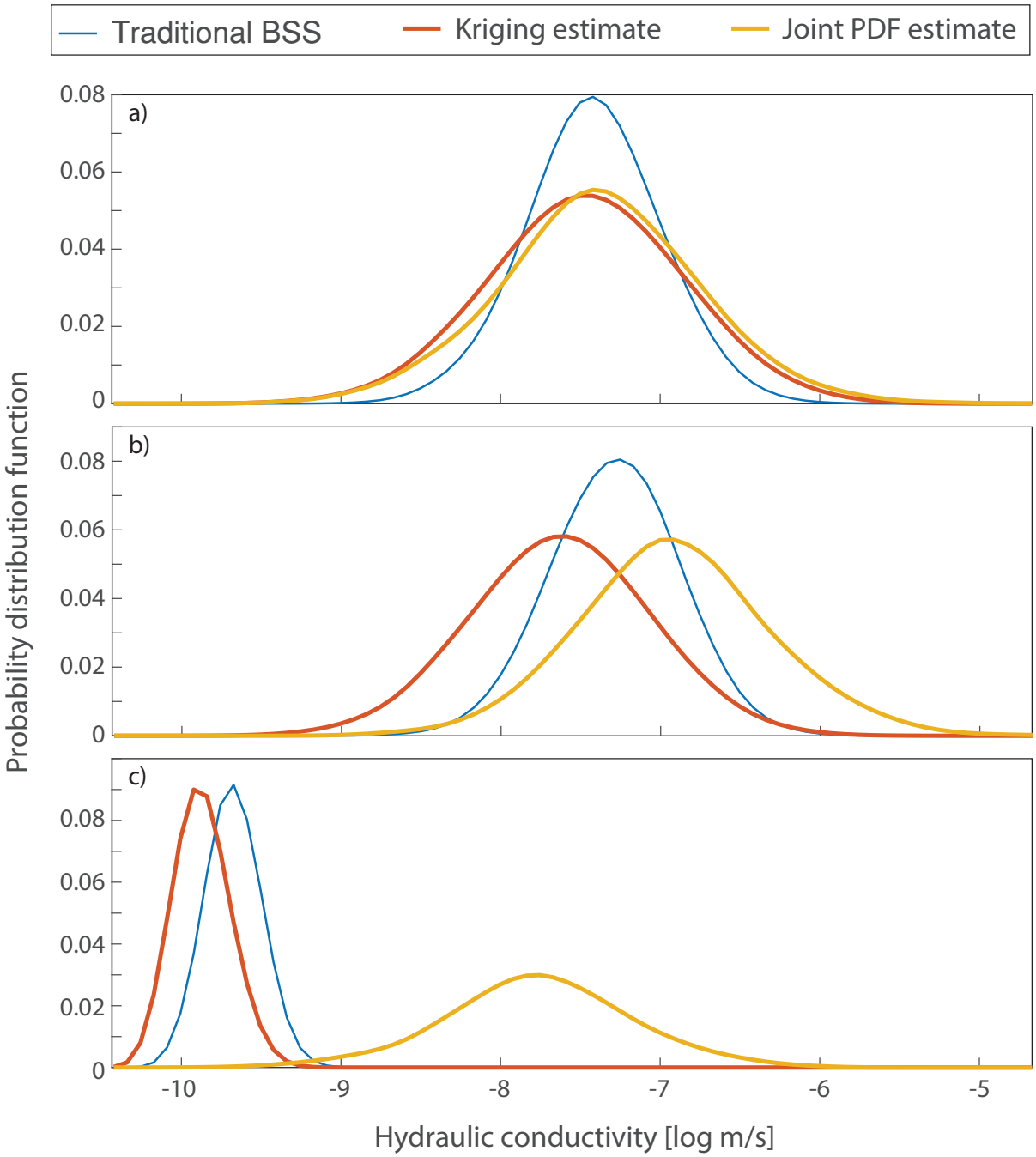


Figure 4.2 – Illustration of the effect of assuming conditional independence in traditional BSS for (a) a node simulated at the beginning of the sequential simulation at the bottom of the domain, (b) a node simulated near the surface and also at the beginning of the simulation and (c) a node simulated towards the end of the simulation.

Allard et al. (2012) describe several pooling operators$F$, of which the log-linear operator is one of the most general and popular approaches based on the multiplication of probabilities

$$P(A, D_1, \ldots, D_n) \propto P(A)^{1-\sum_i w_i} \prod_{i=1}^{n} P(A \mid D_i)^{w_i} \tag{4.11}$$

Allard et al. (2012) show that the log-linear pooling decomposition is exact, that is, it accounts for all dependences, when the weights w_i are computed as

$$w_i = \frac{\ln P(D_i|A, D_{<i})}{\ln P(D_i|A)} \tag{4.12}$$

Yet, as the nominator of Eq.(12) is usually unknown, the weights have to be approximated. If $w_i = 1$, the nominator and denominator of Eq. (12) become equal, such that the data $\{D_1, \ldots, D_n\}$ are assumed independent conditionally to A. This kind of aggregator corresponds to the so-called conjunction of probability (Tarantola, 2005).

If $\sum_i w_i = 1$, the prior distributionPA) vanishes in Eq.(11) and the pooling preserves unanimity. Unanimity implies that, if all individual conditional probabilities are equal, the pooling aggregation results in the same probability. In the context of BSS, log-linear pooling can be used to combine the information provided by kriging $P(X_i \mid X_{<i})$ and by the joint probability $P(X_i \mid Z_i)$ such that the assumption of independence is relaxed through the use of the weights w_X and w_Z. This amounts to rewriting Eq.(11) as

$$P(X_i \mid Z_i, X_{<i}) \propto P(X_i)^{1-w_X-w_Z} \cdot P(X_i \mid Z_i)^{w_Z} \cdot P(X_i \mid X_{<i})^{w_X} \tag{4.13}$$

Note that, in this paper, $P(X_i)$ is referred to as the prior, while $P(X_i \mid X_{<i})$, which is considered as the prior in the traditional BSS, is referred to as a conditional probability because it is based on the information of the neighbours. The prior can either be assumed to be unknown and chosen as a uniform distribution, thus having no influence, or it can be assumed to correspond to the marginal distribution.

4.4 Weighting scheme and numerical tests

In the following section, we illustrate the capacity of the proposed probability aggregation strategies for improving the inclusion of geophysical data in the simulation of hydraulic conductivity.

4.4.1 Experimental setting

To test our approach, we consider a synthetic case study, which closely follows that of Ruggeri et al. (2013). Figure 4.3 illustrates the procedure we used to generate the hydrogeophysical database. First, a Gaussian porosity field is produced with the Fast Fourier Transform Moving Average method (FFT-MA) (Le Ravalec-Dupin et al., 2000) on a 240x20 m grid with 513x65 nodes. This heterogeneous porosity distribution is characterized by a mean value of 0.27, a standard deviation of 0.05, and a theoretical 2D exponential variogram having horizontal and vertical ranges of 27 m and 2.7 m, respectively. Under the assumption of full saturation, the electrical conductivity field is then computed using Archie's law (Archie, 1942), with an electrical conductivity of the pore water of 43 mS/m and a cementation exponent of 1.4. The fine-scale hydraulic conductivity structure is obtained through an empirical power law relationship of the form Heinz et al. (2003). The large-scale electrical conductivity structure is inferred by tomographically inverting a synthetic ERT-type geoelectric electrical survey over the conductivity field using the software R2 Binley and Kemna (2005). Four equally spaced boreholes are placed on the grid providing detailed synthetic in situ measurements of the hydraulic conductivity.

In the context of BSS, the primary variable corresponds to the hydraulic conductivity and the secondary variable to the coarse-scale electrical conductivity structure interpolated on the grid of the primary variable. The joint distribution of the primary and secondary variable is traditionally built using only the collocated samples at the borehole locations. However, this assumes that numerous samples are available and representative of the whole area of interest. In practice, it can be difficult to estimate the joint distribution with a limited number of

samples. In this study, we therefore use an alternative approach, which consists of generating several unconditional simulations of both variables, which are then used as a basis to build the joint distribution. This allows quantifying the performance of our approach without contamination from other sources of uncertainty, in particular, a biased joint pdf resulting from a sub-optimal number of samples. Figure 4.4 schematically presents the approach used to generate the reference data.

A particular feature of the surface-based ERT is the variation of resolution with depth (Figure 4.4). Because of the inherent smoothing associated with the regularization of the inversion procedure and the decreasing resolution with depth, the resulting tomogram is computed on a relatively coarse grid, whose cell size increases depth. In order to use the ERT tomogram in BSS, a nearest-neighbour downscaling was performed to the resolution of the simulated model. However, the smoothness generated by the regularization of the inversion procedure has some complicating effects. The joint pdf is not able to account for the resulting non-stationarity of the secondary variable and, when sampled, produces estimates with larger variance for nodes with high resolution (close to the surface) and smaller variance for the nodes with lower resolution (near the bottom of the model). A possible avenue for addressing this issue could be to use a non-stationary joint pdf with, for instance, depth as the third variable. This is, however, a rather complex and largely self-contained endeavour, which goes beyond the scope of this work and should be retained as a topic for future research.

The neighbourhood search strategy consist of a two-part search with a spiral search of 40 nodes for the previously simulated nodes and a superblock search of 20 nodes for the hard data, if available Deutsch and Journel (1992). Superblock search finds the closest hard data by building a coarser grid and assigning at the beginning of the simulation all nearby hard data to cells of this supergrid. Finding the corresponding superblock of the simulated nodes allows directly obtaining the closest hard data. Spiral search looks for the closest previously simulated nodes by visiting all neighbouring nodes sorted by their distance to the simulated node, recording the existing neighbours until the predefined maximum number is reached. Despite its name, the search does not necessarily follow spiral path, as the sequence is fixed at the beginning of the simulation by sorting with Euclidian distances.

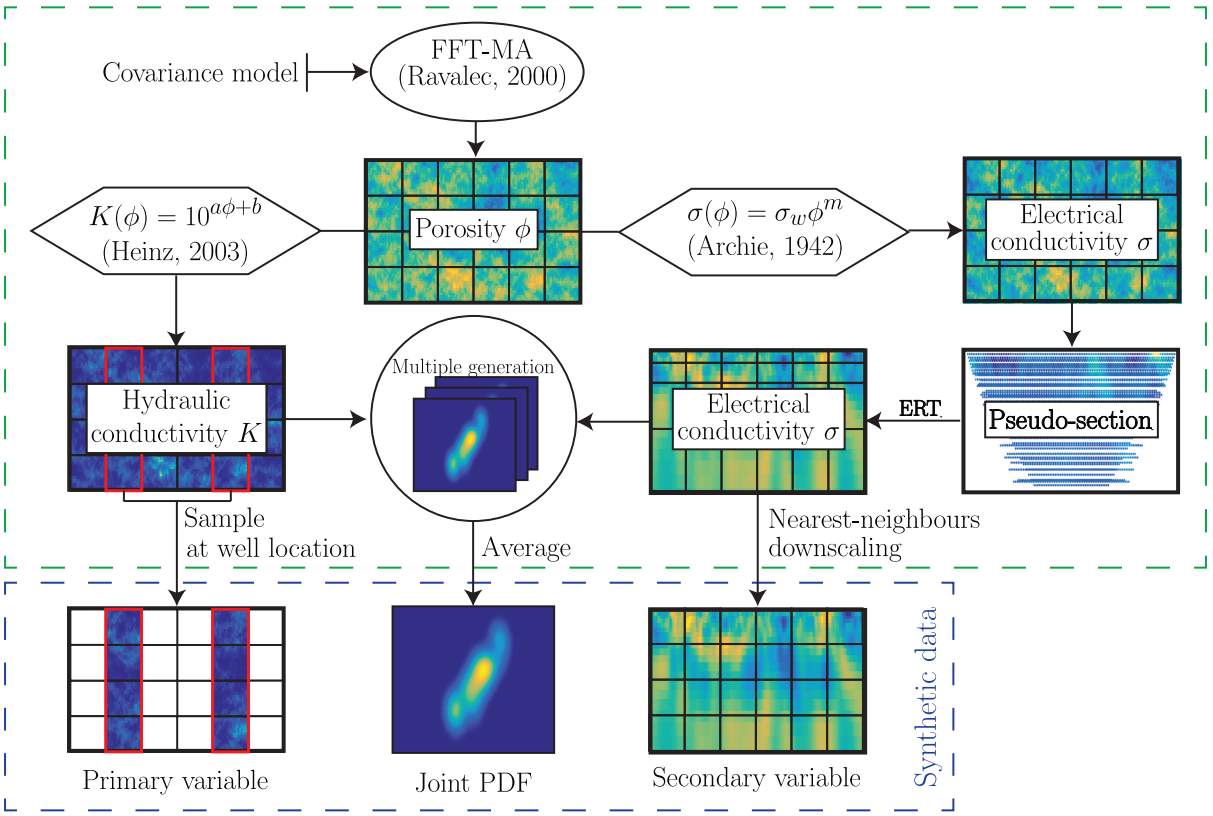


Figure 4.3 – Illustration of the procedure used to generate the synthetic hydrogeophysical database considered in this study. A heterogeneous porosity field is generated through FFT-MA and transformed into hydraulic and electrical conductivities. Estimation of the large-scale electrical conductivity structure is performed through surface-based ERT geoelectric measurements and their subsequent tomographic inversion. The fine-scale hydraulic conductivity structure is sampled along isolated boreholes.

The nodes are visited with a multi-grid path to minimize the effect of using a limited neighbourhood Nussbaumer et al. (2018a). Our implementation uses a constant path for several realizations, which allows to re-use the same kriging weights. This reduces drastically the computational cost while generating very small biases Nussbaumer et al. (2018b). The algorithm is parallelized over 48 cores using AMD Opteron processors (2300 MHz).

4.4.2 Simple weighting schemes

In this section, log-linear pooling is presented as a generalization of traditional BSS and the effect of using different weights is analysed. Table 4.1 presents a summary of the basic weighting schemes considered in this section.

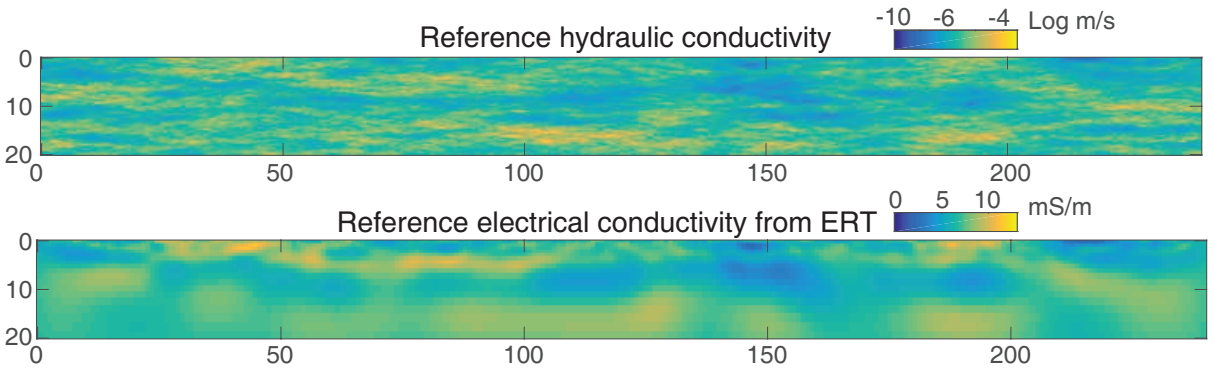


Figure 4.4 – (Top) Reference hydraulic conductivity field; (bottom) low-resolution electrical conductivity structure estimated from surface-based ERT.

Table 4.1 – Basic weighting schemes considered in this study.

	Aggregating $P(X_i \mid Z_i, X_{<i})$	w_X	w_Z
Traditional BSS	$P(X_i \mid Z_i) \cdot P(X_i \mid X_{<i})$	1	1
SGS	$P(X_i \mid X_{<i})$	1	0
White cosimulation	$P(X_i \mid Z_i)$	0	1
BSS-0.5	$\sqrt{P(X_i \mid Z_i) \cdot P(X_i \mid X_{<i})}$	0.5	0.5
BSS-1	$\frac{P(X_i \mid Z_i) \cdot P(X_i \mid X_{<i})}{P(X_i)}$	1	1

Traditional BSS is obtained when setting $w_Z = w_X = 1$ and assuming a uniform prior. For $w_X = 1$ and $w_Z = 0$, only the kriging information is used such that the procedure becomes identical to SGS. Conversely, $w_X = 0$ and $w_Z = 1$ results in a so-called white cosimulation, where only the information of the secondary variable is used. Additionally, we consider two simple weighting schemes, in the following referred to as BSS-0.5 and BSS-1, which honour the condition of log-linear pooling. Neither of these schemes applies any preference to $P(X_i \mid X_{<i})$ or $P(X_i \mid Z_i)$. The first scheme, referred as BSS-0.5, sets $w_Z = w_X = 0.5$ so that the prior is not included as $1 - w_X - w_z = 0$ and, consequently, the property of unanimity is preserved. The second scheme, referred to as BSS-1, sets the weights to $w_Z = w_X = 1$ and uses the marginal distribution of the primary variable as the prior with a weight of $1 - w_X - w_z = -1$.

Figure 4.4.2 takes the same examples as Figure 4.3.1 and illustrates the distribution $P(X_i \mid Z_i, X_{<i})$ resulting from aggregation with the weighting schemes described in Table 1. As SGS and white cosimulation result in the same distributions as the kriging estimate $P(X_i \mid X_{<i})$ and the joint distribution $P(X_i \mid Z_i)$, respectively, they are displayed by thicker lines.

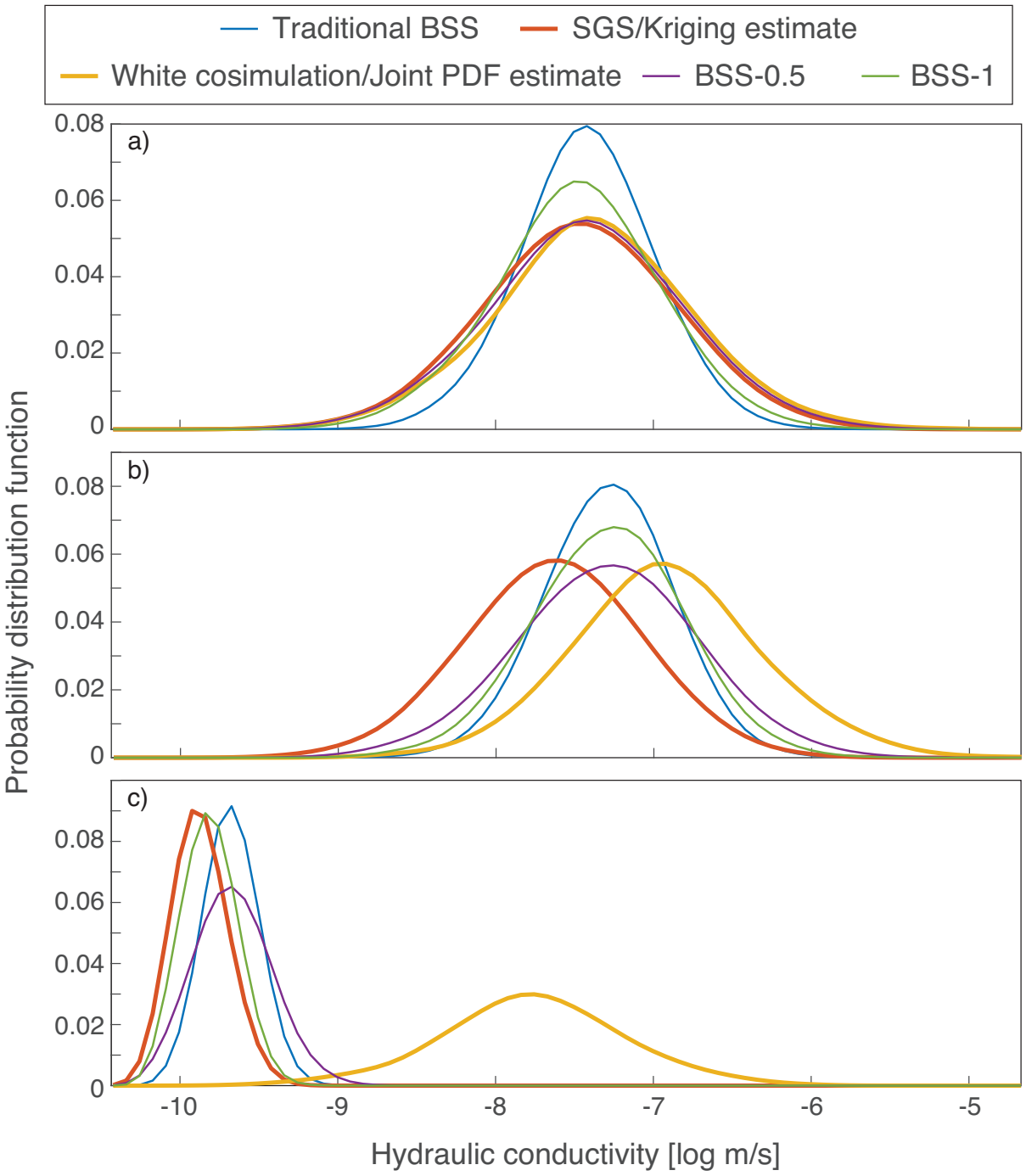


Figure 4.5 – Resulting probability distributions for the aggregation schemes considered in this study (Table 1) for three scenarios related to $P(X_i \mid X_{<i})$ and $P(X_i \mid Z_i)$: a) equal mean and variance, b) equal variance but different mean, and c) different mean and different variance.

In the first scenario (Figure 4.4.2a), the aggregation used in traditional BSS leads to a reinforcement of the central value as both sources of information are considered to be independent. The other aggregation methods mitigate this effect and result in estimates with larger variance. For instance, because of its unanimity property, the estimation provided by BSS-0.5 is almost identical to both $P(X_i \mid X_{<i})$ and $P(X_i \mid Z_i)$.

In the second scenario (Figure 4.4.2b), $P(X_i \mid X_{<i})$ and $P(X_i \mid Z_i)$have a similar variance, but a different expected value. This scenario shows how the various weighting schemes deal with such disagreement. If $P(X_i \mid X_{<i})$ and $P(X_i \mid Z_i)$ provide independent information, traditional BSS succeeds in finding the overlapping estimate. However, when there is interdependence, it overestimates the agreement between $P(X_i \mid X_{<i})$ and $P(X_i \mid Z_i)$ and, hence, supresses all values present in only one of the distributions. As a consequence, the estimate has a small variance and tends to the marginal distribution $P_0(X)$. BSS-1 corrects for this overlapping information by including the marginal distribution with a weight of -1 and produces an estimate with a larger variance.

