A Practical Framework to Characterize Non-Stationary Regionalized Variables

by

Jianan Qu

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Mining Engineering

Department of Civil and Environmental Engineering

University of Alberta

Abstract

Geostatistics aims at applying statistics to quantitatively describe geological deposits and assess the uncertainty due to incomplete sampling. Strong assumptions are required regarding the location-independence of statistical parameters to construct numerical models with geostatistical tools. Most geological data often exhibit trends or non-stationary location-dependent features. Such location-dependence in the average grade violates common geostatistical assumptions and the results may be suboptimal. Non-stationary geostatistical techniques have been developed; however, there are concerns with all current approaches of dealing with location-dependent mean values. Developing a practical framework that accounts for location-dependent features would improve geostatistical models in these situations.

This dissertation develops a practical geostatistical modeling workflow to account for the deterministic and stochastic features of continuous regionalized grade variables. The main contributions of the thesis include: (1) the construction of a deterministic trend model that realistically represents the primary features of the geological process. A mathematical form for the trend is established; (2) the development of an objective function to optimize the calculation of the large-scale trend features. The objective function is established with a synthetic

example to minimize the mean squared error values between the modeled trend and the true trend for cases when the true trend is known; (3) the incorporation of the non-stationary features into geostatistical modeling. A parametric stepwise conditional transformation is considered to provide a stable and artifact-free numerical model. The complex features of the regionalized variable in the presence of a trend are removed in the forward transformation and restored in the back transformation; (4) the demonstration of the proposed techniques with real data. The trend is modeled with corrected parameters in an objective manner, and further, considered in subsequent geostatistical modeling for a more robust and reliable result.

The importance of the trend for improving performance in resource estimates is demonstrated in this thesis. The ultimate goal of this research is to improve resource management in the presence of a deterministic trend.

Preface

This thesis is an original work by Jianan Qu. Parts of the research project have been previously published, or are ready for the journal submission.

Chapter 3 is composed in part by Qu, J. and Deutsch, C. V., *Practical Trend Modeling for Geostatistical Simulation*, submitted as an original research article.

Chapters 4 and 6 are composed in part by Qu, J. and Deutsch, C. V., (2018), *Geostatistical Simulation with a Trend Using Gaussian Mixture Models*, published by Natural Resources Research as an original research article.

Chapter 6 is composed in part by Qu, J. and Deutsch, C. V., (2017), *A Case Study on Geostatistical Simulation with a Trend*, presented at 38th International Symposium on the Application of Computers and Operations Research in the Mineral Industry (APCOM 2017), as an original work.

Dedication

To my parents, **Shaoan** and **Shufen**, for their support and encouragement.

To my husband, **Yimin**, for his caring and infinite love.

Everything you do mean the world to me!

Acknowledgements

I would like to special express my thank to my supervisor Dr. Clayton V. Deutsch, the professor and the director of the Centre for Computational Geostatistics (CCG), for his continuous support, guidance, patience, and invaluable advice in my research. I appreciate his vast knowledge and skills in many areas. I also thank CCG sponsors from the industry for the financial support of my research.

Besides my supervisor, I would like to thank the rest of my thesis committee: Dr. Derek Apel, Dr. Jeffery B. Boisvert, Dr. Yashar Pourrahimian, Dr. Wei Victor Liu and Dr. Jorge Kazuo Yamamoto for their encouragement and insightful comments. The comments provide improvements as I start the next stage of my academic life.

It is a pleasure to thank Alice Da Silva and Arlene Figley for providing a stress-free research environment and support during my time at University of Alberta.

I would like to thanks all CCG colleagues for providing an enjoyable research environment and expanding my geostatistical knowledge beyond the classroom. We are not only able to support each other by deliberating over academic knowledge, but also by sharing the daily life.

Contents

List of Tables

List of Figures

Abbreviations

1-D	**1** (**One**) - **D**imensional
2-D	**2** (**Two**) - **D**imensional
3-D	**3** (**Three**) - **D**imensional
BLUE	**B**est **L**inear **U**nbiased **E**stimation
CDF	**C**umulative **D**istribution **F**unction
CS	**C**onditional **S**imulation
DH	**D**rill**H**ole
EM	**E**xpectation **M**aximization
GMM	**G**aussian **M**ixture **M**odel
GSLIB	**G**eostatistical **S**oftware **LIB**rary
IDW	**I**nverse **D**istance **W**eighting
IRF-k	**I**ntrinsic **R**andom **F**unction with Order k
KC	**K**riging with **C**ondtional Estimate
KDE	**K**ernel **D**ensity **E**stimation
KED	**K**riging with an **E**xternal **D**rift
KR	**K**riging the **R**esiduals
KT	**K**riging with a **T**rend
KU	**K**riging with **U**ncondtional Simulation
LD	**L**ag **D**istance
LU	**L**ower Triangular Matrix - **U**pper Triangular Matrix Simulation

LVM	**L**ocall **V**arying **M**ean
MAF	**M**inimum/Maximum **A**utocorrelation **F**actors
MCS	**M**onte **C**arlo **S**imulation
MSE	**M**ean **S**quared **E**rror
NS	**N**ormal **S**core Transformation
OK	**O**rdinary **K**riging
PDF	**P**robability **D**istribution **F**unction
PPMT	**P**rojection **P**ursuit **M**ultivariate **T**ransform
RF	**R**andom **F**unction
RV	**R**andom **V**ariable
ReV	**Re**gionalized **V**ariable
st.dev.	**S**tandard **D**eviation
SC	**S**tepwise **C**conditional Transformation
SGS	**S**equential **G**aussian **S**imulation
SK	**S**imple **K**riging
SK-LVM	**S**imple **K**riging with **L**ocally **V**arying **M**ean
SP	**S**moothing **P**arameter
TBS	**T**urning **B**ands **S**imulation
UC	**U**n**C**ondition Simulation
UK	**U**niversal **K**riging

Chapter 1

Introduction

Resource management should be undertaken in the presence of the best possible understanding of the natural resources under consideration. A numerical model is an essential tool to describe and communicate the behaviour of the natural resources. These numerical models should represent the natural resources with the greatest possible geological realism. Resource management depends on such realistic numerical models for the best possible decisions.

Geostatistics provides tools to construct realistic numerical models and supports mineral and petroleum resource estimates. Geostatistical simulation techniques generate multiple realizations that are used to characterize the geological heterogeneity and uncertainty. The difference between multiple realizations represents the geological uncertainty. This uncertainty is combined with other aspects of a project to provide a quantified basis for decision-making.

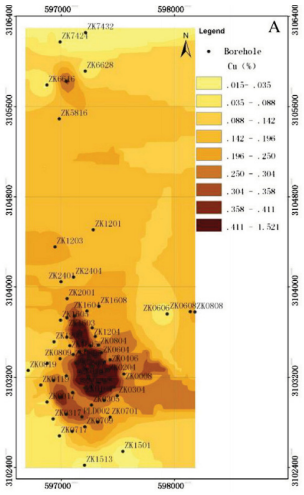
Geostatistical modeling requires parameters and assumptions because of the limited data available for a particular deposit. The parameters include global probability distributions, variograms and training images. A common assumption of geostatistical modeling is that the parameters are location invariant, for example, the expected value or the mean value of grades is constant for all

locations within each domain. This assumption is part of the decision of stationarity. Many techniques are available for conventional geostatistical modeling of stationary regionalized variables.

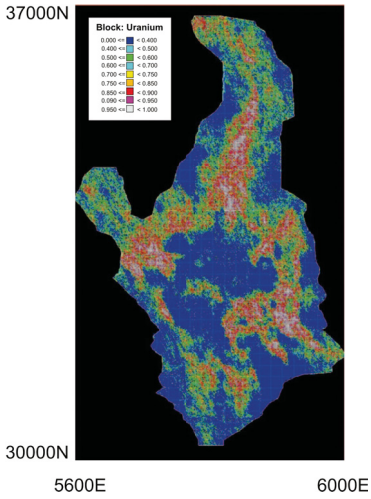
1.1 Problem Statement

A deterministic trend is often seen in geological phenomena, for example, the Copper distribution in the Pulang Deposit (Wang et al., 2012) and the plan view of Uranium recovery at the 450-meter elevation in the Olympic Dam Deposit (Boisvert et al., 2013; Rossi and Deutsch, 2014), see Figure 1.1. The geological variables show a considerable amount of variations in addition to the deterministic component. This deterministic component represents large-scale spatially continuous geological information. The small-scale stochastic component carries information on the high-frequency variations of the geological process. The numerical geological models must consider both deterministic and stochastic features.