The third scenario (Figure 4.4.2c) depicts a situation where $P(X_i \mid X_{<i})$ disagrees with$P(X_i \mid Z_i)$and has a smaller variance. Both BSS-0.5 and BSS-1 give estimates that are similar to$P(X_i \mid X_{<i})$ because they favour estimations with smaller variance. On the other hand, traditional BSS favours overlaps between estimations, which in turn results in more centred distributions. This explains why traditional BSS tends to underrepresent extreme values in the realizations.

To further illustrate the differences among these approaches, Figure 4.4.2 shows a single realization for each of the weighting schemed presented, Figure 4.4.2 compares their ability to reproduce the relation between primary and secondary variables with the joint distribution, and Figure 4.4.2 shows a comparison in terms of variograms to assess the ability to reproduce the underlying geostatistical model.

The simulations generated with SGS have a similar fine-scale structure as the reference field and, hence, the corresponding empirical variogram follows closely that of the reference (Figure 4.4.2). Yet, as the hard date along the boreholes represent the only constraint, the values of the field are varied freely between boreholes and, correspondingly, their joint distribution is not well reproduced (Figure 4.4.2). In contrast, white cosimulation offers the

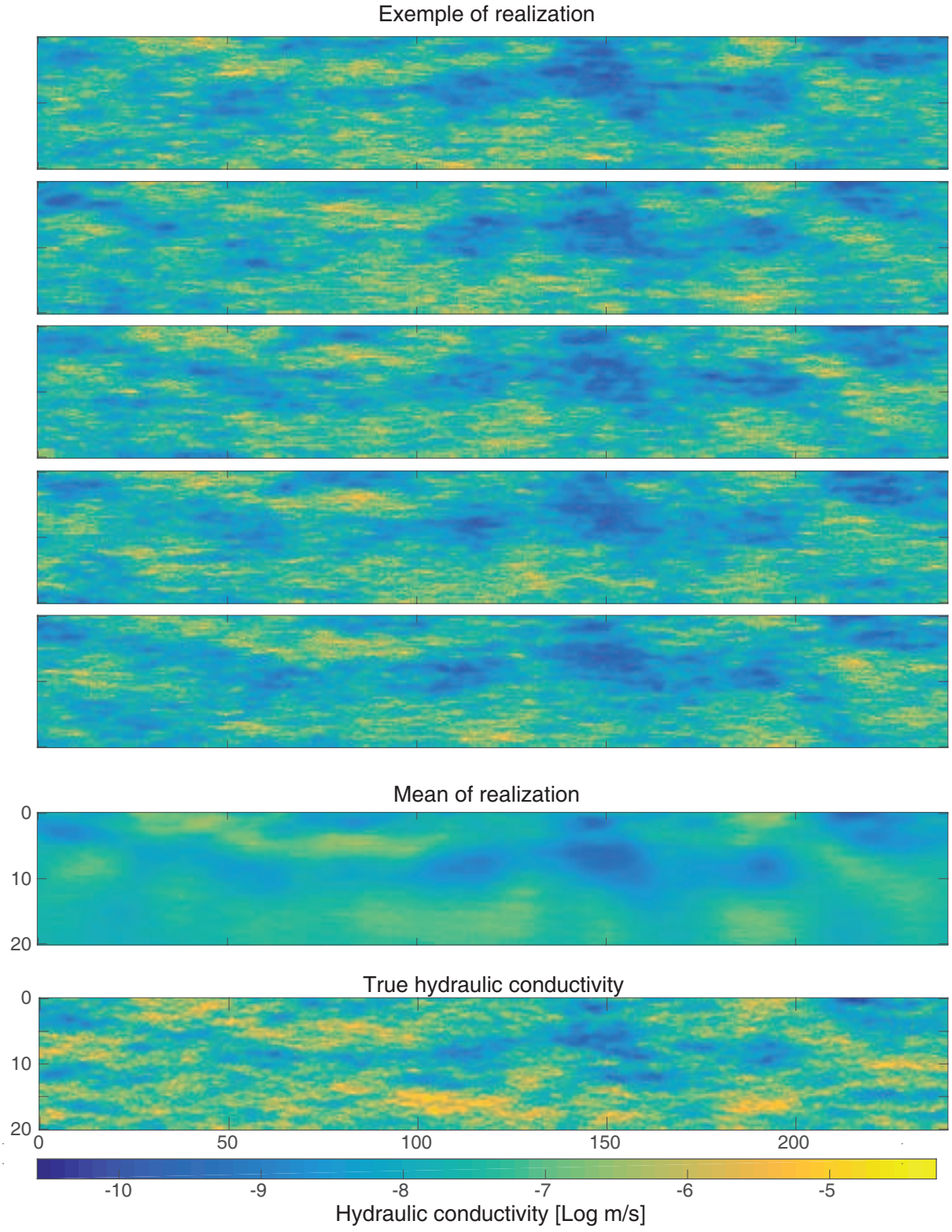


Figure 4.6 – Individual realizations for each of the scenarios described in Table 1

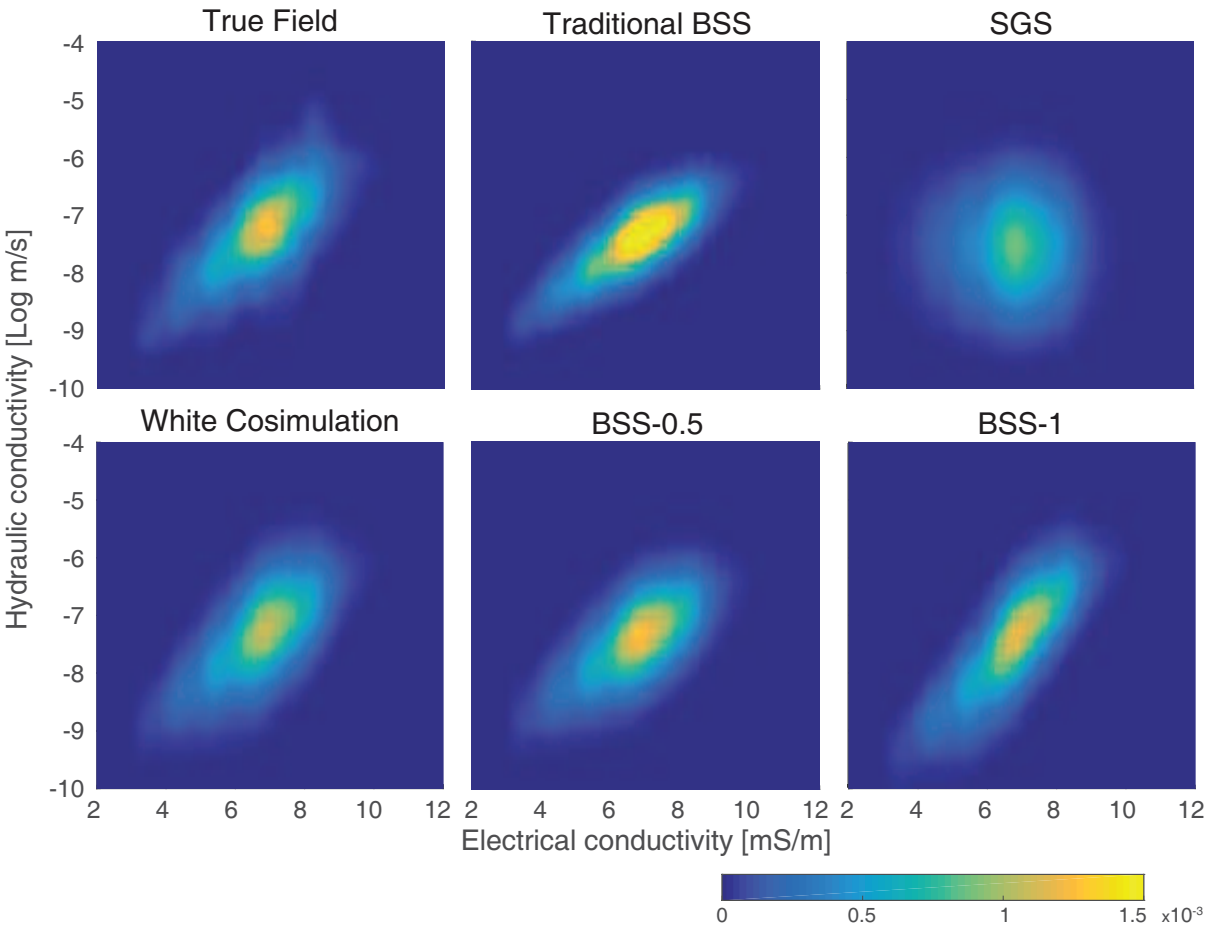


Figure 4.7 – Average of joint distributions errors of 480 realizations for each of the scenarios described in Table 1.

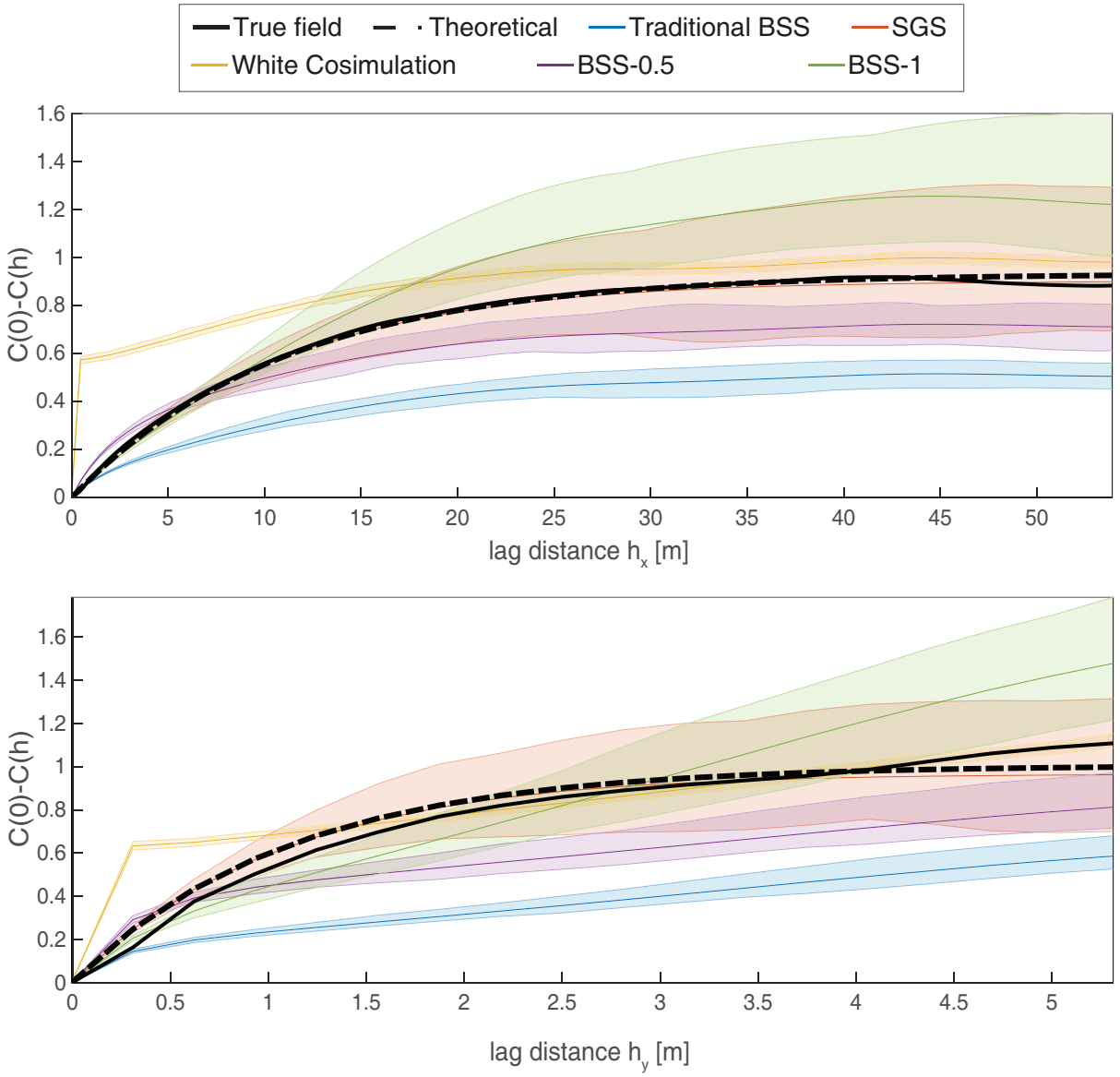


Figure 4.8 – Variograms of 480 realizations for each of the scenarios described in Table 1. The solid coloured lines denote the mean variograms of the 480 realizations, while the corresponding shaded areas correspond to the ranges of the variograms resulting from these 480 realizations. The variograms of the true field and of the underlying geostatistical model are shown as solid and dotted black lines, respectively.

best reproduction of the overall large-scale features (Figure 4.4.2), but ignores the fine-scale structure, which results in noisy realizations that fail to reproduce the target variogram at small lags (Figure 4.4.2).

Traditional BSS combines both sources of information and, therefore, reproduces the joint distribution better than SGS (Figure 4.4.2) and the variogram better than white cosimulation (Figure 4.4.2). However, as it assumes conditional independence of the primary and secondary variable, traditional BSS produces fields with a bias towards a lower variance, as illustrated in Figure 4.4.2. This effect can be observed in Figure 4.4.2 and results in a low variogram sill in Figure 4.4.2. Both log-linear pooling weighting schemes outlined in Table 1 (BSS-0.5 and BSS-1) correct for the information redundancy by limiting the sum of the weights to 1, either by setting both weights to 0.5 in the case of BSS-0.5 or by including the prior distribution with a weight of -1 in the case of BSS-1. These schemes improve the realizations with respect to the reproduction of the joint distribution (Figure 4.4.2) and the variogram (Figure 4.4.2). BSS-0.5 results in a slightly better joint distribution, while BSS-1 results in a slightly better reproduction of the target variogram, mainly for small lag distances.

4.4.3 Calibration of the weights

In order to quantitatively assess a chosen combination of weights, two objective functions are defined. The first objective function OF_X assesses reproduction of the spatial model and is evaluated by comparing the empirical variograms of all m realizations γ_j, $j = \{1,\ldots,m\}$ with the model variogram $\gamma_X(\mathbf{h})$. A root-mean square-type error (RMSE) of the variogram at each discrete vertical and horizontal lag distance $h_x$and h_y for lags up to two times the variogram range is used

$$OF_X = \frac{1}{K_x}\sqrt{\sum_{x=1}^{n_x} k(h_x)\left(\frac{1}{m}\sum_{i=1}^{m}\left(\gamma_i(h_x)-\gamma_X(h_x)\right)\right)^2} + \frac{1}{K_y}\sqrt{\sum_{y=1}^{n_y} k(h_y)\left(\frac{1}{m}\sum_{i=1}^{m}\left(\gamma_i(h_y)-\gamma_X(h_y)\right)\right)^2}, \tag{4.14}$$

where $K_i = \sum_{i=1}^{n_i} k(h_i)$ and $k(h_i) = 1-\gamma_X(h_i)$ weighs the misfit of each lag-distance based on its model variogram value so that short lags have more importance than longer ones. The

second objective function OF_Z evaluates the relationship between the realizations and the secondary variable. This is achieved by computing the discrete joint distribution $p_{x^{(i)}}(X,Z)$ between the realizations $x^{(i)}$ and the secondary variable and by comparing it to the joint distribution $p(X,Z)$ of the underlying model. The errors of each realization are averaged and the discrete joint distribution error is again aggregated with a RMSE estimate

$$OF_Z = \frac{1}{n_u n_v}\sqrt{\sum_{u=1}^{n_u}\sum_{v=1}^{n_v}\left(\frac{1}{n}\sum_{i=1}^{n} p_{x^{(i)}}\left(X=x_u, Z=y_v\right) - p\left(X=x_u, Z=y_v\right)\right)^2}. \tag{4.15}$$

Note that both OF_X and OF_Z do not depend on the true field so that they can readily be computed in real-world scenarios where the corresponding values are unknown.

The values tested for w_X and w_Z range from 0 to 2 and 96 realizations are used for each combination of weights. Figure 4.4.3 shows the resulting values of OF_X and OF_Z. The objective functions are normalized by the two end-member scenarios, SGS ($w_X = 1, w_Z = 0$) and white cosimulation ($w_X = 0, w_Z = 1$). As a result, a Pareto front is identified by minimizing the linear combination of the normalized objective functions weighted by a parameter t varying from 0 to 1.

$$\min_{w_X, w_Z} \quad t\frac{OF_X(w_X,w_Z) - OF_X(0,1)}{OF_X(1,0) - OF_X(0,1)} + (1-t)\frac{OF_Z(w_X,w_Z) - OF_Z(1,0)}{OF_Z(0,1) - OF_Z(1,0)} \quad \forall\, t \in [0,1]\,. \tag{4.16}$$

The Pareto front plotted as a red line on Figure 4.4.3 identifies the optimal weights depending on the relative importance given to each objective function. The continuum along the optimal line is discretised by 6 red dots in Figure 4.4.3 and each dot is illustrated by 2 realizations in Figure 4.4.3. This approach allows to explore various combinations of parameters and making an informed choice based on the importance of reproducing either the relation with the secondary variable or the spatial structure.

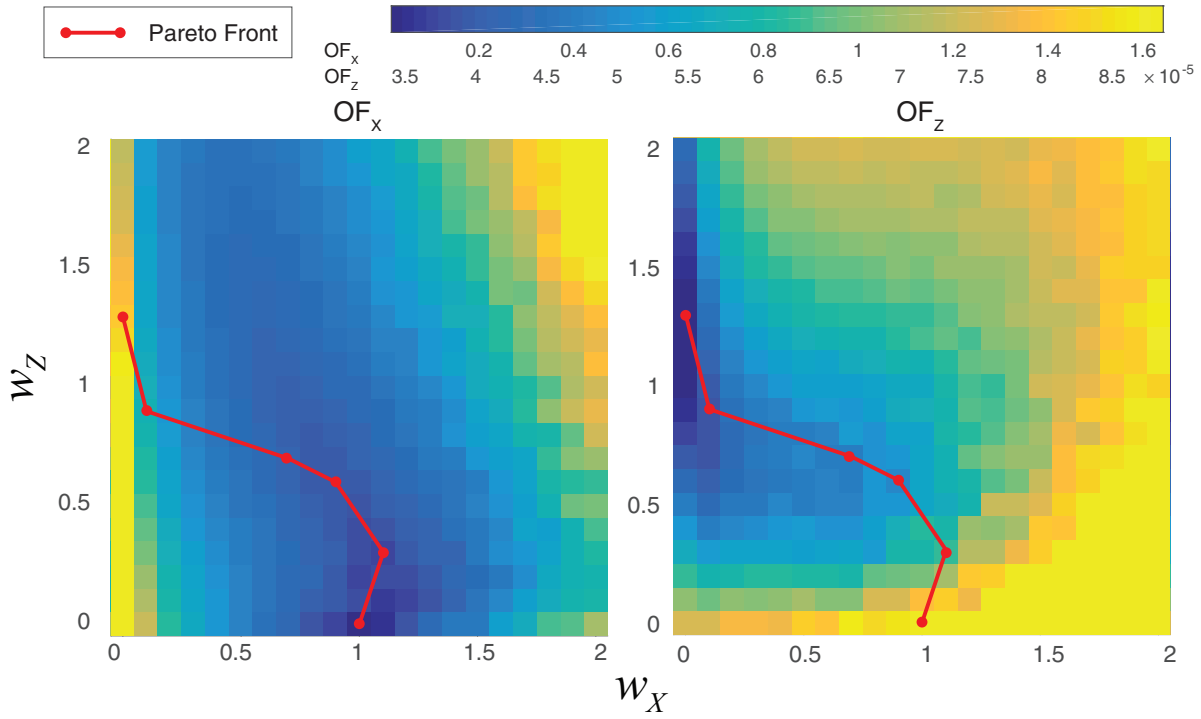


Figure 4.9 – Objective function values as a function of the weights used in the simulations. The colormap is scaled such that the minima and maxima correspond to the values of the objectif functions of SGS and white cosimulation, respectively.

4.4.4 Step weighting schemes

This section considers an alternative parameterization of the weights, whereby, instead of using constant weights, there is an abrupt switch from white cosimulation to SGS during the simulation. This strategy, which we refer to as Step-BSS in the following, is defined as

$$w_Z(i) = \begin{cases} 1 & \text{if} \quad i < T \\ 0 & \text{else} \end{cases} \qquad w_X(i) = 1 - w_Z(i) \tag{4.17}$$

where i denotes the node's order in the simulation path and T the switching threshold. The reasoning behind this approach has its origins in section 4.3.1, where the information redundancy in BSS is explained by the reuse of neighbouring values, which were previously simulated using the same secondary information. As the simulation proceeds, each newly simulated grid node thus compounds the amount of redundant information. This, in turn, implies that there is less redundant information for the first simulated nodes and more redundancy at the later stages of the simulation.