The non-stationary spatial features should be modeled properly so that final predictions are accurate and precise. This would lead to improved decision making. The trend model should be smooth and contain the large-scale deterministic knowledge of the geologic data. There are several techniques for modeling the large-scale deterministic trend including regression, kriging, moving window averages and others (Krige, 1976; Journel and Huijbregts, 1978; Journel and Rossi, 1989; Goovaerts, 1997; Brunsdon et al., 1998; Gonzalez et al., 2006; Hong and Deutsch, 2009; Rossi and Deutsch, 2014). However, the inference of the deterministic component is difficult since there is no unique objective approach to define the trend parameters. In addition, there is no approach to quantitatively judge the goodness of a trend model, see Figure 1.2. An underfitting may not capture important features from the geological property, while



(a) Copper grades (Wang et al., 2012)



(b) Uranium recovery (Rossi and Deutsch, 2014). This example is based on the paper by Boisvert et al. (2013)

Figure 1.1: Trend appearance in two real case studies showing high and low values in large regions

an over-fitting can result in a lower error but an unrealistic model. The trend model should be constructed without under-fitting or over-fitting.

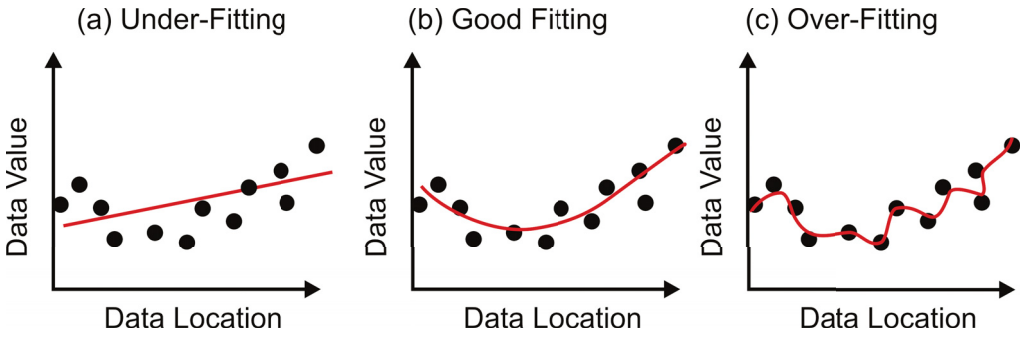


Figure 1.2: Schematic illustration of the judgment on the goodness of a trend

Different theories for non-stationary geostatistics have been developed. One common approach to non-stationary geostatistics is data decomposition into a deterministic trend model and a random fluctuation (Krumbein, 1959; Watson, 1972; Delfiner, 1976; Journel and Rossi, 1989; Wackernagel, 2003; Chiles and

Delfiner, 2012). The deterministic trend is modeled and then subtracted from the data values. The conventional kriging estimation is applied to the residuals assuming they are stationary. The local trend values are added back after the residuals have been estimated. The final result strongly depends on how the deterministic component has been estimated.

Another approach of dealing with non-stationary variables is to consider a conditional transformation. A modified normal score transformation was developed by Leuangthong and Deutsch (2003). The residual component is transformed conditional to the deterministic trend to arrive at a more stationary variable. The reproduction of complicated relationship between the trend and its residual is reasonable. However, the application has encountered difficulties that using the residuals is unstable and the results show binning artifacts. A new non-stationary geostatistical algorithm should be developed to increase the reliability of resource management.

Proposed Thesis Statement: *A flexible trend modeling implementation will improve high-resolution geostatistical models; therefore, improving the performance of the local estimates and the prediction of resources.*

1.2 Dissertation Contributions

This research focuses on developing numerical geological models of non-stationary variables. The goal is to improve resource estimates by ensuring the reproduction of the large-scale and short-scale variabilities of realistic geological features. Four main areas of research are presented to address current shortcomings: (1) develop a flexible trend modeling algorithm accounting for the large-scale continuous features; (2) establish an optimality criterion for trend modeling; (3) propose an

improved numerical geostatistical modeling approach in the presence of a deterministic trend; and (4) develop a case study that demonstrates the improvement brought by the proposed methods. The objectives will be briefly discussed below:

- A flexible mathematical form of the deterministic trend will be considered. The weighting function, anisotropy and other considerations will be established. The purpose is to characterize the deterministic trend with the correct parameters.