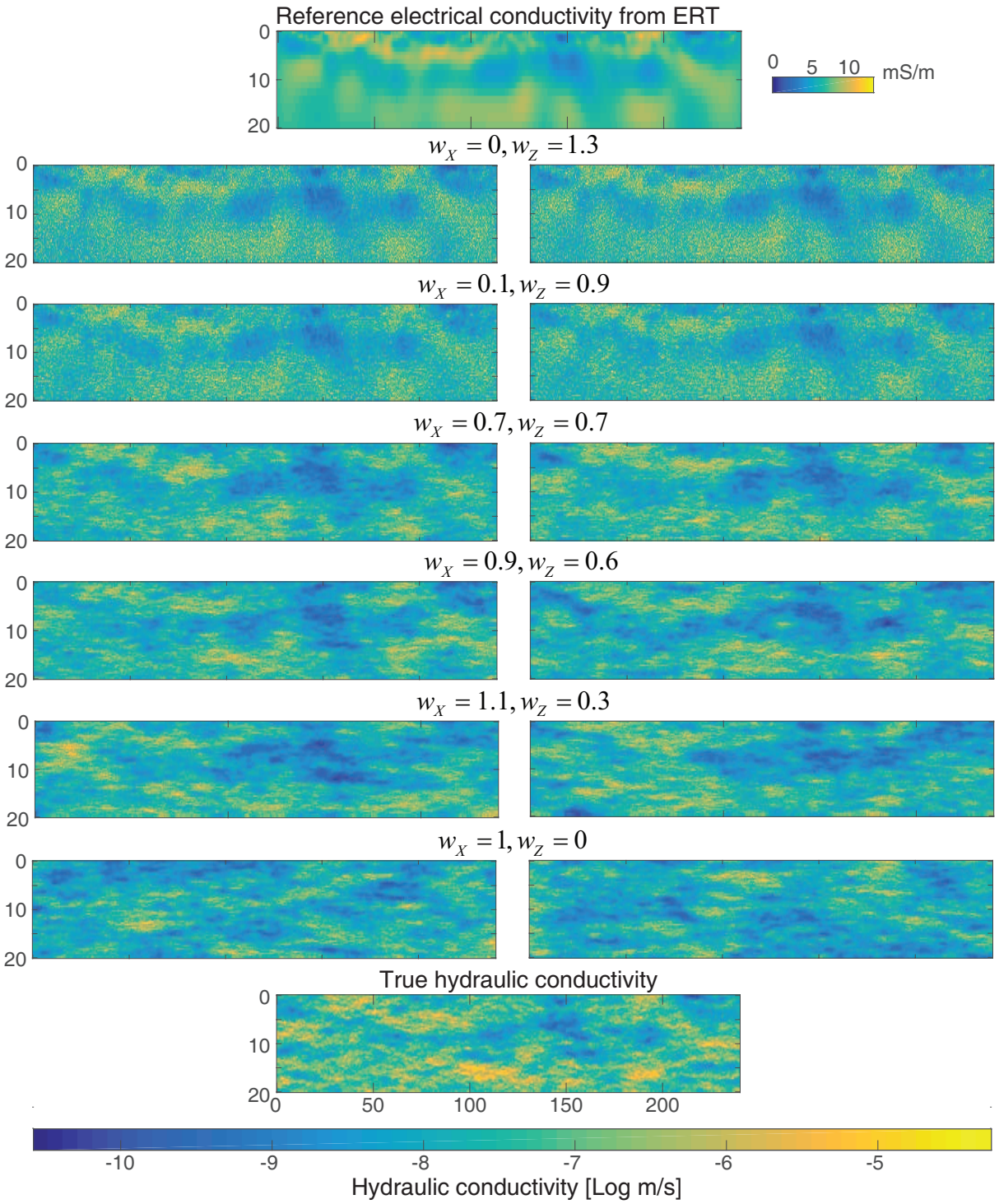


Figure 4.10 – Two realizations for different combinations of weights along the Pareto front.

This can be illustrated by further analysing the three scenarios shown in Figure 4.4.2. The first scenario (Figure 4.4.2a) corresponds to a situation where either of the two conditional pdfs could be used alone or they could be combined using a weighting scheme preserving unanimity. Including the prior in the in BSS-1 accounts for the marginal distribution of both sources of information and to reduce the overestimation. The second scenario (Figure 4.4.2b) using $P(X_i \mid Z_i)$is clearly more appropriate. In the third scenario (Figure 4.4.2c) it is better to rely on the kriging estimate to generate a realization respecting the spatial model.

These examples clearly motivate the use of a dynamic weighting scheme, which changes the aggregation weights as a function of the number of previously simulated nodes. The order of the simulation path i gives an indication of how many times the secondary variable has been included in the simulation, thus providing a proxy for the information redundancy.

This weighting scheme was tested against the ones presented in the previous sections. Figure 4.4.4 quantitatively compares all schemes considered in this study by representing as a scatter plot the values of both objective functions OF_X and OF_Z, quantifying to the reproduction of the variogram and joint pdf, respectively. The true field shows the smallest error for both objective functions, albeit with a small error in the reproduction of variogram, as quantified by OF_X, due to the ergodic fluctuations (e.g., Emery 2004).

Figure 4.4.4 illustrates that SGS and white cosimulation represent the two extreme scenarios with high values of OF_Z and low values of$OF_X$for SGS and vice versa for white cosimulation. In comparison, BSS-0.5 and BSS-1 are characterized by trade-offs between OF_Z and OF_X. Calibration of the weights using constant weights, denoted as Cst-BSS, allows for further exploring this trade-off along the lines of the Pareto front discussed in section 4.4.3.

The result of using a Step-BSS approach is similar to Cst-BSS as it also explores different optimal solutions depending on the relative importance given to OF_X and OF_Z. It is important to note that the reproduction of the joint pdf is improved by simulating only a few percent of the grid with white cosimulation. This can be explained by the fact that the first few simulated nodes constrain the large-scale structures and thus enforce a local correspondence between the simulated primary variable and the smooth secondary variable. In terms of objective functions, Step-BSS generates slightly better realizations for all combinations compared to

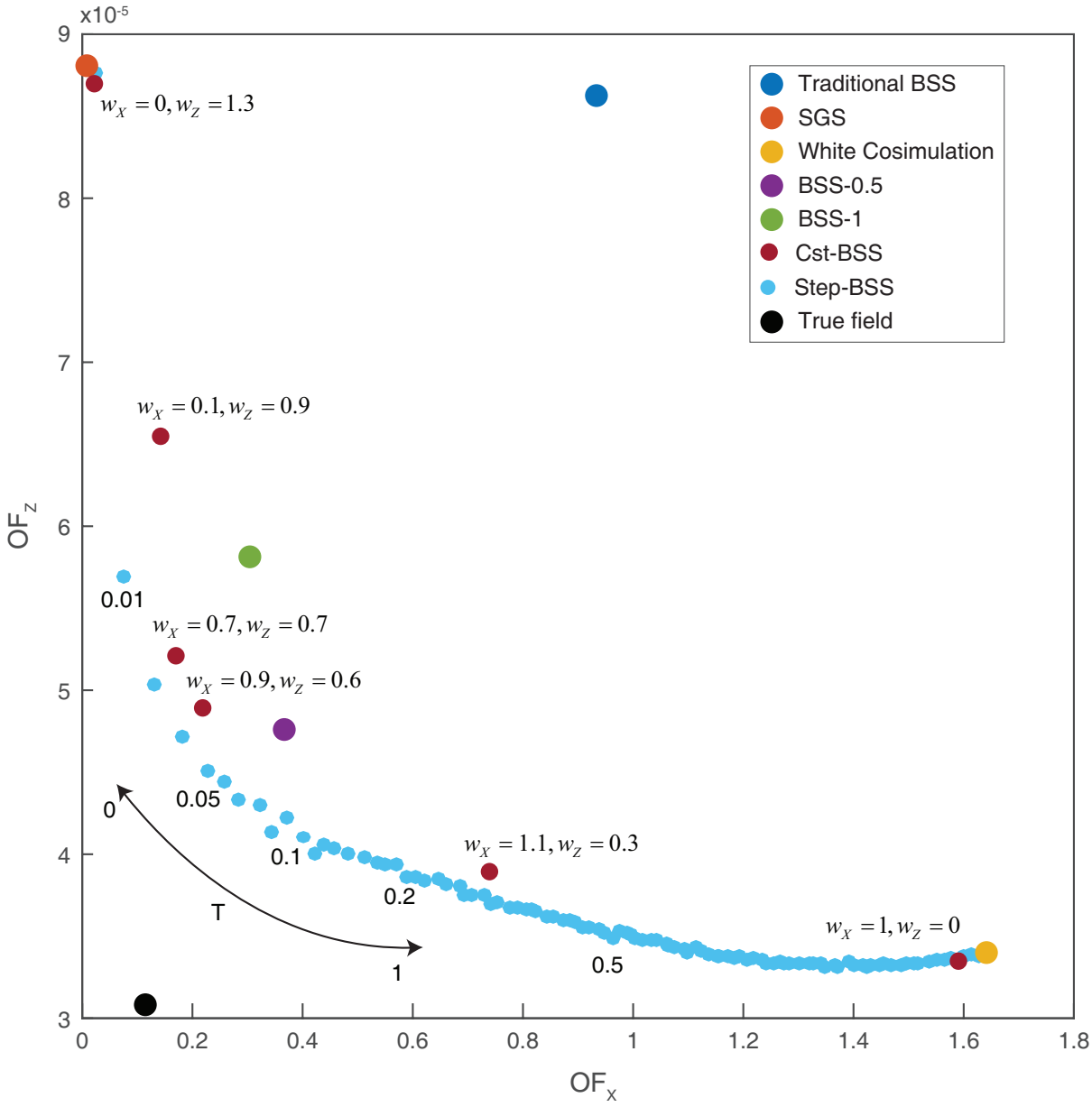


Figure 4.11 – Performance of different weighting schemes with regard to the reproduction of the variogram and the joint distribution, as quantified by the objective functions OF_X and OF_Z, respectively. For each weighting scheme, the objective functions are computed based on 400 realizations. The numbers associated with the Step-BSS simulations correspond to the proportion of the grid that is simulated before switching from white cosimulation to SGS.

the other weighting schemes explored. Instead of using a weight of 1 in Equation (18), various other values were also tested but not shown here because similar results were obtained for values ranging between 0.9 and 1.3, while simulations with values above or below this range resulted in the deterioration of both objective functions.

4.4.5 Multi-step weighting schemes

In this last section, a flexible weighting scheme is tested to explore the hypothetical existence of optimal weights. For this purpose, a multiple-step weighting scheme, in the following referred to as Multi-step-BSS, is defined using 10 parameters: 9 different step values located at 1%, 2%, 3%, 5%, 8%, 10%, 20%, and 50% of the simulation path and another parameter determining a constant sum of weights. The locations of the steps were chosen so as to divide the simulation path in a more or less equally log-spaced manner. Figure 4.4.5 illustrates the three weighting schemes with their respective optimal weights as functions of the simulation path order. A Metropolis-Hastings algorithm (Chib and Greenberg, 1995) is used to explore the posterior probability of the different parameters. The likelihood function is computed as

$$P(\mathbf{x}) = \exp\{-OF(\mathbf{x})/\sigma\}, \tag{4.18}$$

where the vector $\mathbf{x}$ comprises the 10 parameters of the weighting scheme, $OF(\mathbf{x})$ is the normalized sum of objective functions defined in Eq. (14) with $t = 0.5$ such that equal importance is given to each objective function, and $\sigma = 0.02$ corresponds to the errors in determining the objective function. The value of σ was "manually" calibrated such that the acceptance ratio of the Metropolis algorithm was around 40%. The generation of new candidates starts by defining the new sum of weights corresponding to the 10th parameter followed by the step values corresponding to the remaining 9 parameters. The proposal distribution is a Gaussian function with a variance of 0.05. The sum of weights is not constrained and can evolve freely, while the other 9 parameters are bounded between 0 and the sum of the weights. In order to keep a symmetric proposal, a modulus equal to the sum of the weights is applied on them. The initial parameters are chosen based on the optimal Step-BSS with $T = 0.06$.

The Metropolis-Hastings sampler was stopped after 2500 iterations with an acceptance rate of 52%. Figure 4.4.5 shows the accepted parameters $w_Z(i)$ of the Metropolis-Hastings sampler together with the corresponding value of the objective function. The minimum found with Multi-step-BSS is $OF = 0.13$ and is illustrated in Figure 4.4.5. In comparison, the best Step-BSS was found for $T = 0.06$ with $OF = 0.16$ and the best Cst-BSS corresponds to

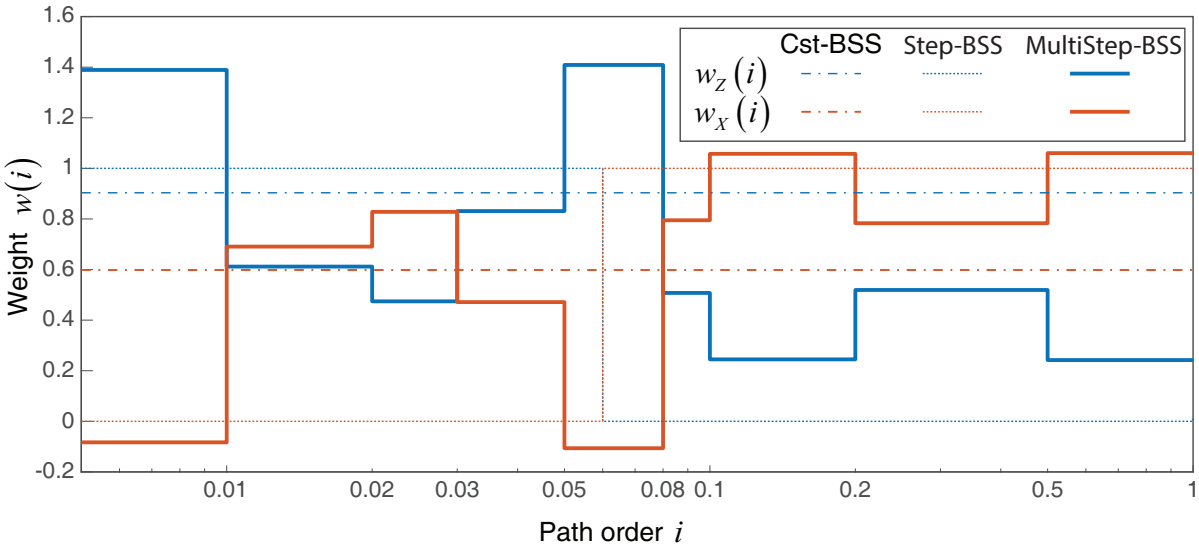


Figure 4.12 – Values of the weights as the simulation progresses for the three weighting schemes considered. Note the log-scale for the simulation path order. The values of the weights shown for the three schemes correspond to those minimizing the objective function of their relative weighting scheme.

$w_X = 0.9$, $w_Z = 0.6$ with $OF = 0.21$(Figure 4.4.5). The posteriori distribution of the accepted sum of weights (10th parameter) ranges between 1.2 and 1.4 and, thus, is similar to that of Cst-BSS. As hypothesised when using step-BSS, $w_Z(i)$ should decrease during the simulation, showing the redundant information contributed by the secondary variable. However, the best weightings schemes present a surprising oscillation in the values of consecutive steps. The origins of this phenomenon remain largely enigmatic at this point.

4.5 Conclusions

The objective of this study was to improve BSS by accounting for the notorious interdependence of the primary and secondary variables through the implementation of a log-linear pooling operator. To this end, a pertinent synthetic case study was considered involving the simulation of a fine-scale hydraulic conductivity field conditioned by highly resolved, but sparse in situ measurements of this property and by a coarse-scale, but spatially extensive estimate of the electrical conductivity structure from the tomographic inversion of a surface-based ERT-type geoelectric survey. Several weighting schemes for the considered log-linear

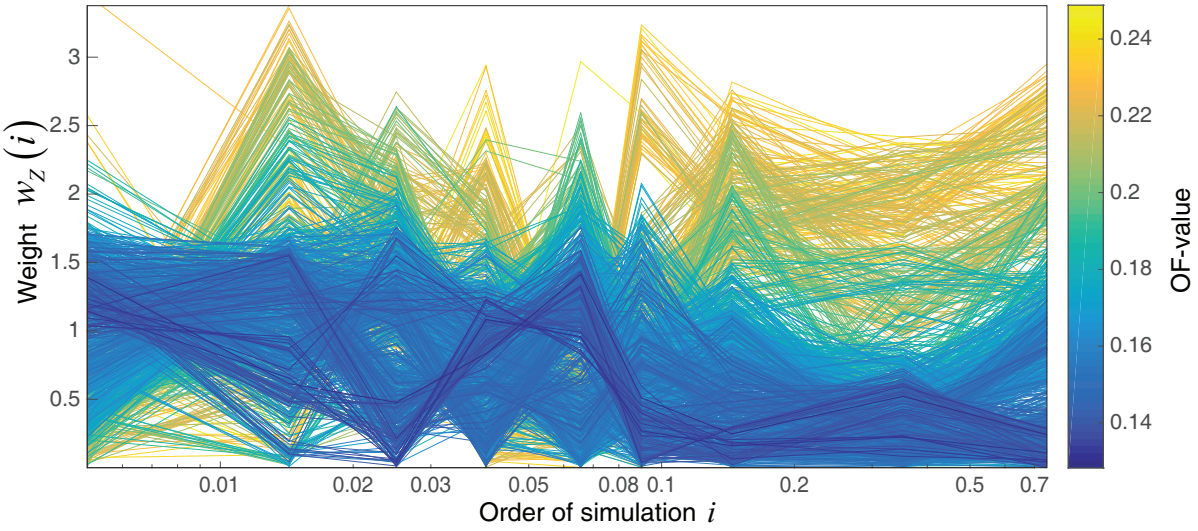


Figure 4.13 – Weighting schemes accepted by the Metropolis-Hastings sampling scheme together with their corresponding OF-values. Each parameterization is shown as a line rather than as a step for appropriate visualization.

pooling approach were compared by evaluating the ability of the simulation to reproduce both the spatial statistics and the joint distribution. It was first shown that using log-linear pooling with a constant weight of 0.5 or including the marginal distribution with a weight of -1 significantly improves the realizations compared to the traditional implementation of BSS. Then, we thoroughly tested all combination of the two weights to explore the optimal trade-offs between either honouring the joint distribution or the variogram structure. This allowed defining a Pareto front, along which subjective choices with regard to the relative importance of the two criteria can be made. Furthermore, we found that abruptly changing the weights with a step function during the simulation procedure proved to be slightly more efficient. Finally, a Metropolis-Hastings sampling algorithm was used to freely explore the space of all possible weighting schemes considering 9 different step values. This allowed improving the realizations further, albeit at the price of a heavy parameterization and a complex inference procedure. Designing and testing these different strategies clearly demonstrated that it is possible to reduce the redundancy of information with a flexible weighting scheme. Yet, the generalization of any quantitative result is limited because of the wide variability of possible scenarios, such as differences in resolution between the primary and secondary variable, the form of the joint distribution, the correspondence between the variograms of the primary and the secondary variable, or the location and quantity of hard data. Nevertheless, it is possible

to provide some fundamental recommendations. Our results suggest that the most attractive approach is to use a simple calibrated weighting scheme, such as Step-BSS or Cst-BSS. The primary reasons are that these types of schemes are computationally inexpensive with only one or two parameters to calibrate and able to account for a large part of the redundancy, while at the same time allowing to effectively explore trade-offs between constraints. For scenarios for which the calibration of the weights is difficult, for example due to computational limitations, our study indicates that traditional BSS should be avoided in preference of basic log-linear schemes, such as BSS-0.5 or BSS-1. Future research on this topic should consider putting the focus on stochastic simulation techniques which are capable of dealing explicitly with scale-dependent variables such as, for example, so-called change of support methods, thus taking into account the typically vastly differing scales and resolutions of the primary and secondary variables as well as the spatial dependency across variables. In addition, the influence of the secondary variable in the simulation of a primary variable should not be limited to collocated values but extended to neighbouring secondary values.

Chapter 5

Simulation of fine-scale electrical conductivity fields using resolution-limited tomograms and area-to-point kriging

---

Raphaël Nussbaumer, Niklas Linde, Grégoire Mariethoz, Klaus Holliger

Under review in *Geophysical Journal International*.

Abstract

Deterministic geophysical inversion approaches yield tomographic images with strong imprints of the regularization terms required to solve otherwise ill-posed inverse problems. While such tomograms generally enable an adequate assessment of the larger-scale features of the probed subsurface, the finer-scale details tend to be unresolved. Yet, representing these fine-scale structural details is generally desirable and for some applications even mandatory. To address this problem, we have developed a two-step methodology based on area-to-point kriging to generate fine-scale multi-Gaussian realizations from smooth tomographic images. Specifically, we use a co-kriging system, in which the smooth, low-resolution tomogram is related to the fine-scale heterogeneity through a linear mapping operation. This mapping is based on the model resolution and the posterior covariance matrices computed using a linearization around the final tomographic model. This, in turn, allows for analytical computations of covariance and cross-covariance models. The methodology is tested on a highly heterogeneous synthetic 2D distribution of the electrical conductivity that is probed with a surface-based electrical resistivity tomography (ERT) survey. In doing so, we demonstrate the ability of this technique to reproduce a known geostatistical model characterizing the fine-scale structure, while simultaneously preserving the large-scale structures identified by the smoothness-constrained tomographic inversion. Finally, we assess the ability of the resulting fine-scale stochastic realizations of the electrical conductivity to reproduce the electrical resistances measured in the underlying ERT-type geoelectric survey.

5.1 Introduction

Most geophysical inversion approaches are formalized as deterministic smoothness-constrained problems, in which solutions are sought that maximize the regularization weight under the constraint that the observations are fitted to a pre-defined noise level (DeGroot-Hedlin and Constable, 1990, e.g.,). In such approaches, the subsurface is discretized with cells that are much finer than the resolution of the resulting tomographic images. An advantage of this inherent over-parameterization is that the resolved features in the tomograms are essentially independent of the model parameterization. For example, in the case of electrical resistivity tomography (ERT), which relies on multiple, partially overlapping surface- and/or borehole-based measurements of electrical resistances to infer the electrical resistivity distribution of the probed subsurface region, a smoothness-constraining regularization term ensures uniqueness of an otherwise ill-posed problem (Binley and Kemna, 2005, e.g.,). However, this comes at the cost of tomographic images that only retain the larger-scale structures that are resolved by the data. The absence of the finer-scale heterogeneity can be problematic, notably in the context of aquifer and reservoir characterization where tomographic images of geophysical properties are commonly used to extrapolate point-type measurements of the hydraulic target properties (Rubin et al., 1992; Copty et al., 1993; Hubbard et al., 2001, e.g.,). Accounting for small-scale fluctuations in hydraulic conductivity is indeed essential to appropriately describe transport and mixing processes (Dentz et al., 2011, e.g.,). In this context, geostatistical methods can be helpful to impose small-scale properties on the solution under the assumption of a known underlying covariance model. There are two general approaches to reproduce the fine-scale structure in the geophysical images: modifying the inverse problem or applying downscaling methods a posteriori.

The first and, arguably, most generic approach is to replace the deterministic smoothness-constrained solution and find alternative methods that directly account for the geostatistical model of interest. In the case of a linear forward problem, Gaussian linear inverse theory (Hansen et al., 2006) can be used to derive the mean model and the covariance of the model parameters. This estimate is closely related to the cokriging estimate Gloaguen et al. (2005, 2007). Simulations can then be generated with Gaussian error simulation (Journel and Hui-

jbregts, 1978; Gloaguen et al., 2004) or sequential simulation (Hansen and Mosegaard, 2008). In the case of non-linear problems, a probabilistic formulation of the inverse problem can be used to generate an ensemble of model realizations that collectively characterize a posterior probability density function (pdf). By relying on advanced global search strategies, such as, for example, Markov chain Monte Carlo algorithms, it is possible to consider very generic classes of geostatistical models (Ramirez et al., 2005; Hansen et al., 2012; Linde et al., 2015, e.g.,). The associated computational costs do, however, tend to be very high. Using a least-squares measure of data fit and assuming a stationary multi-Gaussian prior, it is possible to solve a non-linear optimization problem based on iterative linearization, which is reminiscent of classical deterministic inversion, albeit with a solution that enables multiple model realizations to be obtained that are consistent with a geostatistical covariance model (Tarantola and Valette, 1982; Kitanidis, 1995; Yeh et al., 2002; Englert et al., 2016). Here, a covariance model essentially replaces the regularization term and the regularization weight is unity. This technique is widely used for addressing hydraulic inverse problems Kitanidis and Vomvoris (1983); Hoeksema and Kitanidis (1984); Li et al. (2005), but is seldom considered in geophysical inversion methods, which generally rely on smoothness-constrained deterministic techniques.

The second approach is to either downscale the tomogram resulting from the deterministic smoothness-constrained inversion, based on some ad hoc method or, more commonly, to directly simulate the property of interest as a primary variable and using the tomogram as a secondary variable. For example, McKenna and Poeter (1995) and Cassiani et al. (1998) use seismic tomograms as secondary variables in co-kriging systems to estimate the hydraulic conductivity and hydrofacies, respectively. These approaches involve fitting variograms and cross-variograms to known hard data. Doyen and Boer (1996) and Chen et al. (2001) do not fit a cross-variogram model, but instead, construct a joint pdf of the primary and secondary variables with collocated hard data. During the simulation of the primary variable, the kriging estimation is updated with the joint pdf sampled conditionally to the known collocated secondary variable. However, the difference in resolution between the primary and secondary variables is not explicitly taken into account in these studies. Ruggeri et al. (2013, 2014) use a two-step approach to first downscale an ERT image and then use the resulting fine-scale

information to simulate stochastically the associated small-scale distribution of the hydraulic conductivity. This approach does, however, suffer from an inability to accurately account for the conditional dependence of information (Mariethoz et al., 2009b). The basic underlying problem is that the model estimates of the tomograms are treated as independent data, while the number of degrees of freedom is much smaller than the number of model parameters. The impact of this loss of resolution inherent to smoothness-constrained inversions has been extensively studied in the hydrogeophysics literature, thereby highlighting the need to properly account for this phenomenon when inferring hydraulic properties or state variables Moysey and Knight (2004); Day-Lewis and Lane (2004); Moysey et al. (2005); Singha et al. (2007). Our study focuses on the second approach with the geophysical inversion being performed using smoothness constraints. Since many geophysical software only provide regularizations in terms of smoothness or damping constraints, the second approach becomes essential if the output of such codes is to be used in a multi-Gaussian framework.

The objective of this study is to stochastically downscale the smooth tomogram by adding fine-scale structure consistent with a known covariance model, while accounting for the resolution limitations of the tomogram. In the geostatistical community, several methods exist for downscaling a geostatistical variable. This is known as the change of support problem (Gotway and Young, 2002; Atkinson, 2013). Among these methods, area-to-point kriging (Kyriakidis, 2004) is an adapted co-kriging technique for the particular case when the known secondary variable is an upscaled description of the unknown primary variable. Kyriakidis (2004) shows that the area-to-point estimator is unbiased with the minimum error variance. Furthermore, it is a coherent estimation, that is, upscaling the area-to-point estimation of the primary variable reproduces the secondary variable. Kyriakidis and Yoo (2005) extended this method to generate stochastic realizations through Gaussian error simulation. This approach is equivalent to the one presented by Liu and Journel (2009). The method has been further adapted to include inequality constrains (Yoo and Kyriakidis, 2006) and was applied to population density downscaling (Liu et al., 2008) as well as to house price models with an external drift (Yoo and Kyriakidis, 2009). It has been combined with point data within the area-and-point kriging framework Goovaerts (2010, 2011). It was also extended to the multivariate case in the context of satellite image downscaling Pardo-Igúzquiza et al. (2006);

Atkinson et al. (2008); Pardo-Iguzquiza et al. (2010). In order to be able to use area-to-point kriging to downscale the tomogram, a linear relationship between the target fine-scale realization and the large-scale and smooth tomogram needs to be quantified. More specifically, the smoothing resulting from the inversion can be assessed by traditional appraisal tomographic method (i) the model resolution matrix, which describes the averaging filter relating the resulting tomographic image to the "true" subsurface structure, and (ii) the posterior model covariance matrix, which quantifies parameter errors and their correlation (e.g., Friedel, 2003; Menke, 1989). For the predominant non-linear case, the same theoretical framework can be readily applied for a local uncertainty assessment by linearizing the problem around the final model (Alumbaugh and Newman, 2000). These matrices have the potential to bridge the gap to use area-to-point kriging with the tomogram. In this work, we describe a methodology to generate stochastic fine-scale realizations of electrical conductivity based on deterministic smoothness-constrained ERT images using area-to-point kriging. In doing so, we account for the averaging process inherent to the deterministic inversion process by incorporating the information from the resolution and posterior model covariance matrices. We focus on electrical conductivity because of the overriding importance of this rock physical property and of ERT-based imaging for a wide variety of important applications. It is, however, important to note that the methodological framework presented in this study can be readily applied to any other smoothness-constrained geophysical inverse problem. The paper is structured as follows. In section 2, the methodological background is presented by introducing smoothness-constrained geophysical inversion and area-to-point kriging followed by an elaboration of how to combine them. Section 3 then presents the testing and assessment of the proposed developed technique with regard to a pertinent synthetic case study.

5.2 Methodology

5.2.1 Problem set-up

Figure 5.1 shows the fine-scale heterogeneous electrical conductivity field $\boldsymbol{\sigma}^{\text{true}}$ (Figure 5.1a) considered in this study together with the corresponding result (details follow later) at a coarser scale of a deterministic smoothness-constrained ERT-type inversion $\boldsymbol{\sigma}^{\text{est}}$ (Figure 5.1b). Please note that, in the following, uppercase characters are used when referring to the large-scale variable, while lowercase characters are used for the fine-scale variable. Bold symbols denote vectors or matrices. In our formulation of the problem, wireline-type high-resolution sampling of hard data from $\boldsymbol{\sigma}^{\text{true}}$ is performed along a borehole, the location $\mathbf{u}^{\text{hd}}$ of which is denoted by a red line. Comparing Figures 5.1a and 5.1b illustrates the generally lower and spatially variable resolution of the ERT image. Notably, the systematic decrease of the resolution with depth is clearly evident. Retrieving the underlying fine-scale structure from such a tomogram, which can be regarded as representing upscaled information with a spatially variable support, is an inherently non-unique problem. Indeed, this particular smooth upscaled representation of the underlying electrical conductivity structure could have arisen from a wide variety of fine-scale structures. In this study, we attempt to provide fine-scale electrical conductivity fields denoted as $\boldsymbol{\sigma}_{\text{c}}^{\text{sim}}$ constrained to local in situ measurements $\boldsymbol{\sigma}^{\text{true}}\left(\mathbf{u}^{\text{hd}}\right)$ and the smooth ERT image $\boldsymbol{\sigma}^{\text{est}}$ under the assumption of a known geostatistical model.

5.2.2 Geophysical inversion and model appraisal

A forward model $f()$ is used to simulate the response of a given electrical conductivity distribution in terms of the electrical resistances $\mathbf{r}^{\text{obs}}$ measured with the electrodes located on the surface, based on which the inversion procedure estimates the distribution of the electrical conductivity $\boldsymbol{\sigma}^{\text{est}}$ (in logarithmic scale) throughout the probed subsurface. Inverse problems of this type tend to be ill-posed and, hence, are typically solved with a regularized

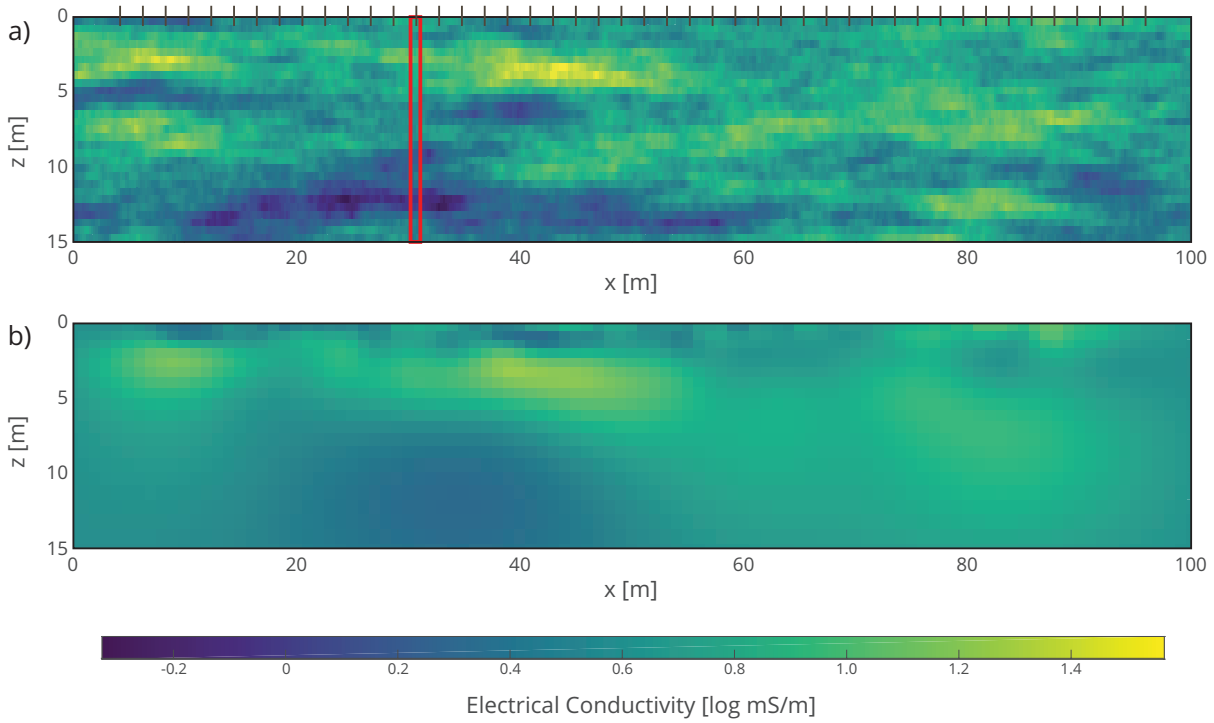


Figure 5.1 – (a) True fine-scale electrical conductivity field and (b) corresponding surface-based ERT estimate. The red rectangle denotes the location of a borehole, along which the true electrical conductivity is known. The brown vertical line segments in (a) indicate the locations of the 47 electrodes used.

least-squares optimization of an objective function consisting of the sum of the data misfit and a model regularization term (e.g., DeGroot-Hedlin and Constable, 1990; Menke, 1989)

$$\Psi(\boldsymbol{\sigma}) = \left\| \mathbf{W}_r \left[\mathbf{r}^{\text{obs}} - f(\boldsymbol{\sigma}) \right] \right\|_2^2 + \alpha \left\| \mathbf{W}_\Sigma \boldsymbol{\sigma} \right\|_2^2, \tag{5.1}$$

where $\mathbf{W}_r$ is a data weighting matrix related to the observational errors and their correlations and $\mathbf{W}_\Sigma$ is the model regularization operator, which is typically the discrete first-order derivative. During the inversion process, the regularization parameter α is maximized under the constraint that the data are fitted to a predefined error level (Constable et al., 1987). Non-linear inverse problems are solved iteratively based on successive linearization around the model obtained in the previous iteration.

The model resolution matrix $\mathbf{R}$ relates the unknown true coarse model parameters to the estimated parameters

$$\boldsymbol{\sigma}^{\text{est}} = \mathbf{R}\boldsymbol{\sigma}^{\text{true}}. \tag{5.2}$$

It can be approximated based on the last iteration p Newman and Alumbaugh (2000)

$$\mathbf{R} = \left(\mathbf{J}_p^T \mathbf{W}_r^T \mathbf{W}_r \mathbf{J}_p + \alpha_p \mathbf{W}_\Sigma^T \mathbf{W}_\Sigma \right)^{-1} \mathbf{J}_p^T \mathbf{W}_r^T \mathbf{W}_r \mathbf{J}_p, \tag{5.3}$$

where the sensitivity matrix $\mathbf{J}_p$ is obtained by linearizing $f()$ around the model obtained in the pth biteration. The model regularization results in an artificially high spatial correlation of the inverted values, which is characterized by the posterior covariance

$$\mathbf{C}_\Sigma^{\text{est}} = \left(\mathbf{J}_p^T \mathbf{W}_r^T \mathbf{W}_r \mathbf{J}_p + \alpha_p \mathbf{W}_\Sigma^T \mathbf{W}_\Sigma \right)^{-1}. \tag{5.4}$$

Note that this covariance matrix contains the information about the correlation among parameters due to data errors and the averaging created by the regularization operator.

5.2.3 Area-to-point kriging and stochastic simulation

The following subsections provide a summary of the general formulation of area-to-point kriging in the discrete case (section 2.3.1) as well as its extension for conditional estimation (2.3.2) and for conditional simulations (2.3.3). Area-to-point kriging estimation and simulation are described in detail in Kyriakidis (2004) and Kyriakidis and Yoo (2005).

5.2.3.1 Area-to-point kriging

Area-to-point kriging is a particular case of the co-kriging estimation of a primary variable **z** with a secondary variable **Z** that can be related to **z**by a linear operator **G**

$$\mathbf{Z} = \mathbf{G}\mathbf{z}. \tag{5.5}$$

In this paper, the **G**-transform of a variable denotes the result of applying the operator **G** to the variable in question. Because **z** is a Gaussian variable with a known covariance function, its covariance matrix, denoted as $\mathbf{C}_z$, is known. Thus, the cross-covariance with respect to **Z** can be computed as

$$\mathbf{C}_{zZ} = \mathbf{C}_z\mathbf{G}^T, \tag{5.6}$$

and the covariance matrix of **Z** is

$$\mathbf{C}_Z = \mathbf{G}\mathbf{C}_z\mathbf{G}^T. \tag{5.7}$$

The estimation of **z** by area-to-point kriging, denoted as $\hat{\mathbf{z}}$, is given by the linear combination of **Z**weighted by the coefficients Λ

$$\hat{\mathbf{z}} = \Lambda\mathbf{Z}. \tag{5.8}$$

In analogy to traditional kriging, the weights Λ are obtained by solving a linear system of equations constructed from the covariance matrices and $\mathbf{C}_{zZ}$

$$\mathbf{C}_Z \Lambda = \mathbf{C}_{zZ}. \tag{5.9}$$

5.2.3.2 Area-to-point kriging with point conditioning

In the presence of known data $\mathbf{z}^{\text{hd}}$, commonly referred to as hard data, the kriging estimation

$$\hat{\mathbf{z}} = \Lambda_z \mathbf{z}^{\text{hd}} + \Lambda_Z \mathbf{Z}, \tag{5.10}$$

and the kriging system in equation (9) becomes

$$\begin{bmatrix} \mathbf{C}_Z & \mathbf{C}_{Z,z^{\text{hd}}} \\ \mathbf{C}_{z^{\text{hd}},Z} & \mathbf{C}_{z^{\text{hd}}} \end{bmatrix} \begin{bmatrix} \Lambda_Z \\ \Lambda_z \end{bmatrix} = \begin{bmatrix} C_{\hat{z},Z} \\ C_{\hat{z},\, z^{\text{hd}}} \end{bmatrix}. \tag{5.11}$$

5.2.3.3 Area-to-point conditional simulation

Gaussian error simulation is used to generate conditional realizations. For completeness, the procedure is briefly recalled below for a single realization. See Kyriakidis and Yoo (2005) for more detail.

1. Construct an unconditional realization $\mathbf{z}^{\text{sim}}$ with the Fast-Fourier Transform Moving Average (FFT-MA) method Le Ravalec-Dupin et al. (2000)

$$\mathbf{z}^{\text{sim}} = \text{FFTMA}(0, \mathbf{C}_z) \tag{5.12}$$

2. Compute the corresponding G-transform $\mathbf{Z}^{\text{sim}}$

$$\mathbf{Z}^{\text{sim}} = \mathbf{G}\mathbf{z}^{\text{sim}} \tag{5.13}$$

3. Compute the co-kriging estimate based on $\mathbf{Z}^{\text{sim}}$ and the values of $\mathbf{z}^{\text{sim}}$at the locations of the hard data $\mathbf{u}^{\text{hd}}$

$$\hat{\mathbf{z}}^{\text{sim}} = \Lambda_z \mathbf{z}^{\text{sim}}(\mathbf{u}^{\text{hd}}) + \Lambda_Z \mathbf{Z}^{\text{sim}} \tag{5.14}$$

4. Compute the co-kriging predictor of the known data $\mathbf{Z}^{\text{est}}$ and hard data $\mathbf{z}^{\text{hd}}$

$$\hat{\mathbf{z}} = \Lambda_z \mathbf{z}^{\text{hd}} + \Lambda_Z \mathrm{Z}^{\text{est}} \tag{5.15}$$

5. Finally, compute the conditional realization

$$\mathbf{z}_{\text{c}}^{\text{sim}} = \hat{\mathbf{z}} + \left[\mathbf{z}^{\text{sim}} - \hat{\mathbf{z}}^{\text{sim}}\right] \tag{5.16}$$

5.2.4 Area-to-point simulation based on ERT tomogram

In the following, the information gathered by the ERT inversion (section 2.2) is included in area-to-point kriging (section 2.3). First, the electrical conductivity at the fine scale $\boldsymbol{\sigma}$ is transformed into the standard normally distributed variable $\mathbf{z}$

$$\mathbf{z} = \frac{\log(\boldsymbol{\sigma}) - \boldsymbol{\mu}_\sigma}{s_\sigma}, \tag{5.17}$$

where $\boldsymbol{\mu}_\sigma$ is a vector with the same size as $\mathbf{z}$ containing the mean electrical conductivity and s_σ is the standard deviation of the fine-scale electrical conductivity. Similarly, the variable $\mathbf{Z}$ is defined as

$$\mathbf{Z} = \frac{\boldsymbol{\sigma} - \boldsymbol{\mu}_\sigma}{s_\sigma}. \tag{5.18}$$

Because $\boldsymbol{\mu}_\sigma$ and s_σ are the fine-scale mean and standard deviation, $\mathbf{Z}$ has a mean of zero but not a standard deviation of one. Note that the fine-scale variable $\boldsymbol{\sigma}$ describing electrical conductivity has a log-normal distribution, while the larger-scale variable $\mathbf{Z}$ describing the logarithm of electrical conductivity has a normal distribution.

Then, the operator **G** is found to be equivalent to **RU**, where **U** is the linear operator that performs upscaling from the resolution of $\boldsymbol{\sigma}$ to the model parameter cell size of $\boldsymbol{\sigma}$ (**U** is constructed as a linear average, where $\mathbf{U}(i,j)$ is the weight of $\log(\boldsymbol{\sigma}(j))$ in defining the average $\boldsymbol{\sigma}(i)$, such that the sum of each row is equal to one). This is shown by combining equations (2), (17), and (18)

$$\mathbf{Z}^{\text{est}} = \frac{\boldsymbol{\sigma}^{\text{est}} - \boldsymbol{\mu}_\sigma}{s_\sigma} = \frac{\mathbf{R}\boldsymbol{\sigma}^{\text{true}} - \boldsymbol{\mu}_\sigma}{s_\sigma} = \frac{\mathbf{RU}\log(\boldsymbol{\sigma}^{\text{true}}) - \boldsymbol{\mu}_\sigma}{s_\sigma}. \tag{5.19}$$

Since **RU**is an operator with all sums of rows equal to one, applying it to a vector of equal values does not affect this vector, such that $\boldsymbol{\mu}_\sigma = \mathbf{RU}\boldsymbol{\mu}_\sigma$, which added to equation (19) gives

$$\mathbf{Z}^{\text{est}} = \frac{\mathbf{RU}\log(\boldsymbol{\sigma}^{\text{true}}) - \mathbf{RU}\boldsymbol{\mu}_\sigma}{s_\sigma} = \mathbf{RU}\frac{\log(\boldsymbol{\sigma}^{\text{true}}) - \boldsymbol{\mu}_\sigma}{s_\sigma} = \mathbf{RUz}^{\text{true}}. \tag{5.20}$$

Finally, the covariance matrix of **Z**needs to be updated to account for the effects of the data errors and the regularization used in the inversion process. Indeed, the model regularization matrix and data errors add correlation and uncertainty to $\boldsymbol{\sigma}^{\text{est}}$. These effects are accounted for by the covariance $\mathbf{C}_\Sigma^{\text{est}}$ of equation (4), and once normalized, it can be added to the covariance of the spatial structure $\mathbf{C}_Z$

$$\mathbf{C}_Z^{\text{est}} = \mathbf{C}_Z + \left(\frac{1}{s_\sigma}\right)^2 \mathbf{C}_\Sigma^{\text{est}} \tag{5.21}$$

The resulting covariance matrix combines the spatial information and the effect of smoothing and data errors on the tomogram.