- An optimality criterion for trend modeling will be proposed. The goal is to propose a robust and reliable measure that leads to reliable trend models and improved predictions at unsampled locations.

- A geostatistical simulation algorithm in the presence of a deterministic trend will be developed. The developed geostatistical algorithm must account for the deterministic component in an artifact-free fashion. The developed algorithm will improve the high-resolution geological modeling.

- The improved geological models will be evaluated by case studies to validate the proposed methods. The primary goal is to evaluate the performance of the proposed model relative to conventional techniques.

The aim of the proposed research is to increase the robustness and reliability of geostatistical modeling algorithms and reduce uncertainty. Realistic data will be considered to show the improvement brought by the proposed methods in the estimation.

1.3 Assumptions and Limitations

Some assumptions related to the proposed non-stationary algorithm should be mentioned.

The research aims to extract as much useful information as possible from the geologic data. The complexity of real-life geological phenomena may not be amenable to an expression in terms of simplified mathematical algorithms. An assumption that the mathematical notions can be used to approximate the geologic data is required. In addition, an assumption that the observed data are representative of the true conditions of the geologic processes is made by geostatistical modeling techniques.

Geologic data with non-stationary features are considered as locally stationary with a varying mean over the interest. Such geologic data could be dissociated into two components. The trend component can be represented as a gradual change over the domain(Watson, 1972; Journel and Huijbregts, 1978; Myers, 1989; Deutsch and Journel, 1998; Lloyd, 2011). This assumption of a trend model is often made in many fields. The residual component exhibits a random variation around the local mean. Some particular assumptions are also made about the structure of the residual field: (1) the residuals are uncorrelated; and (2) the average value of the residuals over the entire set of the observations is zero.

The third assumption in trend modeling is that the trend can be reasonably approximated by a smooth continuous model. Such smooth model can be calculated by a simple mathematical expression. Additionally, the trend is assumed to represent the large-scale variability beyond the length scales of the data spacing. The short-scale features trends will not significantly affect the large-scale trend modes, see Figure 1.3.

In this thesis, the trend is modeled deterministically and assumed to have no uncertainty. In addition, besides the description of the gradually changing mean, the real trend could abrupt changes at specific locations within the domain. Figure 1.4 shows a 1-D schematic illustration of possible trends that a general increasing value $z(\boldsymbol{u})$ with a coordinate $\boldsymbol{u}$. The trend can be seen as a

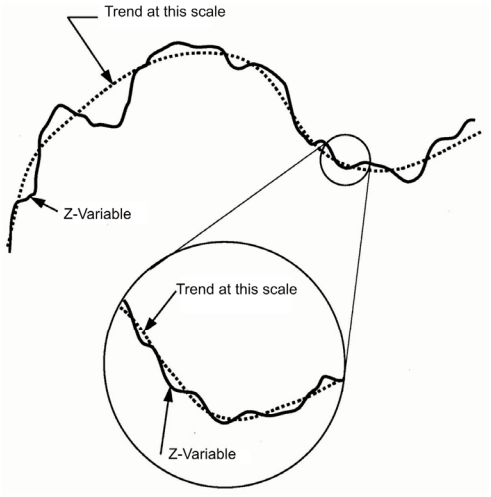


Figure 1.3: Schematic illustration of the decision on the scale of the trend (Deutsch, 2010a)

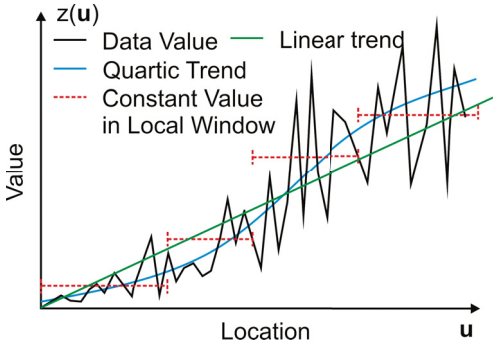


Figure 1.4: Schematic illustration of possible trends for a general increasing attribute $z(\boldsymbol{u})$ with the coordinate $\boldsymbol{u}$ (Journel and Rossi, 1989)

single function plus the fluctuations (a red line or a blue line) or by a constant value plus fluctuations in each segment (red dash lines). In this research, these sudden changes, such as the red-dash line in Figure 1.4, are not considered.