5.3 Results

In the following, the methodology outlined above is tested for a synthetic case study of a heterogeneous alluvial aquifer, where the field parameters are inspired by those recently reported by Pirot et al. (2017). Figure 5.2 provides an overview of methodology, which consists of four parts: (1) generate the synthetic data (yellow background), (2) perform the inversion (blue background), (3) calculate the kriging estimation (green background), and (4) produce

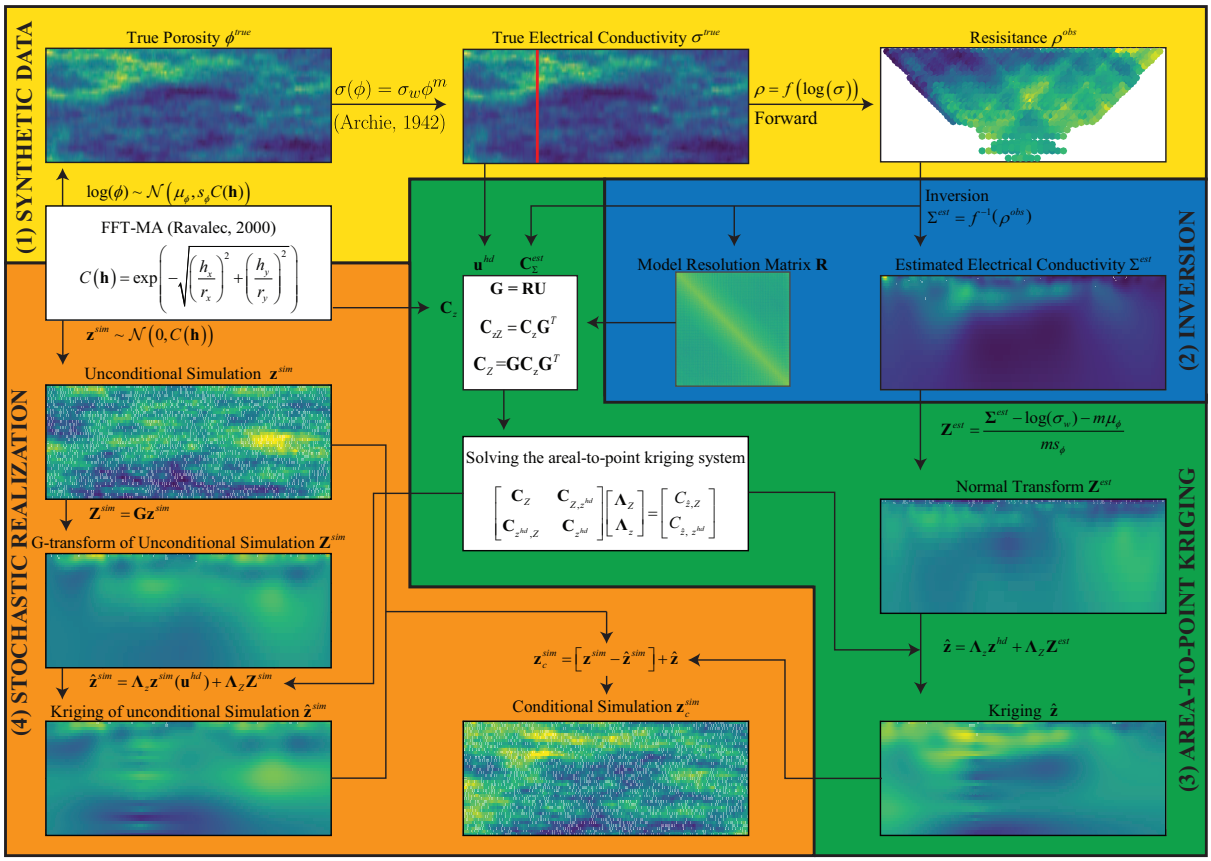


Figure 5.2 – Schematic overview of the key elements of the proposed methodology involving (1) generation of synthetic data (yellow), (2) geophysical inversion (blue), (3) area-to-point kriging (green), and (4) conditional stochastic simulation (orange).

stochastic realizations (orange background). These parts are described in the following subsections.

5.3.1 Synthetic data

The fine-scale synthetic electrical conductivity structure (Figure 5.1a) is generated using the FFT-MA method (Le Ravalec-Dupin et al., 2000) from a log-normally distributed porosity field ϕ^{true} (Pirot et al., 2017). Using Archie's law (Archie, 1942),

$$\sigma(\phi) = \sigma_w \phi^m, \tag{5.22}$$

this porosity field is transformed into an electrical conductivity field, which thereafter is considered as the true electrical conductivity model 40×4. The cementation exponent m

and the electrical conductivity of pore water σ_w are assumed to be known and constant throughout the study area, which is a common approximation for a given hydrogeological unit (e.g., Ruggeri et al., 2014). The corresponding simulated resistances that would have been measured in the course of a noise-free surface-based ERT-type geoelectric survey are evaluated using the forward modeling part of the software R2 Binley and Kemna (2005), which corresponds to a 2.5D solver of the Poisson equation. The boundary conditions along all four boundaries are of the Neumann type (zero flux). A surface-based dipole-dipole survey is used with 47 electrodes spaced every 2 m, from $x = 4m$ to $x = 96m$. To avoid artifacts associated with boundary conditions, a large buffer zone of 10 log-spaced nodes (700 m) is added to the left and right model edges as well as to the bottom of the domain. For the forward computation, the mean value of the true field is used in this buffer zone. To mimic realistic measurement errors, heteroskedastic normally distributed noise with a standard deviation s_ε is added to the simulated resistances to create the observed data $\mathbf{r}^{\text{obs}}$. Table 1 provides a summary of the parameters used to generate the synthetic electrical conductivity structure and the corresponding ERT-type geolectric measurements.

5.3.2 Geophysical inversion

The inversion of the observed resistances is performed with the inversion subroutine of R2 Binley and Kemna (2005) using the mean value of the apparent electrical conductivity as the initial model. To be consistent with the considered ratio of covariance ranges, the model regularization operator penalizes horizontal variations ten times more strongly than vertical variations. Figure 5.1b represents the estimated electrical conductivity field $\boldsymbol{\sigma}^{\text{est}}$ resulting from the inversion of the observed resistance data. $\boldsymbol{\sigma}^{\text{est}}$ is characterized by a significant loss in resolution with regard to $\boldsymbol{\sigma}^{\text{true}}$ (Figure 5.1a). Moreover, the resolution of $\boldsymbol{\sigma}^{\text{est}}$ is spatially variable and decreases significantly with depth.

The resolution matrix $\mathbf{R}$ is computed according to equation (3). Figure 5.3a displays the diagonal of $\mathbf{R}$ reshaped into a 2D grid, thus indicating to which degree each cell of the domain is resolved. The pixels in the first top row are resolved between 60-70%. This value decreases with depth reaching near-zero values at the bottom and of the lateral edges of

the domain. Each row of the resolution matrix, also known as the point-spread function, corresponds to the relative spreading of the information associated with a given node over the other nodes, so that $\boldsymbol{\sigma}^{\text{est}}(i) = \mathbf{R}(i,:)\boldsymbol{\sigma}^{\text{true}}$. Figures 5.3b through 5.3d display the point-spread-functions corresponding to 3 different locations marked with a red dot in Figure 5.3a. Figures5.3b though 5.3d only show sub-domains of Figure 5.3a, as denoted by red rectangles, because the remaining domain has a relatively limited influence and thus low values. These three examples demonstrate the ability of the resolution matrix to provide information for each individual cell of the grid with regard to the specific averaging pattern imposed by the inversion on each cell. For instance, a small and well-defined radius of influence for cells located near the surface (Figure 5.3b), a larger and more diffuse contribution for deeper cells with oscillating values of weights from positive to negative (Figure 5.3c), and an asymmetric shape of influence for cells located near the lateral edges of the domain (Figure 5.3d).

Similar to the forward modeling, a buffer zone is used during the inversion to limit the effects of the boundary of the domain. Consequently, the rows and columns of the resolution matrix also include nodes corresponding to this buffer zone. Since the zone of interest to simulate does not include this buffer zone, the resolution matrix is split in two parts, $\mathbf{R}_{\text{in}}$ for the columns of $\mathbf{R}$ pertaining to model cells in the interior and $\mathbf{R}_{\text{out}}$ for the columns of R pertaining to model cells in the buffer zone. However, the true field $\boldsymbol{\sigma}^{\text{true}}$related to the zone to be simulated does not include the buffer zone, whereas the known estimated tomogram $\boldsymbol{\sigma}^{\text{est}}$does. Thus, the contribution of the buffer zone $\mathbf{R}_{\text{out}}(i,:)$ to $\boldsymbol{\sigma}^{\text{est}}(i)$is subtracted by the estimated tomogram in the buffer zone $\boldsymbol{\sigma}^{\text{est}}_{\text{out}}$ in order to remove its influence on the estimation process, such that

$$\boldsymbol{\sigma}^{\text{est}}(i) - \mathbf{R}_{\text{out}}(i,:)\,\boldsymbol{\sigma}^{\text{est}}_{\text{out}} = \mathbf{R}_{\text{in}}(i,:)\,\boldsymbol{\sigma}^{\text{true}}. \tag{5.23}$$

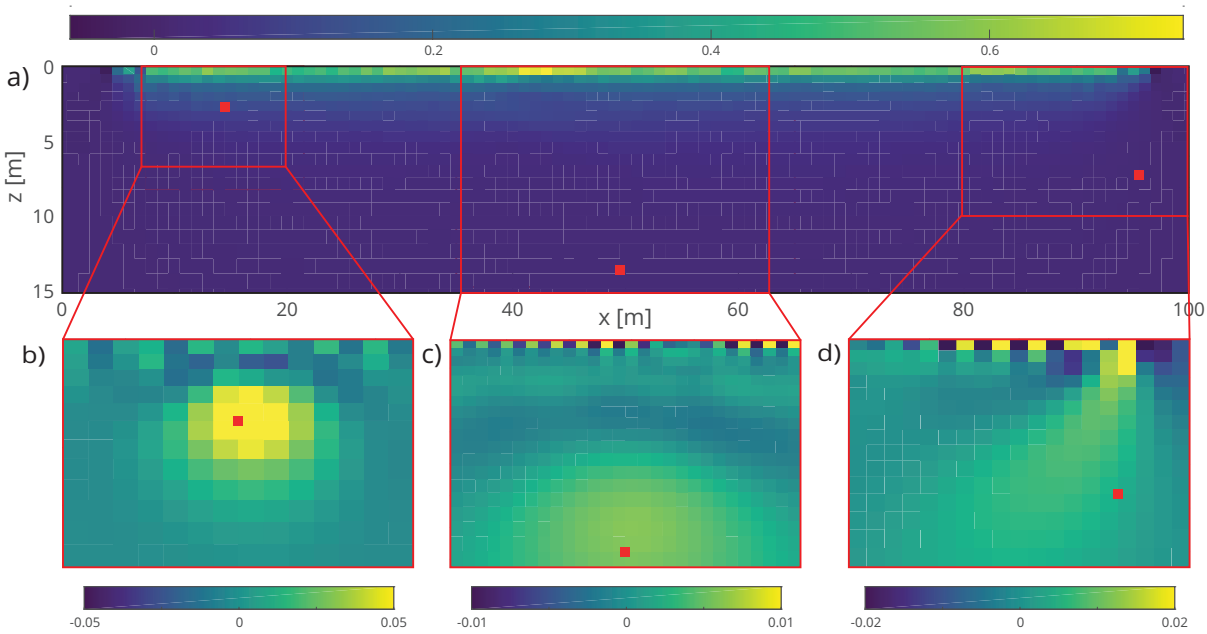


Figure 5.3 – a) Diagonal of the resolution matrix reshaped into a 2D grid. b)-d) Rows of the resolution matrix (point-spread functions) corresponding to the locations denoted by the red dots in the zoomed sub-figures. Note that different color scales are used in each panel in order to highlight the characteristic features.

5.3.3 Area-to-point kriging

Figure 5.4a presents the normalized secondary variable $\mathbf{Z}^{\text{est}}$ computed from $\boldsymbol{\sigma}^{\text{est}}$ using equation (18). Here, the mean μ_σ and standard deviation s_σ of the electrical conductivity is related to those of the porosity using Archie's law

$$\mu_\sigma = \log(\sigma_w) + m\mu_\phi \tag{5.24}$$

and

$$s_\sigma = m s_\phi \tag{5.25}$$

The upscaling operator **U**is constructed with a gridded linear 2D interpolation, which maps the fine-scale to the large-scale. Figure 5.4b shows the result of the **G**-transform $\mathbf{G} = \mathbf{RU}$ of the normalized true electrical conductivity $\mathbf{z}^{\text{true}}$. The strong resemblance between $\mathbf{Z}^{\text{est}}$ (Figure 5.4a) and $\mathbf{G}\mathbf{z}^{\text{true}}$ (Figure 5.4b) as well as the comparatively small residuals (Figure 5.4c) demonstrate

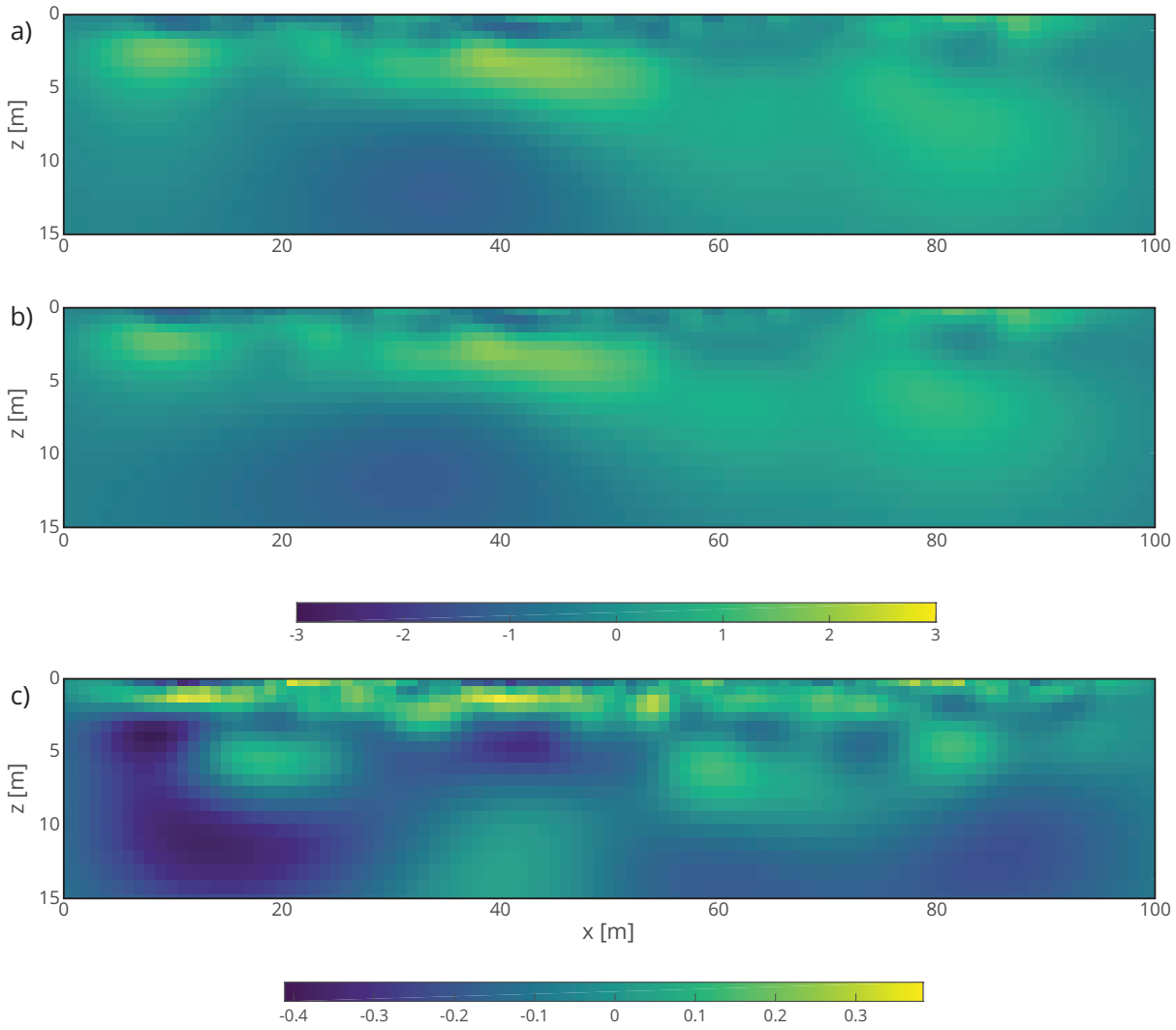


Figure 5.4 – a) $\mathbf{Z}^{\text{est}}$ resulting from the tomographic inversion, b) G-transform of the true field $\mathbf{Gz}^{\text{true}} = \mathbf{RUz}^{\text{true}}$, and c) difference between the two.

the ability of the resolution matrix to correctly mimic the smoothing effects inherent to the inversion.

The diagonal of the covariance matrices $\mathbf{C}_Z$ and $\mathbf{C}_\Sigma^{\text{est}}$ of equation (21) are displayed in Figure 5.5. The difference in magnitude points to the relatively small importance of the error estimates of the tomogram compared to the variability of the geostatistical model. Note that values above one are possible because the top rows of $\mathbf{Z}$ have smaller grid sizes than $\mathbf{z}$, thus increasing the variance.

The area-to-point kriging system is constructed using equation (11). The hard data are assumed to come from a single borehole located at a lateral distance of 33 m (Figure 5.1a),

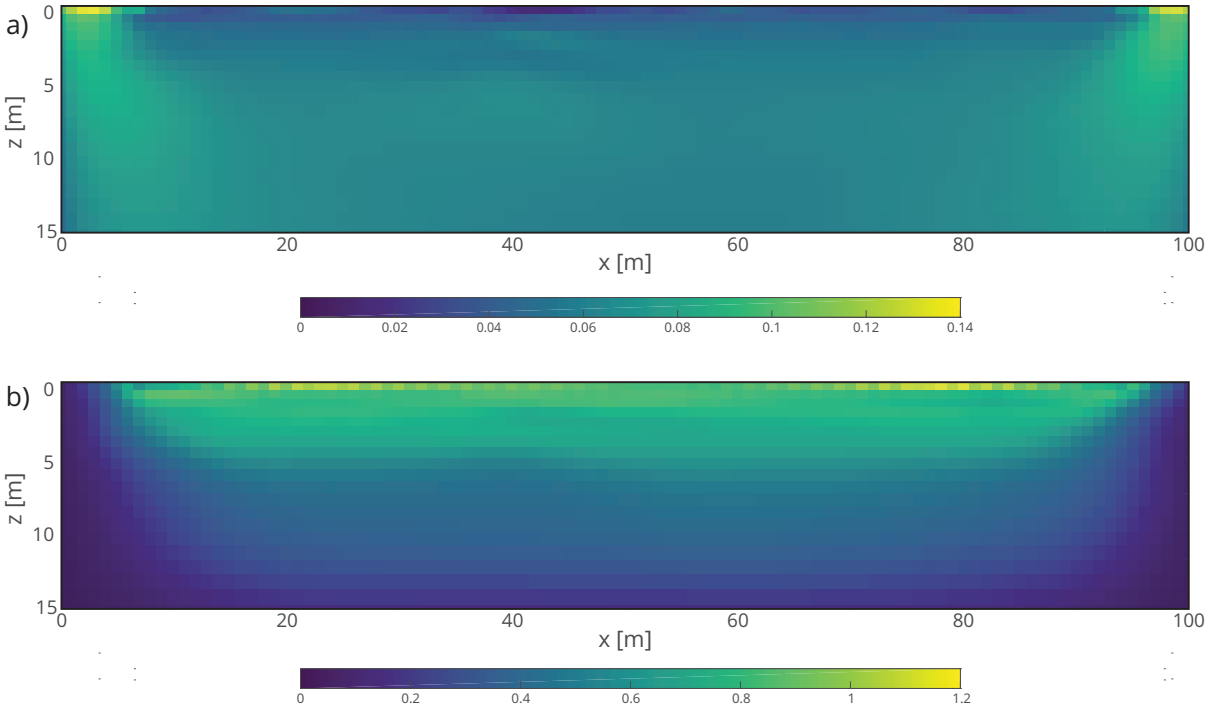


Figure 5.5 – Diagonals of the covariance matrices (a) resulting from the inversion $\left(\frac{1}{s_\sigma}\right)^2 \mathbf{C}_\Sigma^{\text{est}}$ and (b) of the geostatistical model $\mathbf{C}_Z$.

along which error-free measurements of the electrical conductivity at the resolution of the true field are available. Figure 5.6 shows the resulting co-kriging estimation $\hat{\mathbf{z}}$ based on $\mathbf{Z}^{\text{est}}$ and the conditioning data $\mathbf{z}^{\text{hd}}$.

5.3.4 Stochastic simulation

Based on the workflow outlined in section 2.3.3, 500 posterior realizations are generated to assess if the variability of fine-scale conductivity realizations are consistent with the tomogram and the underlying geostatistical model. Figure 5.7 compares the normalized true field $\mathbf{z}^{\text{true}}$ to a randomly selected conditional realization $\mathbf{z}_c^{\text{sim}}$, the mean of the 500 realizations $\overline{\mathbf{z}_c^{\text{sim}}}$, and the corresponding standard deviation std$\left(\mathbf{z}_c^{\text{sim}}\right)$. Visual appraisal of the considered example realization (Figure 5.7b) indicates that the stochastic simulations correctly reproduce the texture of $\mathbf{z}^{\text{true}}$, while the mean field $\overline{\mathbf{z}_c^{\text{sim}}}$ (Figure 5.7c) shows that the realizations capture the larger-scale features, with increasing details towards the surface and the borehole location, along which the hard data have been sampled. $\overline{\mathbf{z}_c^{\text{sim}}}$ is very similar to Figure 5.6a, which is

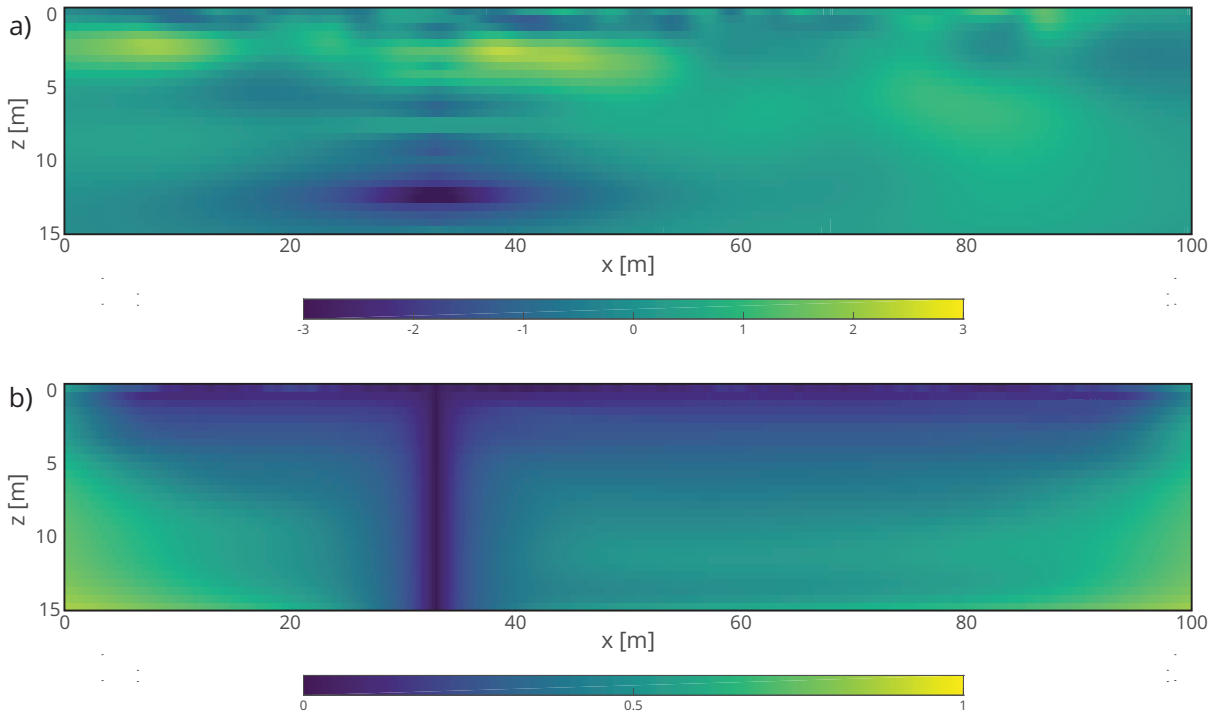


Figure 5.6 – a) Co-kriging estimation based on the tomographic image and the hard data $\hat{\mathbf{z}} = \Lambda_z \mathbf{z}^{hd} + \Lambda_Z \mathbf{Z}$ and b) corresponding co-kriging error variance.

expected as the mean of realizations tends towards the kriging estimation. The standard deviation field (Figure 5.7d) shows that the realizations are less variable near the surface and near the borehole.

Figure 5.8 displays an analysis of the 500 realizations using semi-variograms and probability density to enable a comparison between the true field and the theoretical model. These results confirm that, with some ergodic fluctuations, the spatial statistics are reasonably well reproduced (Figure 5.8a through b). Empirical semi-variograms of the realizations show an overall good reproduction of the spatial structure of the theoretical model. However, as the spatial structure of the realizations is also influenced by the spatial information contained in the tomogram, and ultimately by the true field, the empirical variograms of the realizations appear to differ from the theoretical model in a similar manner as those of the true field.

Figure 5.9 compares the **G**-transform of the 500 conditional realizations (Figures 5.8b though5.8d) to the inverted field $\mathbf{Z}^{\text{est}}$ (Figure 5.9a). Both the single upscaled realization (Figure 5.9b) and the mean of the 500 realizations (Figure 5.9c) present a good match to the inverted field

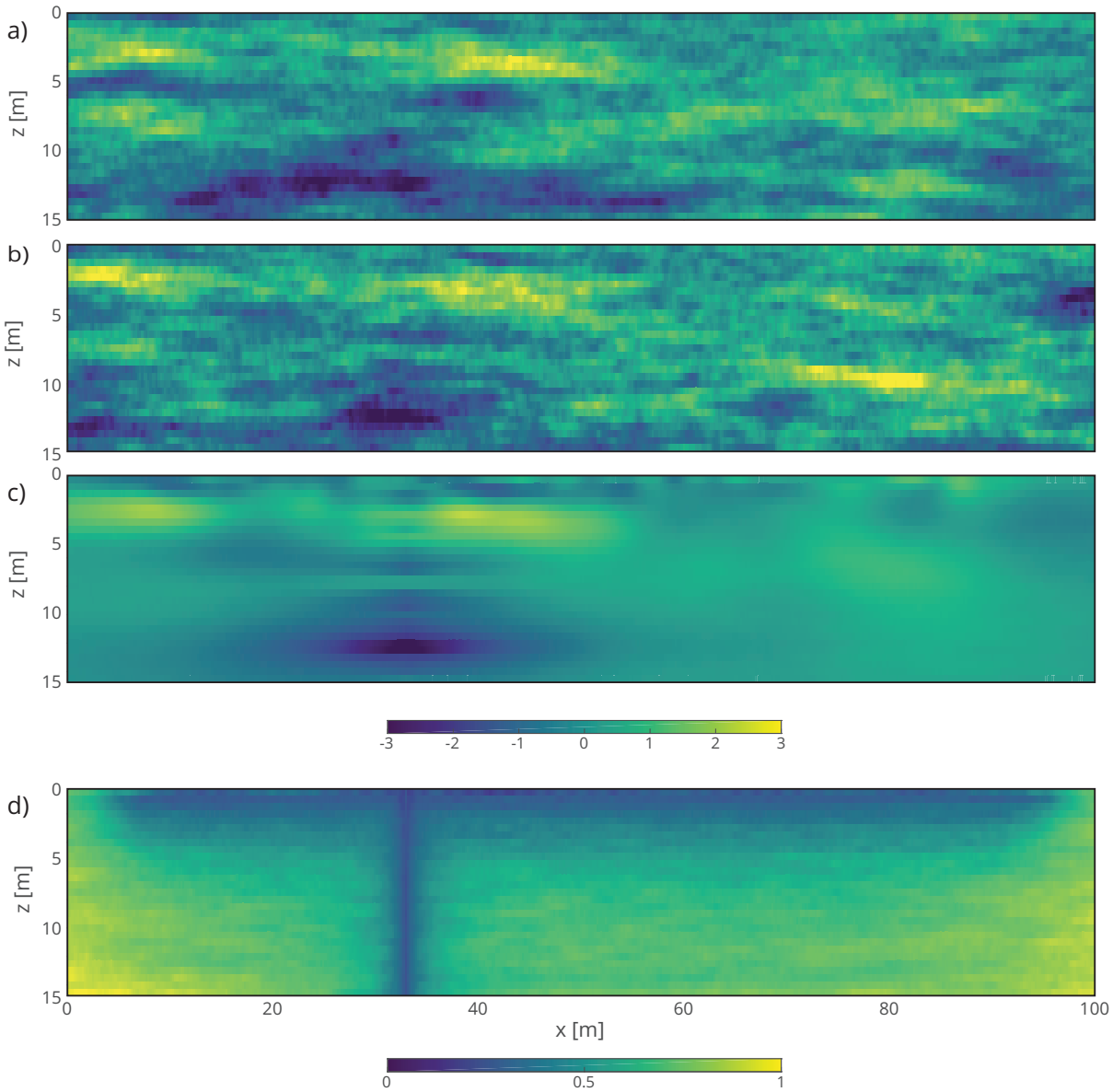


Figure 5.7 – Results of stochastic simulations. a) True initial field $\mathbf{z}^{\text{true}}$, b) example of a realization $\mathbf{z}_c^{\text{sim}}$, c) mean of 500 realizations $\overline{\mathbf{z}_c^{\text{sim}}}$, and d) standard deviation of the same 500 realizations $\text{std}\left(\mathbf{z}_c^{\text{sim}}\right)$.

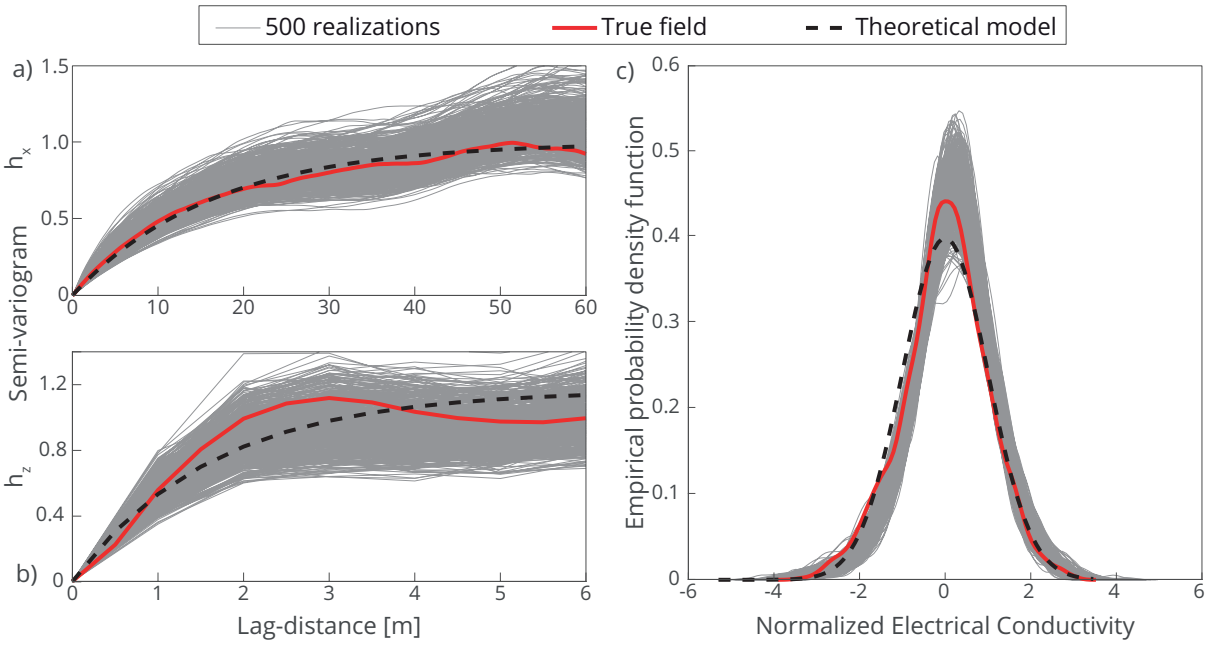


Figure 5.8 – Horizontal and b) vertical semi-variograms, and c) probability density function based on 500 realizations (grey lines) with the values for true field (red line) and the theoretical model (dashed black line) superimposed.

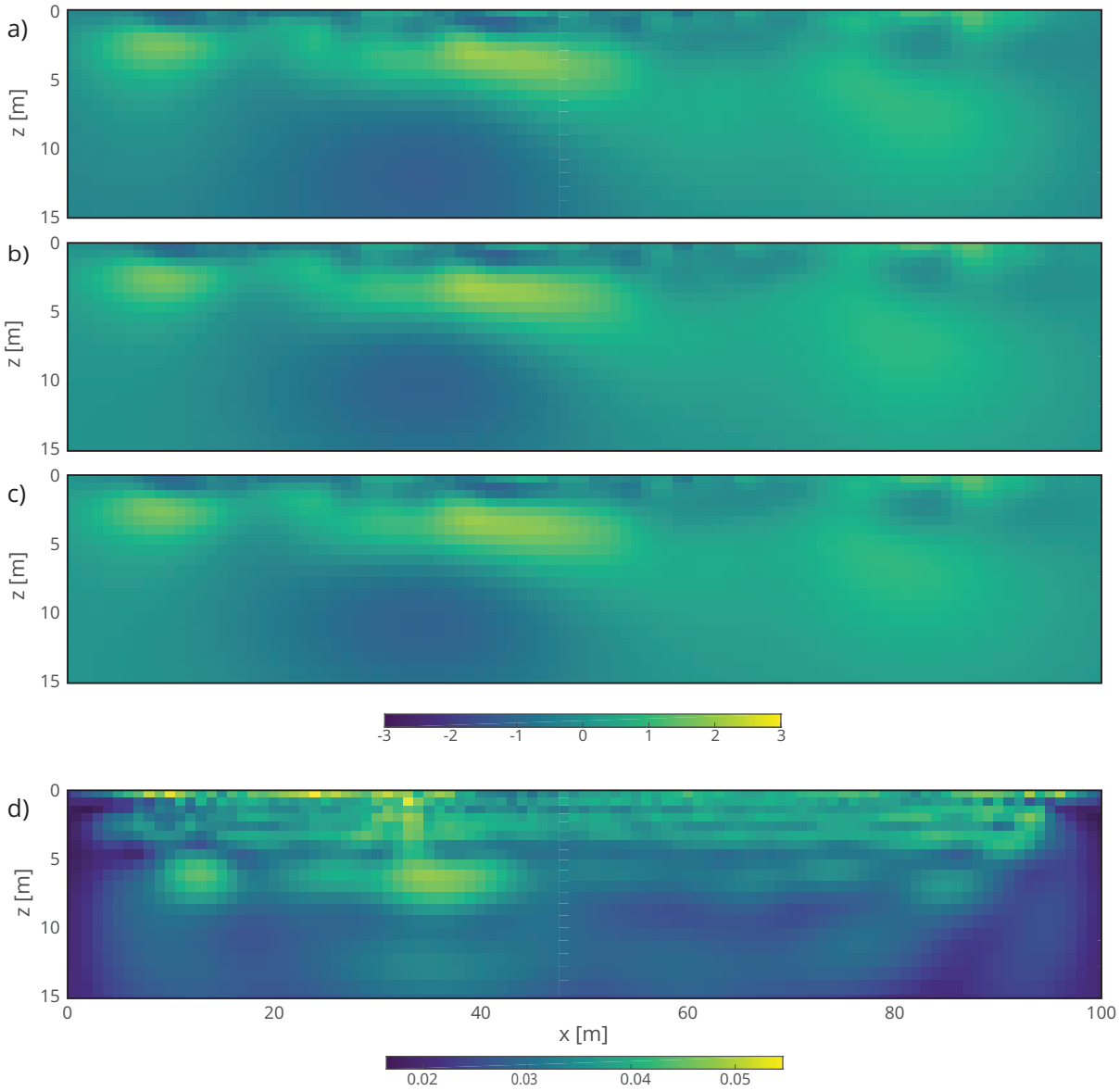


Figure 5.9 – Results of the stochastic simulations. a) Inverted field $\mathbf{Z}^{\text{est}}$, b) upscaled version or G-transform of a single example of realizations $\mathbf{RUz}_{\text{c}}^{\text{sim}}$, c) mean of the G-transforms of 500 realizations $\overline{\mathbf{RUz}_{\text{c}}^{\text{sim}}}$, and d) standard deviation of the G-transforms of 500 realizations $\text{std}\left(\mathbf{RUz}_{\text{c}}^{\text{sim}}\right)$.

(Figure 5.9a). Although, a detailed analysis of the standard deviation of the **G**-transform in Figure 5.9d highlights some variability close to surface.

5.3.5 Corroboration with apparent resistivity

When applying a two-step approach to a non-linear problem, there is no guarantee that the resulting geostatistical realizations are fully compatible with the observed geophysical data (e.g. Bosch, 2004). One way to assess the information loss is to simulate the forward response from the various realizations and then to compare them with the observed data (Linde et al., 2015a). Here, we calculate the forward response of our 500 geostatistical realizations. The pseudo-sections of the observed apparent resistivity (Figure 5.10a) is overall well reproduced by the average simulated apparent resistivity of all 500 model realizations (Figure 5.10b). The normalized error $\left(\overline{f\left(\boldsymbol{\sigma}_{\mathrm{c}}^{\mathrm{sim}}\right)}-\mathbf{r}^{\mathrm{obs}}\right)/\mathbf{r}^{\mathrm{obs}}$ (Figure 5.10c) indicates some discrepancies, particularly near the surface. This is expected as 2% noise was added in creating the observed data. In Figure 5.11, it is observed that the forward response of the true field is not perfectly aligned with the 1:1 curve (corresponding to the observed apparent resistivities), while the variability in the apparent resistivity responses of the 500 realizations is much larger. This data misfit can be quantified with the weighted root-mean square error (WRMSE)

$$\mathrm{WRMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{f\left(\boldsymbol{\sigma}_{\mathrm{c}}^{\mathrm{sim}}\right)_i-\mathbf{r}_i^{\mathrm{obs}}}{s_\varepsilon \mathbf{r}_i^{\mathrm{obs}}}\right)^2}. \tag{5.26}$$

The origin of this information loss can be traced back to the difference between the inverted field $\mathbf{Z}^{\mathrm{est}}$ and the RU transform of the true field $\mathbf{RUz}^{\mathrm{true}}$ (Figure 5.4). Indeed, while the forward response of $\mathbf{Z}^{\mathrm{est}}$ produce a WRME equal to one, the one of $\mathbf{RUz}^{\mathrm{true}}$is equal to 2.5. The first reason is that the forward response is affected by the upscaling. Indeed, the forward response of the upscaled true field $\mathbf{Uz}^{\mathrm{true}}$has a WRMSE of 1.75; without upscaling, the forward response of the true field $\mathbf{z}^{\mathrm{true}}$ leads to a WRMSE of 0.98. The second reason is that the linearization around the final iteration used to compute the Jacobian matrix leads to an approximate resolution matrix $\mathbf{R}$. To investigate this effect, the Jacobian and posterior covariance matrix were computed directly based on the true model at the fine scale, such that the linearization is computed on the true fine-scale model. Here, the resulting WRMSE based on $\mathbf{R}^{\mathrm{true}}\mathbf{z}^{\mathrm{true}}$ is 1.53. By combining the error of these two contributions $\sqrt{1.75^2+1.53^2}=2.32$ leads to an error close to the error of 2.5 found in the model realizations.

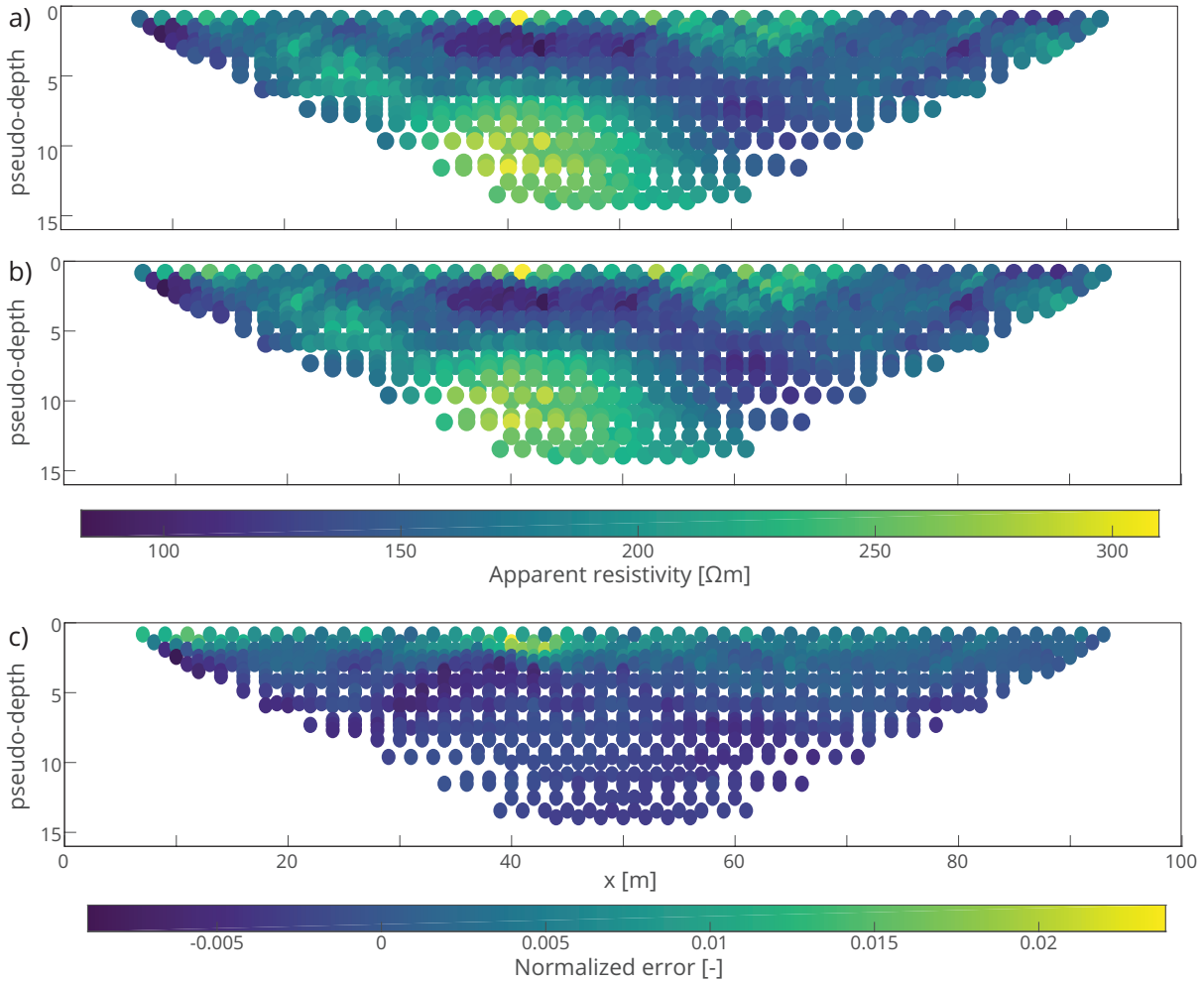


Figure 5.10 – Comparison of the pseudo-sections of the apparent resistivity. a) Observed apparent resistivity and b) the mean forward response of 500 stochastic realizations, and c) the normalized relative error of the realizations $\left(\overline{f\left(\boldsymbol{\sigma}_{\mathrm{c}}^{\mathrm{sim}}\right)}-\mathbf{r}^{\mathrm{obs}}\right)/\mathbf{r}^{\mathrm{obs}}$.

Figure 5.11 – Comparison of the simulated and observed apparent resistivities for the 500 realizations.

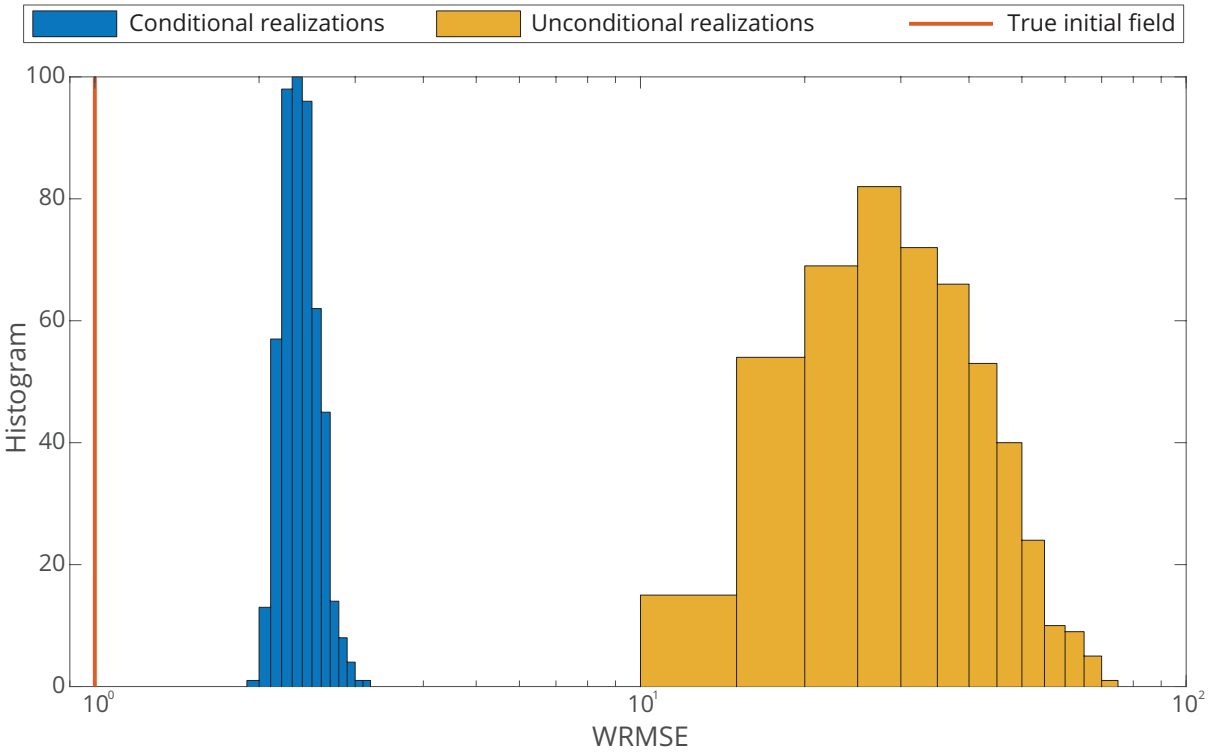


Figure 5.12 – Histogram of the WRMSE misfit of 500 realizations.