1.4 Dissertation Outline

The goal for this dissertation is to develop a practical algorithm accounting for non-stationary features in continuous regionalized variables. The outline of the dissertation is as follows.

Chapter 2 begins with a brief review on the basic concepts of the essentials of geostatistics. Most geostatistical modeling techniques are addressed with the decision of stationarity; however, they are inappropriate to be considered when non-stationary features from the regionalized variable exist. Several current geostatistical methods regarding non-stationary features are then presented. The limitations of these methods are discussed.

Chapter 3 is the core chapter on trend modeling. The goal is to develop a reasonable function to mathematically describe the trend-like features in the geological phenomenon. This chapter also discusses how to choose these trend parameters. The choices of trend parameters in practice come from a synthetic example. After that, an objective function is determined so that the resulting trend model best represents the large-scale observed data. The novel weighting function for trend extrapolation is proposed and applied with a small example at the end of the chapter.

Chapter 4 addresses geostatistical modeling in the presence of a trend. The stepwise conditional transformation and the expectation maximization algorithm for Gaussian mixtures are reviewed and implemented with an example in the presence of location-dependent features. The results show the benefits with the incorporation of Gaussian mixtures. Additional information from the trend is also discussed and would be accounted for in Gaussian mixture models. The consideration of the incorporation of the missing trend values is introduced so that all trend features would be considered.

Chapters 5 and 6 illustrate the proposed methods with actual examples

and show the importance of considering the trend features into geostatistical predictions. The demonstrations are discussed using case studies with: (1) a summary of the choices of the trend parameters; (2) a review of the conventional geostatistical prediction; and (3) the implementation of the proposed technique combining the trend model into the geostatistical modeling. The results show the improved performance by the proposed method.

Chapter 7 wraps up the thesis with conclusions and future work. The benefits and limitations of using the proposed approaches in trend modeling and modeling with a trend are evaluated. Future developments in trend extrapolation and other implementations in details are contemplated.

Appendix A contains the derivatives of the weighting function for a 1-D trend extrapolation. Appendix B provides the descriptions of the Fortran coded programs related to the trend modeling and modeling with a trend. Methodology that is developed in this thesis has been implemented with the stand-alone programs. The programs use text-based data files and parameters. The results in this thesis could be repeated.

Chapter 2

Theoretical Background

This chapter briefly outlines a literature review of the material that is relevant to standard geostatistical modeling. Numerous geostatistical modeling algorithms have been developed and implemented. Care should be taken to choose the appropriate geostatistical algorithm that reasonably describes the geological property being modeled.

2.1 Essentials of Geostatistics

The fundamental decision in geostatistical analysis is to model the geologic variables within stationary domains. The concept of stationarity is reviewed with the notation that will be used in the thesis.

2.1.1 Random Function Model

Consider an attribute z in the domain A where z is measured on a constant volume or mass support. The actual value at a location is considered as a Regionalized Variable (ReV) that may be One-, Two- or Three-dimensional (1-D, 2-D or 3-D) in real space. A certain measurement, for example, a copper grade value at location $\boldsymbol{u}$, is characterized by a ReV that is denoted with a lower

case $z(\boldsymbol{u})$ (Wackernagel, 2003; Oliver and Webster, 2015). A random mechanism viewpoint could be considered for a ReV. A measured value $z(\boldsymbol{u})$ represents one draw from a distribution. Such a variable at location of $\boldsymbol{u}$ is called a Random Variable (RV) which is denoted with an upper case $Z(\boldsymbol{u})$. There are many locations within the domain of interest A. Our knowledge of each continuous RV $Z(\boldsymbol{u})$ is fully characterized by a Cumulative Distribution Function (CDF) (Goovaerts, 1997; Deutsch and Journel, 1998; Wackernagel, 2003):

$$F_Z(\boldsymbol{u}; z) = \text{Prob}\{Z(\boldsymbol{u}) \leq z\} \quad \forall \boldsymbol{u} \in A \tag{2.1}$$

here $F_Z(\boldsymbol{u})$ represents the CDF of a continuous RV $Z(\boldsymbol{u})$.