5.4 Conclusions

We have presented a two-step approach for producing fine-scale stochastic realizations of a rock physical property based on corresponding low-resolution tomographic images resulting from smoothness-constrained deterministic inversions, borehole data, and an underlying geostatistical model. The proposed technique relies on area-to-point kriging, which requires a linear support function between the targeted fine-scale structure and the coarse resolution of the tomogram. This is achieved through the model resolution and posterior covariance matrices computed for the tomographic image at the last iteration of a linearized inversion. The method was tested and verified with regard to a synthetic ERT-type case study for a strongly heterogeneous subsurface model. This choice was motivated by the widespread use of geoelectric measurements, the importance of the electrical conductivity as a rock physical property, and the generally low and spatially variable resolution of the corresponding tomographic images. It is, however, important to note that this method could be applied to smoothness-constrained inversions of virtually any other type of geophysical

measurements and/or with any other type of model regularization, such as, for example, damping or curvature smoothing, as long as the corresponding resolution and posterior covariance matrices can be calculated. One important advantage of the method presented is the ability to provide uncertainty ranges and posterior realizations at a low computational cost. For comparison, with the forward model taking the order of several seconds to run, a Markov chain Monte Carlo (MCMC) chain would require several weeks of calculations to provide such information without any underlying simplifications, while this took a matter of hours for the method presented. Preliminary tests using MCMC (not shown) indicated that we could not reach convergence of the MCMC chain after 100'000 iterations. Our results demonstrate that the proposed methodology is able to generate realizations of the underlying fine-scale electrical conductivity structure that faithfully reproduce the tomogram through the support function, while also being constrained to the considered geostatistical model.

Chapter 6

Conclusion

As stated in the introduction, forecasting hydrogeological systems is a pre-requisite to informed decisions on groundwater management. This is performed through groundwater modelling, which strongly relies on the adequate characterization of the corresponding aquifers in terms of the spatial distributions of key parameters, such as hydraulic conductivity. A relatively recent advance in the progress of estimating these parameters consists in using geophysical methods to constrain them.

The general objective of this thesis was therefore to develop computationally efficient methodologies to stochastically simulate hydrological parameters constrained by geophysical data. To this end, the first two chapters sought to optimize the computational efficiency and accuracy of Sequential Gaussian Simulation (SGS). In Chapter 4, SGS was adapted into Bayesian Sequential Simulation (BSS) in order to simulate stochastic hydraulic conductivity fields integrating the information of electrical resistivity tomography with log linear pooling. Finally, in Chapter 5, we developed a methodology to stochastically downscale the electrical resistivity tomogram while accounting explicitly for the inherent smoothing of the tomogram. The main contributions of each chapter are summarized below.

6.1 Simulation path in Sequential Gaussian Simulation (SGS)

6.1.1 Which path to choose in SGS?

The objective of this work is (1) to explain the origin of biases in SGS and their link with the simulation path, (2) to present a method to quantitatively evaluate a simulation path, and (3) to assess the most common paths (row-by-row, spiral, multi-grid, mid-point, random or quasi-random). Results and recommendations can be summarized as follows:

- Biases in SGS are due to using a limited neighbourhood. Therefore, increasing the neighbourhood size is the most efficient way to reduce these biases.
- The corresponding evaluation results in the classification of simulation paths into clustering and declustering paths, i.e., simulating consecutively nearby nodes or maximizing the distance between consecutively simulated nodes.
- While clustering paths correctly reproduce covariance associated with short lag-distance, the larger lag-distances are poorly reproduced. Conversely, declustering paths maintain a more diversified neighbourhood and minimize the re-use of nodes associated with large biases.
- Finally, we recommend using a multi-grid path, or, alternatively, a random path in case a simple implementation is required.

6.1.2 Accelerating SGS with a constant path

In this study, we analysed the use of a constant simulation path in SGS when multiple realizations are generated by (1) presenting a mathematical framework to explain the reason for using a randomized path, (2) assessing the algorithmic implications together with numerical considerations, and (3) comparing the resulting computational gain and biases when using a constant path. Results and recommendations can be summarized as follows:

- The use of a constant simulation path results in additional numerical considerations, which are especially pertinent for large grids and/or large neighbourhoods.
- Parallelization is also possible with a constant path.
- For each additional realization, the computational gain of using a constant path is equal to the time spent to compute the kriging weights on the first realization. This can correspond to an estimated 60 to 99% reduction in computation time depending primarily on the grid size, neighbourhood size, and/or search method used.
- The minor biases incurred by using a constant path are easily overcome by increasing the neighbourhood size owing to the computational savings mentioned above.
- The optimal setting is to use a multi-grid path and change from randomized to constant path after a few grid levels.

6.1.3 Perspective

In the course of my work, I developed a good understanding of SGS and the implications of each parameter involved as well as a sense of the computational expenses of each subroutine. Some key learnings which might be useful for SGS users are listed below:

- Although, ideal neighbourhood size varies with the type of variogram use, it is good practice to have a least 30 neighbours. For multi-grid path, the gains are marginal for more than 100 neighbors.
- Parallelization can and should be implemented at the path level.
- Neighbourhood search strategy can be essential. For most settings, spiral search is the recommended method. Spiral search can take advantage of a multi-grid path to search only the grid with previously simulated nodes. This implies that hard data are treated separately with either a partial sort or a superblock search if they are numerous and widespread. In case of very large grids and few neighbours, a partial sort method might be slightly faster.
- A trade-off between the time spent on improving the code and the time spent on waiting for a slow code always needs to be assessed and optimized.
- The suggested implementation varies depending on the main concerns:
  - For a simple and intuitive solution, use a random path without constant path.

  - For computationally efficient code, use a parallelized version of SGS with a constant path. It would be also useful to use a covariance table and a spiral path. A multi-grid path will allow you to reduce slightly the neighbourhood size (50 neighbours) and reduce computational time.
  - For highly accurate realizations, use a multi-grid path and a very large neighbourhood size. If you need many realizations, change the path from randomized to constant during the course of the multi-grid path. Using a covariance table and a spiral path might not be essential at this stage if your grid is not very large.
- Template matlab codes for all these versions of SGS are available on https://raphael-nussbaumer-phd.github.io/SGS/

6.2 Bayesian Sequential Simulation (BSS) with log-linear pooling

The objective of this study is to improve BSS by accounting for the interdependence of the primary and secondary variables through the implementation of a log-linear pooling operator. A major novelty of this work was to use various weighting schemes in the pooling operator, specifically, weights varying along the simulation path. This method is applied to the simulation of the fine-scale hydraulic conductivity distribution based on a coarse electrical conductivity tomogram as the secondary variable. Results show that:

- Accuracy increases with a more sophisticated weighting scheme and, correspondingly, so does computational time due to the calibration required.
- Our recommendation would be to avoid using traditional BSS and at use a constant weight of 0.5 or to include the marginal distribution. If possible, we recommend using a calibrated weighting scheme with a single parameter.
- The quantitative results might not be the same for all applications of BSS, but remain valid for applications with a smooth secondary variable.

6.3 Area-to-point kriging for resolution-limited tomogram

This work presents a methodology to simulate fine-scale realizations of a parameter based on a tomographic image of this parameter. It takes advantage of the known regularization of the tomographic inversion to inform an area-to-point kriging system of the linear relationship between the fine-scale parameters and the smooth tomogram. In this study, the method was applied to a synthetic case study of electrical resistivity. The results of the case study show that the realizations generated have the correct spatial structure and their upscaling reproduces faithfully the tomogram. However, when the forward model of the tomogram is applied to these realizations, a small error is observed in the data as compared to the synthetic data. An important advantage of this method is its computational efficiency compared to other inversion methods such as MCMC. This method could, in theory, be applied to any type of smoothness-constrained inversion for any type of measurement. For instance, we successfully performed preliminary tests to downscale hydraulic tomograms.

6.4 Final remarks

In this thesis, I have worked on four different projects which share the common objective to develop computationally efficient and accurate methodologies for the integration of spatial datasets which are variable in terms of coverage and resolution, and present complex relationships (e.g. non-linearity, non-uniqueness and partially redundant information). In particular, these contributions aim to improve imaging the subsurface through integrating geophysical datasets. Ultimately, these technical advances have the potential to improve the management of water resources.

Bibliography

Abdu, H., Robinson, D. A., Seyfried, M., and Jones, S. B. (2008). Geophysical imaging of watershed subsurface patterns and prediction of soil texture and water holding capacity. *Water Resources Research*, 44(4):W00D18.

Albaugh, J., Dye, P., and King, J. (2013). Eucalyptus and water use in SA. *Journal of forestry research*, 2013:1–11.

Allard, D. (2018). No Title. personal communication.

Allard, D., Comunian, A., and Renard, P. (2012). Probability Aggregation Methods in Geoscience. *Mathematical Geosciences*, 44(5):545–581.

Altwegg, K., Balsiger, H., Bar-Nun, A., Berthelier, J. J., Bieler, A., Bochsler, P., Briois, C., Calmonte, U., Combi, M., De Keyser, J., Eberhardt, P., Fiethe, B., Fuselier, S., Gasc, S., Gombosi, T. I., Hansen, K. C., Hassig, M., Jackel, A., Kopp, E., Korth, A., LeRoy, L., Mall, U., Marty, B., Mousis, O., Neefs, E., Owen, T., Reme, H., Rubin, M., Semon, T., Tzou, C. Y., Waite, H., and Wurz, P. (2015). 67P/Churyumov-Gerasimenko, a Jupiter family comet with a high D/H ratio. *Science*, 347(6220):1261952–1261952.

Alumbaugh, D. L. and Newman, G. A. (2000). Image appraisal for 2-D and 3-D electromagnetic inversion. *Geophysics*, 65(5):1455.

Anderson, M. P., Woessner, W. W., and Hunt, R. J. (2015). *Applied groundwater modeling: simulation of flow and advective transport*. Academic press.

Archie, G. (1942). The electrical resistivity log as an aid in determining some reservoir characteristics. *Transactions of the AIME*, 146(01):54–62.

Atibu, E. K., Devarajan, N., Thevenon, F., Mwanamoki, P. M., Tshibanda, J. B., Mpiana, P. T., Prabakar, K., Mubedi, J. I., Wildi, W., and Poté, J. (2013). Concentration of metals in surface water and sediment of Luilu and Musonoie Rivers, Kolwezi-Katanga, Democratic Republic of Congo. *Applied Geochemistry*, 39:26–32.

Atkinson, P. M. (2013). Downscaling in remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 22(1):106–114.

Atkinson, P. M., Pardo-Igúzquiza, E., and Chica-Olmo, M. (2008). Downscaling cokriging for super-resolution mapping of continua in remotely sensed images. *IEEE Transactions on Geoscience and Remote Sensing*, 46(2):573–580.

Baalousha, H. (2011). Fundamental of Grounwater Modeling. In Konig, L. and Weiss, J., editors, *Groundwater: Modelling, Management and Contamination*, chapter 4, pages 113–130. Nova Science Publishers, Inc.

Bair, E. S. and Metheny, M. A. (2011). Lessons learned from the landmark "A Civil Action" trial. *Ground Water*, 49(5):764–769.

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(4):825–848.

Barnsley, M. F., Devaney, R. L., Mandelbrot, B. B., Peitgen, H.-O., Saupe, D., and Voss, R. F. (1988). *The Science of Fractal Images*. Springer New York, New York, NY.

Barry, R. P. and Kelley Pace, R. (1997). Kriging with large data sets using sparse matrix techniques. *Communications in Statistics - Simulation and Computation*, 26(2):619–629.

Billings, S. D., Newsam, G. N., Beatson, R. K., and Newsam, G. N. (2002). Interpolation of geophysical data using continuous global surfaces. *Geophysics*, 67(6):1823–1834.

Binley, A., Hubbard, S. S., Huisman, J. a., Revil, A., Robinson, D. A., Singha, K., and Slater, L. D. (2015). The emergence of hydrogeophysics for improved understanding of subsurface processes over multiple scales. *Water Resources Research*, 51(6):3837–3866.

Binley, A. and Kemna, A. (2005). DC resistivity and induced polarization methods. In Rubin, Y. and Hubbard, S. S., editors, *Hydrogeophysics*, volume 50 of *Water Science and Technology Library*, pages 129–156. Springer Netherlands, Dordrecht.

Boisvert, J. B. and Deutsch, C. V. (2011). Programs for kriging and sequential Gaussian simulation with locally varying anisotropy using non-Euclidean distances. *Computers & Geosciences*, 37(4):495–510.

Boucher, A. (2007). Random Function : A New Formalism for Applied Geostatistics. Technical report, Stanford University.

Boulanger, F. (1990). *Modélisation et simulation de variables régionalisées par des fonctions aléatoires stables*. PhD thesis, Ecole des Mines de Paris, Fontainebleau.

Box, G. E. P. and Draper, N. R. (1987). *Empirical model-building and response surfaces*. John Wiley & Sons.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (2008). *Time Series Analysis*, volume 37. John Wiley & Sons, Inc., Hoboken, NJ.

Butler, J. J. (2005). Hydrogeological Methods for Estimation of Spatial Variations in Hydraulic Conductivity. In *Hydrogeophysics*, pages 23–58. Springer Netherlands, Dordrecht.

Cáceres, A., Emery, X., and Godoy, M. (2010). Speeding up conditional simulation: using Sequential Gaussian Simulaton with residual Substitution. Technical report.

Cassiani, G., Böhm, G., Vesnaver, A., and Nicolich, R. (1998). A geostatistical framework for incorporating seismic tomography auxiliary data into hydraulic conductivity estimation. *Journal of Hydrology*, 206(1-2):58–74.

Chen, J., Hubbard, S., and Rubin, Y. (2001). Estimating the hydraulic conductivity at the South Oyster Site from geophysical tomographic data using Bayesian techniques based on the normal linear regression model. *Water Resources Research*, 37(6):1603–1613.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, 49(4):327–335.

Chilès, J.-P. and Delfiner, P. (1999). *Geostatistics*, volume 497 of *Wiley Series in Probability and Statistics*. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Comunian, A., Renard, P., and Straubhaar, J. (2012). 3D multiple-point statistics simulation using 2D training images. *Computers & Geosciences*, 40:49–65.

Constable, S. C., Parker, R. L., and Constable, C. G. (1987). Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data. *GEOPHYSICS*, 52(3):289–300.

Copty, N., Rubin, Y., and Mavko, G. (1993). Geophysical-hydrological identification of field permeabilities through Bayesian updating. *Water Resources Research*, 29(8):2813–2825.

Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):209–226.

Dafflon, B., Irving, J., and Holliger, K. (2009). Simulated-annealing-based conditional simulation for the local-scale characterization of heterogeneous aquifers. *Journal of Applied Geophysics*, 68(1):60–70.

Daly, C. (2005). Higher order models using entropy, Markov random fields and sequential simulation. In *Geostatistics Banff 2004*, pages 215–224. Springer.

Darcy, H. (1856). *Les fontaines publiques de la ville de Dijon: exposition et application...* Victor Dalmont.

Day-Lewis, F. D. and Lane, J. W. (2004). Assessing the resolution-dependent utility of tomograms for geostatistics. *Geophysical Research Letters*, 31(7).

DeGroot-Hedlin, C. and Constable, S. (1990). Occam's inversion to generate smooth, two-dimensional models from magnetotelluric data. *GEOPHYSICS*, 55(12):1613–1624.

Delbari, M., Afrasiab, P., and Loiskandl, W. (2009). Using sequential Gaussian simulation to assess the field-scale spatial uncertainty of soil water content. *Catena*, 79(2):163–169.

Dentz, M., Le Borgne, T., Englert, A., and Bijeljic, B. (2011). Mixing, spreading and reaction in heterogeneous media: A brief review. *Journal of Contaminant Hydrology*, 120-121(C):1–17.

Deutsch, C. V. and Journel, A. G. (1992). *GSLIB: Geostatistical Software Library and User's Guide*. Oxford University Press, New York.

Deutsch, C. V. and Wen, X. H. (2000). Integrating Large-Scale Soft Data by Simulated Annealing and Probability Constraints. *Mathematical Geology*, 32(1):49–67.

Dimitrakopoulos, R., Farrelly, C. T., and Godoy, M. (2002). Moving forward from traditional optimization: grade uncertainty and risk effects in open-pit design. *Mining Technology*, 111(1):82–88.

Dimitrakopoulos, R. and Luo, X. (2004). Generalized Sequential Gaussian Simulation on Group Size and Screen-Effect Approximations for Large Field Simulations. *Mathematical Geology*, 36(5):567–591.

Doligez, B., Le Ravalec, M., Bouquet, S., and Adelinet, M. (2015). A review of three geostatistical techniques for realistic geological reservoir modeling integrating multi-scale data. *Bulletin of Canadian Petroleum Geology*, 63(4):277–286.

Doyen, P. M. and Boer, L. D. (1996). Bayesian sequential Gaussian simulation of lithology with non-linear data. page 11.

Dubreuil-Boisclair, C., Gloaguen, E., Marcotte, D., and Giroux, B. (2011). Heterogeneous aquifer characterization from ground-penetrating radar tomography and borehole hydrogeophysical data using nonlinear Bayesian simulations. *Geophysics*, 76(4):J13.

Emery, X. (2004). Testing the correctness of the sequential algorithm for simulating Gaussian random fields. *Stochastic Environmental Research and Risk Assessment*, 18(6):401–413.

Emery, X. and Peláez, M. (2011). Assessing the accuracy of sequential Gaussian simulation and cosimulation. *Computational Geosciences*, 15(4):673–689.

Englert, A., Kemna, A., Zhu, J.-f., Vanderborght, J., Vereecken, H., and Yeh, T.-C. J. (2016). Comparison of smoothness-constrained and geostatistically based cross-borehole electrical resistivity tomography for characterization of solute tracer plumes. *Water Science and Engineering*, 9(4):274–286.

EPA (2013). The Importance of Water to the U.S. Economy. Technical Report November, United States Environmental Protection Agency.

Essink, O. G. (2000). Groundwater Modelling.

Ezzedine, S., Rubin, Y., and Chen, J. (1999). Bayesian Method for hydrogeological site characterization using borehole and geophysical survey data: Theory and application to the Lawrence Livermore National Laboratory Superfund Site. *Water Resources Research*, 35(9):2671.

FAO (2016). AQUASTAT.

Fick, A. (1855). Ueber Diffusion. *Annalen der Physik und Chemie*, 170(1):59–86.

Fournier, A., Fussell, D., and Carpenter, L. (1982). Computer rendering of stochastic models. *Communications of the ACM*, 25(6):371–384.

Freeze, R. and Cherry, J. (1979). *Groundwater*.

Friedel, S. (2003). Resolution, stability and efficiency of resistivity tomography estimated from a generalized inverse approach. *Geophysical Journal International*, 153(2):305–316.

Furrer, R., Genton, M. G., and Nychka, D. (2006). Covariance Tapering for Interpolation of Large Spatial Datasets. *Journal of Computational and Graphical Statistics*, 15(3):502–523.

Genest, C. and Zidek, J. V. (1986). Combining Probability Distributions: A Critique and an Annotated Bibliography. *Statistical Science*, 1(1):147–148.

Giordano, M. (2009). Global Groundwater? Issues and Solutions. *Annual Review of Environment and Resources*, 34(1):153–178.

Gleeson, T., Befus, K. M., Jasechko, S., Luijendijk, E., and Cardenas, M. B. (2016). The global volume and distribution of modern groundwater. *Nature Geoscience*, 9(2):161–164.

Gleick, P. (2000). Water Conflict Chronology.

Gloaguen, E., Marcotte, D., and Chouteau, M. (2004). *A non-linear GPR tomographic inversion algorithm based on iterated cokriging and conditional simulations*, pages 409–418. Springer Netherlands, Dordrecht.

Gloaguen, E., Marcotte, D., Chouteau, M., and Perroud, H. (2005). Borehole radar velocity inversion using cokriging and cosimulation. *Journal of Applied Geophysics*, 57(4):242–259.

Gloaguen, E., Marcotte, D., Giroux, B., Dubreuil-Boisclair, C., Chouteau, M., and Aubertin, M. (2007). Stochastic borehole radar velocity and attenuation tomographies using cokriging and cosimulation. *Journal of Applied Geophysics*, 62(2):141–157.

Gómez-Hernández, J. and Wen, X.-H. (1998). To be or not to be multi-Gaussian? A reflection on stochastic hydrogeology. *Advances in Water Resources*, 21(1):47–61.

Gómez-Hernández, J. J. and Cassiraga, E. F. (1994). *Theory and Practice of Sequential Simulation*. Kluwer Academic Publishers, Dordrecht.

Gómez-Hernández, J. J. and Journel, A. G. (1993). Joint Sequential Simulation of Multi-Gaussian Fields. In Soares, A., editor, *Geostatistics Tróia '92*, Quantitative Geology and Geostatistics, pages 85–94. Springer Netherlands, Dordrecht.

Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. Oxford University Press.

Goovaerts, P. (2001). Geostatistical modelling of uncertainty in soil science. *Geoderma*, 103(1-2):3–26.

Goovaerts, P. (2010). Combining areal and point data in geostatistical interpolation: Applications to soil science and medical geography. *Mathematical Geosciences*, 42(5):535–554.

Goovaerts, P. (2011). A coherent geostatistical approach for combining choropleth map and field data in the spatial interpolation of soil properties. *European Journal of Soil Science*, 62(3):371–380.

Gotway, C. A. and Young, L. J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association*, 97(458):632–648.

Gribov, A. and Krivoruchko, K. (2004). Geostatistical Mapping with Continuous Moving Neighborhood. *Mathematical Geology*, 36(2):267–281.

Guardiano, F. B. and Srivastava, R. M. (1993). Multivariate Geostatistics: Beyond Bivariate Moments. In Soares, A., editor, *Geostatistics Tròia '92*, pages 133–144. Springer, Dordrecht.

Halton, J. H. (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, 2(1):84–90.

Hansen, T. M., Cordua, K. S., and Mosegaard, K. (2012). Inverse problems with non-trivial priors: efficient solution through sequential Gibbs sampling. *Computational Geosciences*, 16(3):593–611.

Hansen, T. M., Journel, A. G., Tarantola, A., and Mosegaard, K. (2006). Linear inverse Gaussian theory and geostatistics. *Geophysics*, 71(6):R101–R111.

Hansen, T. M. and Mosegaard, K. (2008). VISIM: Sequential simulation for linear inverse problems. *Computers & Geosciences*, 34(1):53–76.

Harr, J. (1995). *A civil action*. Random House.

Hartman, L. and Hössjer, O. (2008). Fast kriging of large data sets with Gaussian Markov random fields. *Computational Statistics & Data Analysis*, 52(5):2331–2349.

Hassanpour, R. M. and Leuangthong, O. (2006). On the Use of a Quadtree Search for Estimation and Simulation. pages 1–11.

Heinz, J., Kleineidam, S., Teutsch, G., and Aigner, T. (2003). Heterogeneity patterns of Quaternary glaciofluvial gravel bodies (SW-Germany): application to hydrogeology. *Sedimentary Geology*, 158(1-2):1–23.

Hoare, C. A. R. (1962). Quicksort. *The Computer Journal*, 5(1):10–16.

Hoeksema, R. J. and Kitanidis, P. K. (1984). An application of the geostatistical approach to the inverse problem in two-dimensional groundwater modeling. *Water Resources Research*, 20(7):1003–1020.

Hoffimann, J., Scheidt, C., Barfod, A., and Caers, J. (2017). Stochastic simulation by image quilting of process-based geological models. *Computers & Geosciences*, 106(February):18–32.

Holdren, J. P., Lander, E., and Varmus, H. (2010). Report to the president and congress: Designing a digital future: federally funded research and development in networking and information technology. *Executive Office of the President and President's Council of Advisors on Science and Technology*, page 148.

Holman, I. P., Whelan, M. J., Howden, N. J. K., Bellamy, P. H., Willby, N. J., Rivas-Casado, M., and McConvey, P. (2008). Phosphorus in groundwater-an overlooked contributor to eutrophication? *Hydrological Processes*, 22(26):5121–5127.

Hu, L. Y., Le Ravalec, M., and Blanc, G. (2001). Gradual deformation and iterative calibration of truncated Gaussian simulations. *Petroleum Geoscience*, 7(S):S25–S30.

Hubbard, S. S., Chen, J., Peterson, J., Majer, E. L., Williams, K. H., Swift, D. J., Mailloux, B., and Rubin, Y. (2001). Hydrogeological characterization of the south oyster bacterial transport site using geophysical data. *Water Resources Research*, 37(10):2431–2456.

Hubbard, S. S. and Linde, N. (2010). Hydrogeophysics. Technical report, Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States).

Hubbard, S. S. and Rubin, Y. (2005). Introduction to Hydrogeophysics. In Rubin, Y. and Hubbard, S. S., editors, *Hydrogeophysics*, volume 50 of *Water Science and Technology Library*, pages 3–21. Springer Netherlands, Dordrecht.

Hulme, P., Fletcher, S., and Brown, L. (2002). Incorporation of groundwater modelling in the sustainable management of groundwater resources. *Geological Society Special Publication*, 193:83–90.

Hyndman, D. W. and Gorelick, S. M. (1996). Estimating Lithologic and Transport Properties in Three Dimensions Using Seismic and Tracer Data: The Kesterson aquifer. *Water Resources Research*, 32(9):2659–2670.

Ireson, A., van der Kamp, G., Ferguson, G., Nachshon, U., and Wheater, H. S. (2012). Hydrogeological processes in seasonally frozen northern latitudes: understanding, gaps and challenges. *Hydrogeology Journal*, 21(1):53–66.

Isaaks, E. H. (1984). Indicator simulation: application to the simulation of a high grade uranium mineralization. In Verly, G. W., editor, *Geostatistics for natural resources characterization. Part 2*, pages 1057–1069. D. Reidel Publishing.

Isaaks, E. H. (1991). *The application of Monte Carlo methods to the analysis of spatially correlated data*. PhD thesis, Stanford University.

Isaaks, E. H. and Srivastava, R. M. (1989). *An Introduction to Applied Geostatistics*. Oxford University Press, New York.

Johnson, M. E. (1987). *Multivariate Statistical Simulation*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Journel, A. G. (1974). Geostatistics for Conditional Simulation of Ore Bodies. *Economic Geology*, 69(5):673–687.

Journel, A. G. (1989). *Fundamentals of Geostatistics in Five Lessons*, volume 16. American Geophysical Union, Washington, D. C.

Journel, A. G. (2002). Combining Knowledge From Diverse Sources: An Alternative to Traditional Data Independence Hypotheses. *Mathematical Geology*, 34(5):573–596.

Journel, A. G. and Huijbregts, C. J. (1978). *Mining geostatistics*. Academic press, London.

Juang, K.-W., Chen, Y.-S., and Lee, D.-Y. (2004). Using sequential indicator simulation to assess the uncertainty of delineating heavy-metal contaminated soils. *Environmental Pollution*, 127(2):229–238.

Kammann, E. E. and Wand, M. P. (2003). Geoadditive models. *Applied Statistics*, 52(1):1–22.

Khan, A. E., Ireson, A., Kovats, S., Mojumder, S. K., Khusru, A., Rahman, A., and Vineis, P. (2011). Drinking water salinity and maternal health in coastal Bangladesh: Implications of climate change. *Environmental Health Perspectives*, 119(9):1328–1332.

Kitanidis, P. K. (1995). Quasi-linear geostatistical theory for inversing. *Water Resources Research*, 31(10):2411–2419.

Kitanidis, P. K. and Vomvoris, E. G. (1983). A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations. *Water Resources Research*, 19(3):677–690.

Kocis, L. and Whiten, W. J. (1997). Computational investigations of low-discrepancy sequences. *ACM Transactions on Mathematical Software*, 23(2):266–294.

Korzoun, V. I. and Sokolov, A. A. (1978). World water balance and water resources of the earth. *Water Development, Supply and Management (UK)(USA)(Canada)(Australia)(France)(Germany, FR)*.

Kuruppu, N. and Liverman, D. (2011). Mental preparation for climate adaptation: The role of cognition and culture in enhancing adaptive capacity of water management in Kiribati. *Global Environmental Change*, 21(2):657–669.

Kyriakidis, P. C. (2004). A geostatistical framework for area-to-point spatial interpolation. *Geographical Analysis*, 36(3):259–289.

Kyriakidis, P. C. and Yoo, E. H. (2005). Geostatistical prediction and simulation of point values from areal data. *Geographical Analysis*, 37(2):124–151.

Lantuéjoul, C. (1991). Ergodicity and integral range. *Journal of Microscopy*, 161(3):387–403.

Lantuéjoul, C. (2002). *Geostatistical Simulation*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Le Maitre, D. C., Scott, D. F., and Colvin, C. (1999). A review of information on interactions between vegetation and groundwater. *Water SA*, 25(2):137–152.

Le Ravalec-Dupin, M., Noetinger, B., and Hu, L. Y. (2000). The FFT moving average (FFT-MA) generator: An efficient numerical method for generating and conditioning Gaussian simulations. *Mathematical Geology*, 32(6):701–723.

Lee, S.-Y., Carle, S. F., and Fogg, G. E. (2007). Geologic heterogeneity and a comparison of two geostatistical models: Sequential Gaussian and transition probability-based geostatistical simulation. *Advances in Water Resources*, 30(9):1914–1932.

Leuangthong, O., McLennan, J. A., and Deutsch, C. V. (2004). Minimum Acceptance Criteria for Geostatistical Realizations. *Natural Resources Research*, 13(3):131–141.

Li, W., Nowak, W., and Cirpka, O. A. (2005). Geostatistical inverse modeling of transient pumping tests using temporal moments of drawdown. *Water Resources Research*, 41(8):1–13.

Lin, Y.-P., Chang, T.-K., and Teng, T.-P. (2001). Characterization of soil lead by comparing sequential Gaussian simulation, simulated annealing simulation and kriging methods. *Environmental Geology*, 41(1-2):189–199.

Linde, N., Chen, J., Kowalsky, M. B., and Hubbard, S. (2006). HYDROGEOPHYSICAL PARAMETER ESTIMATION APPROACHES FOR FIELD SCALE CHARACTERIZATION. In *Applied Hydrogeophysics*, volume 71, pages 9–44. Springer Netherlands, Dordrecht.

Linde, N., Renard, P., Mukerji, T., and Caers, J. (2015). Geological realism in hydrogeological and geophysical inverse modeling: A review. *Advances in Water Resources*, 86:86–101.

Liu, X. H., Kyriakidis, P. C., and Goodchild, M. F. (2008). Population-density estimation using regression and area-to-point residual kriging. *International Journal of Geographical Information Science*, 22(4):431–447.

Liu, Y. and Journel, A. G. (2009). A package for geostatistical integration of coarse and fine scale data. *Computers & Geosciences*, 35(3):527–547.

Manchuk, J. G. and Deutsch, C. V. (2012). Implementation aspects of sequential Gaussian simulation on irregular points. *Computational Geosciences*, 16(3):625–637.

Mariethoz, G. (2010). A general parallelization strategy for random path based geostatistical simulation methods. *Computers & Geosciences*, 36(7):953–958.

Mariethoz, G., Renard, P., Cornaton, F., and Jaquet, O. (2009a). Truncated Plurigaussian Simulations to Characterize Aquifer Heterogeneity. *Ground Water*, 47(1):13–24.

Mariethoz, G., Renard, P., and Froidevaux, R. (2009b). Integrating collocated auxiliary parameters in geostatistical simulations using joint probability distributions and probability aggregation. *Water Resources Research*, 45(8):n/a–n/a.

Martínez, C. (2004). Partial Quicksort. In *6th ACM-SIAM Workshop on Algorithm Engineering and Experiments and 1st ACM-SIAM Workshop on Analytic Algorithmics and Combinatorics*, page 5.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58(8):1246–1266.

Matheron, G. (1965). *Les Variables Régionalisées et Leur Estimation*. Masson et Cie, Paris.

Matheron, G. (1989). *Estimating and Choosing*. Springer Berlin Heidelberg, Berlin, Heidelberg.

McKenna, S. A. and Poeter, E. P. (1995). Field example of data fusion in site characterization. *Water Resources Research*, 31(12):3229–3240.

McLennan, J. (2002). The Effect of the Simulation Path in Sequential Gaussian Simulation. Technical report, University of Alberta.

Memarsadeghi, N. and Mount, D. M. (2007). Efficient Implementation of an Optimal Interpolator for Large Spatial Data Sets. In *Computational Science–ICCS 2007*, pages 503–510. Springer.

Memarsadeghi, N., Raykar, V. C., Duraiswami, R., and Mount, D. M. (2008). Efficient kriging via fast matrix-vector products. *IEEE Aerospace Conference Proceedings*.

Menke, W. (1989). *Geophysical Data Analysis: Discrete Inverse Theory*, volume 45. Academic press.

Metai, E. (1997). Vulnerability of freshwater lenses on Tarawa: The role of hydrological monitoring in determining sustainable yield. *Proceedings of the Pacific Regional Consultation on Water in Small Island Countries*, pages 65–76.

Meyer, T. H. (2004). The Discontinuous Nature of Kriging Interpolation for Digital Terrain Modeling. *Cartography and Geographic Information Science*, 31(4):209–216.

Moore, G. E. (1998). Cramming more components onto integrated circuits. *Proceedings of the IEEE*, 86(1):82–85.

Morris, B. L., Lawrence, A. R., Chilton, P., Adams, B., Calow, R. C., and Klinck, B. A. (2003). *Groundwater and its susceptibility to degradation: a global assessment of the problem and options for management*, volume 3. United Nations Environment Programme.

Mowrer, H. (1997). Propagating uncertainty through spatial estimation processes for old-growth subalpine forests using sequential Gaussian simulation in GIS. *Ecological Modelling*, 98(1):73–86.

Moysey, S. and Knight, R. (2004). Modeling the field-scale relationship between dielectric constant and water content in heterogeneous systems. *Water Resources Research*, 40(3):1–10.

Moysey, S., Singha, K., and Knight, R. (2005). A framework for inferring field-scale rock physics relationships through numerical simulation. *Geophysical Research Letters*, 32(8):1–4.

Newman, G. A. and Alumbaugh, D. L. (2000). Three-dimensional magnetotelluric inversion using non-linear conjugate gradients. *Geophysical Journal International*, 140(2):410–424.

Nunes, R. and Almeida, J. a. (2010). Parallelization of sequential Gaussian, indicator and direct simulation algorithms. *Computers & Geosciences*, 36(8):1042–1052.

Nussbaumer, R., Mariethoz, G., Gloaguen, E., and Holliger, K. (2018a). Which Path to Choose in Sequential Gaussian Simulation. *Mathematical Geosciences*, 50(1):97–120.

Nussbaumer, R., Mariethoz, G., Gravey, M., Gloaguen, E., and Holliger, K. (2018b). Accelerating Sequential Gaussian Simulation with a constant path. *Computers & Geosciences*, 112(2018):121–132.

Omre, H., Sølna, K., and Tjelmeland, H. (1993). Simulation of Random Functions on Large Lattices. In Soares, A., editor, *Geostatistics Tròia '92*, pages 179–199. Kluwer Academic Publishers, Dordrecht.

Pardo-Iguzquiza, E., Atkinson, P. M., and Chica-Olmo, M. (2010). DSCOKRI: A library of computer programs for downscaling cokriging in support of remote sensing applications. *Computers & Geosciences*, 36(7):881–894.

Pardo-Igúzquiza, E., Chica-Olmo, M., and Atkinson, P. M. (2006). Downscaling cokriging for image sharpening. *Remote Sensing of Environment*, 102(1-2):86–98.

Pebesma, E. J. and Wesseling, C. G. (1998). Gstat: a program for geostatistical modelling, prediction and simulation. *Computers & Geosciences*, 24(1):17–31.

Pirot, G., Linde, N., Mariethoz, G., and Bradford, J. H. (2017). Probabilistic inversion with graph cuts: Application to the Boise Hydrogeophysical Research Site. *Water Resources Research*, 53(2):1231–1250.

Poeter, E. and Gaylord, D. R. (1990). Influence of Aquifer Heterogeneity on Contaminant Transport at the Hanford Site.

Ramirez, A. L., Nitao, J. J., Hanley, W. G., Aines, R., Glaser, R. E., Sengupta, S. K., Dyer, K. M., Hickling, T. L., and Daily, W. D. (2005). Stochastic inversion of electrical resistivity changes using a Markov Chain Monte Carlo approach. *Journal of Geophysical Research: Solid Earth*, 110(B2):1–18.

Rasera, L. G., Machado, P. L., and Costa, J. F. C. L. (2015). A conflict-free, path-level parallelization approach for sequential simulation algorithms. *Computers & Geosciences*, 80:49–61.

Renard, P. and Allard, D. (2013). Connectivity metrics for subsurface flow and transport. *Advances in Water Resources*, 51:168–196.

Rivoirard, J. (1984). *Le comportement des poids de krigeage*. PhD thesis, Ecole des Mines de Paris, Fontainebleau.

Rivoirard, J. and Romary, T. (2011). Continuity for Kriging with Moving Neighborhood. *Mathematical Geosciences*, 43(4):469–481.

Rubin, Y. and Hubbard, S. (2005). Stochastic Forward and Inverse Modeling: The "Hydrogeophysical" Challenge. In *Hydrogeophysics*, pages 487–511. Springer Netherlands, Dordrecht.

Rubin, Y., Mavko, G., and Harris, J. (1992). Mapping permeability in heterogeneous aquifers using hydrologic and seismic data. *Water Resources Research*, 28(7):1809–1816.

Rue, H. and Tjelmeland, H. (2002). Fitting Gaussian Markov Random Fields to Gaussian Fields. *Scandinavian Journal of Statistics*, 29(1):31–49.

Ruggeri, P., Gloaguen, E., Lefebvre, R., Irving, J., and Holliger, K. (2014). Integration of hydrological and geophysical data beyond the local scale: Application of Bayesian sequential simulation to field data from the Saint-Lambert-de-Lauzon site, Québec, Canada. *Journal of Hydrology*, 514:271–280.

Ruggeri, P., Irving, J., Gloaguen, E., and Holliger, K. (2013). Regional-scale integration of multiresolution hydrological and geophysical data using a two-step Bayesian sequential simulation approach. *Geophysical Journal International*, 194(1):289–303.

Rumsfeld, D. H. (2002). Transcript: DoD News Briefing - Secretary Rumsfeld and Gen. Myers.

Ruprecht, J. K. and Stoneman, G. L. (1993). Water yield issues in the jarrah forest of south-western Australia. *Journal of Hydrology*, 150(2-4):369–391.

Safikhani, M., Asghari, O., and Emery, X. (2017). Assessing the accuracy of sequential gaussian simulation through statistical testing. *Stochastic Environmental Research and Risk Assessment*, 31(2):523–533.

Sakata, S., Ashida, F., and Zako, M. (2004). An efficient algorithm for kriging approximation and optimization with large-scale sampling data. *Computer Methods in Applied Mechanics and Engineering*, 193(3-5):385–404.

Scanlon, B. R., Faunt, C. C., Longuevergne, L., Reedy, R. C., Alley, W. M., McGuire, V. L., and McMahon, P. B. (2012). Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley. *Proceedings of the National Academy of Sciences*, 109(24):9320–9325.

Schön, J. H. (2004). *Physical properties of rocks: Fundamentals and principles of petrophysics*, volume 65. Elsevier.

Shiklomanov, I. A. (2000). Appraisal and Assessment of World Water Resources. *Water International*, 25(1):11–32.

Singha, K., Day-Lewis, F. D., and Moysey, S. (2007). Accounting for tomographic resolution in estimating hydrologic properties from geophysical data. In *Subsurface Hydrology: Data Integration for Properties and Processes*, pages 227–241. American Geophysical Union, Washington, D.C.

Srinivasan, B. V., Duraiswami, R., and Murtugudde, R. (2008). Efficient kriging for real-time spatio-temporal interpolation Linear kriging. *20th Conference on Probablility and Statistics in Atmospheric Sciences*, pages 228–235.

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. SIAM.

Tarantola, A. and Valette, B. (1982). Generalized nonlinear inverse problems solved using the least squares criterion. *Reviews of Geophysics*, 20(2):219–232.

Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46:234.

Tóth, J. (1963). A Theoretical Analysis of Groundwater Flow in Small Drainage Basins. *J.Geophys. Res.*, 68(16):4785–4812.

Tran, T. T. (1994). Improving variogram reproduction on dense simulation grids. *Computers & Geosciences*, 20(7-8):1161–1168.

Trefethen, L. N. and Bau III, D. (1997). *Numerical Linear Algebra*, volume 50. SIAM.

UNICEF and WHO (2017). *Progress on Drinking Water, Sanitation and Hygiene*.

van Ast, L., Maclean, R., and Sireyjol, A. (2013). Valuing water to drive more effective decisions. Technical report, Trucost.

Van Genuchten, M. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil science society of America journal*, 44(5):892–898.

Vargas, S. H., Caetano, H., and Filipe, M. (2007). Parallelization of Sequential Simulation Procedures. In *EAGE Conference on Petroleum Geostatistics*.

Verly, G. W. (1993). Sequential Gaussian Cosimulation: A Simulation Method Integrating Several Types of Information. In Soares, A., editor, *Geostatistics Tròia '92*, pages 543–554. Kluwer Academic Publishers.

Vonlanthen, P., Bittner, D., Hudson, A. G., Young, K. A., Müller, R., Lundsgaard-Hansen, B., Roy, D., Di Piazza, S., Largiader, C. R., and Seehausen, O. (2012). Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature*, 482(7385):357–362.

Vorosmarty, C. J. (2000). Global Water Resources: Vulnerability from Climate Change and Population Growth. *Science*, 289(5477):284–288.

Wada, Y., Van Beek, L. P., Van Kempen, C. M., Reckman, J. W., Vasak, S., and Bierkens, M. F. (2010). Global depletion of groundwater resources. *Geophysical Research Letters*, 37(20):1–5.

Warne, K. (2015). Will Pacific Island Nations Disappear as Seas Rise? Maybe Not.

Webb, A. P. and Kench, P. S. (2010). The dynamic response of reef islands to sea-level rise: Evidence from multi-decadal analysis of island change in the Central Pacific. *Global and Planetary Change*, 72(3):234–246.

WHO and UNICEF (2012). *Progress on Drinking Water and Sanitation 2012*. New York.

WHYMAP (2002). World-wide Hydrogeological Mapping and Assessment Programme (WHYMAP).

WWAP (2012). The United Nations World Water Development Report 4: Managing Water under Uncertainty and Risk. In *UN Water Reports*, volume 1, page 909. UNESCO, Paris.

Xu, W., Tran, T., Srivastava, R., and Journel, A. (1992). Integrating Seismic Data in Reservoir Modeling: The Collocated Cokriging Alternative. In *Proceedings of SPE Annual Technical Conference and Exhibition*, pages 833–842. Society of Petroleum Engineers.

Yeh, T.-C. J., Liu, S., Glass, R. J., Baker, K., Brainard, J. R., Alumbaugh, D. L., and LaBrecque, D. (2002). A geostatistically based inverse model for electrical resistivity surveys and its applications to vadose zone hydrology. *Water Resources Research*, 38(12):14–1–14–13.

Yoo, E. H. and Kyriakidis, P. C. (2006). Area-to-point Kriging with inequality-type data. *Journal of Geographical Systems*, 8(4):357–390.

Yoo, E. H. and Kyriakidis, P. C. (2009). Area-to-point Kriging in spatial hedonic pricing models. *Journal of Geographical Systems*, 11(4):381–406.

Zaillian, S. (1999). A Civil Action.

Zektser, I. S. and Everett, L. G. (2004). *Groundwater resources of the world and their use*, volume 6.

Zhao, Y., Xu, X., Huang, B., Sun, W., Shao, X., Shi, X., and Ruan, X. (2007). Using robust kriging and sequential Gaussian simulation to delineate the copper- and lead-contaminated areas of a rapidly industrialized city in Yangtze River Delta, China. *Environmental Geology*, 52(7):1423–1433.

Zinn, B. and Harvey, C. F. (2003). When good statistical models of aquifer heterogeneity go bad: A comparison of flow, dispersion, and mass transfer in connected and multivariate Gaussian hydraulic conductivity fields. *Water Resources Research*, 39(3):1–19.