Multiple spatially related RVs are assembled into a Random Function (RF). Consider a set of attribute values $\{z(\boldsymbol{u}_i), i = 1, \cdots, N\}$ at a set of N locations. The multivariate CDF of the corresponding RF $\{Z(\boldsymbol{u}_i), i = 1, \cdots, N\}$ can be expressed by (Matheron, 1971; Journel, 1986; Deutsch and Journel, 1998):

$$F_{Z(\boldsymbol{u}_1),\cdots,Z(\boldsymbol{u}_N)}(z(\boldsymbol{u}_1), \cdots, z(\boldsymbol{u}_N) = \text{Prob}\{Z(\boldsymbol{u}_1) \leq z(\boldsymbol{u}_1), \cdots, Z(\boldsymbol{u}_N) \leq z(\boldsymbol{u}_N)\} \quad \forall \boldsymbol{u}_i \in A \text{ and } \quad i = 1, \cdots, N \tag{2.2}$$

here N is the number of locations in the domain A. An assumption of stationarity is made to infer this probability function at unsampled locations.

2.1.2 Stationarity

Statistical modeling requires a reasonable decision of stationarity. Stationarity is a critical decision related to the invariance of the multivariate CDF with respect to the translation within the attribute of interest (Journel and Huijbregts, 1978; Davis, 2002). Stationarity is related to a random function instead of the actual data (Isaaks and Srivastava, 1989). The decision of stationarity is made

prior to any statistical prediction. Most decisions are implicitly made with the application in a particular algorithm. Geostatistical modeling is often based on less restrictive assumptions, such as second-order stationarity and intrinsic stationarity.

Geostatistical inference of the random function is often limited to the first two moments that are the expected value and the variance/covariance functions. If the attribute of interest is divided into several sub-regions and the expected values of these sub-regions tend to be the same, this is referred to as first-order stationarity. In addition, the regionalized variable is considered as second-order stationarity if the variance/covariance between two random variables changes only with the relative distance $\boldsymbol{h}$. Considering $Z(\boldsymbol{u})$ to be a second-order stationary random function, we can summarize:

$$\begin{aligned} E\{Z(\boldsymbol{u})\} &= m \\ E\left\{(Z(\boldsymbol{u}) - m)^2\right\} &= C(0) = \sigma^2 \\ E\{(Z(\boldsymbol{u}) \cdot Z(\boldsymbol{u}+\boldsymbol{h}))\} - m^2 &= C(\boldsymbol{u}, \boldsymbol{u}+\boldsymbol{h}) = C(\boldsymbol{h}) \\ &\forall \boldsymbol{u}, \boldsymbol{h}, \boldsymbol{u}+\boldsymbol{h} \in A \end{aligned} \tag{2.3}$$

here m, σ^2 and $C(\boldsymbol{h})$ represent the mean, the variance and the covariance of the variable, respectively. The properties of the mean and the variance/covariance do not depend on location within the domain A. Stationarity of the first two moments is also called weak stationarity.

If the first two moments of the difference between pairs of regionalized values are location invariant, this is referred to as intrinsic stationarity that leads to the notion of the variogram. Intrinsic stationarity only requires increments of the random function $Z(\boldsymbol{u})$ to be stationary. Suppose a pair of points at locations of $\boldsymbol{u}$ and $\boldsymbol{u}+\boldsymbol{h}$, and such pairs of observations are separated by a lag vector $\boldsymbol{h}$. The variance of the increments is given as (Matheron, 1963; Journel and Huijbregts,

1978; Isaaks and Srivastava, 1989; Goovaerts, 1997; Wackernagel, 2003):

$$\begin{aligned} &\text{Var}\left\{(Z(\boldsymbol{u}) - Z(\boldsymbol{u}+\boldsymbol{h}))\right\} = 2\gamma(\boldsymbol{h}) \\ &\Rightarrow \gamma(\boldsymbol{h}) = \frac{1}{2}E\left\{(Z(\boldsymbol{u}) - Z(\boldsymbol{u}+\boldsymbol{h}))^2\right\} \\ &\qquad\qquad \forall \boldsymbol{u}, \boldsymbol{h}, \boldsymbol{u}+\boldsymbol{h} \in A \end{aligned} \tag{2.4}$$

The variogram $\gamma(\boldsymbol{h})$ (without the 2, this is also called the semi-variogram) is defined based on the intrinsic hypothesis that is a half of the expected squared difference between paired observations of the random function. Matheron (1971) proposed the classical experimental variogram calculation that can be estimated from the paired observed values $\{z(\boldsymbol{u}_i), z(\boldsymbol{u}_i+\boldsymbol{h})\}$, that is,

$$\gamma(\boldsymbol{h}) = \frac{1}{2 \cdot n(\boldsymbol{h})} \sum_{i=1}^{n(\boldsymbol{h})} [z(\boldsymbol{u}_i) - z(\boldsymbol{u}_i+\boldsymbol{h})]^2 \quad \forall \boldsymbol{u}_i, \boldsymbol{h}, \boldsymbol{u}_i+\boldsymbol{h} \in A \tag{2.5}$$

here $n(\boldsymbol{h})$ represents the number of paired samples with a distance vector approximately $\boldsymbol{h}$. The variogram is modeled from the available data. The variogram relates to the covariance $C(\boldsymbol{h})$, that is (Journel, 1974; Rossi and Deutsch, 2014),

$$\gamma(\boldsymbol{h}) = \sigma^2 - C(\boldsymbol{h}) \quad \forall \boldsymbol{h} \in A \tag{2.6}$$

The model fitted to the variogram/covariance is used for estimation or simulation to characterize the spatial variability of the regionalized variable.

If all higher moments only depend on the relative distance between locations, the random function has the property of strong stationarity. However, such strong stationarity in natural phenomena rarely exists. The decision of stationarity allows the spatial invariant distribution of data within the domain (Myers, 1989; Goovaerts, 1997; Deutsch and Journel, 1998; Davis, 2002; Wackernagel, 2003; Pyrcz and Deutsch, 2014).

2.2 Geostatistical Techniques for Stationary Prediction

In a mineral deposit, a numerical model is needed to quantify rock properties at all locations. Geostatistical techniques provide tools to predict the values at all unsampled locations. There are two standard geostatistical techniques for the spatial prediction: estimation and simulation. A decision of weak stationarity must be made for these techniques to be applied.

2.2.1 Estimation

Kriging is a Best Linear Unbiased Estimation (BLUE) method in spatial statistics. The value at an unsampled location is estimated based on the structural characteristics of the observed data which are summarized by the variogram/-covariance model (Journel and Huijbregts, 1978; Goovaerts, 2000; Chiles and Delfiner, 2012). The locations with known values are relative to the locations being estimated. The variogram/covariance captures the spatial dependence and accounts for the spatial configuration between samples.

Simple Kriging (SK) provides a minimized estimation error variance when the mean of the regionalized variable is constant and known. Consider n observed measurements $\{z(\boldsymbol{u}_i), i = 1, \cdots, n\}$ at locations $\{\boldsymbol{u}_i, i = 1, \cdots, n\}$ in a spatial domain A. The mean m is constant and known for the area of interest A. In this content, the mean is calculated as the average of the observed data. SK estimate at unsampled location $\boldsymbol{u}_0$ is written as (Deutsch and Journel, 1998; Wackernagel, 2003; Leuangthong et al., 2011):

$$\begin{aligned} z^*_{SK}(\boldsymbol{u}_0) - m &= \sum_{i=1}^{n} \lambda_i^{SK}(\boldsymbol{u}) \cdot [z(\boldsymbol{u}_i) - m] \\ \Longrightarrow z^*_{SK}(\boldsymbol{u}_0) &= \sum_{i=1}^{n} \lambda_i^{SK}(\boldsymbol{u}) \cdot z(\boldsymbol{u}_i) + [1 - \sum_{i=1}^{n} \lambda_i^{SK}(\boldsymbol{u})] \cdot m \quad \forall \boldsymbol{u}, \boldsymbol{u}_0, \boldsymbol{u}_i \in A \end{aligned} \tag{2.7}$$

here $z^*_{SK}(\boldsymbol{u}_0)$ is the estimated value at location of $\boldsymbol{u}_0$, while $\{z(\boldsymbol{u}_i), i = 1, \cdots, n\}$ represent a set of n nearby relevant samples. $\left\{\lambda_i^{SK}(\boldsymbol{u}), i = 1, \cdots, n\right\}$ denote the kriging weights that are assigned to each of the n surrounding data $\{z(\boldsymbol{u}_i), i = 1, \cdots, n\}$. SK does not constrain the kriging weights and works with the stationary residuals from the mean, and therefore, the mean of the regionalized variable has to be known.

The kriging estimator is enforced to be unbiased, that is, $E\{Z^*_{SK}(\boldsymbol{u}_0)\} = E\{Z(\boldsymbol{u}_0)\}$ given the minimum squared error criterion. The simple kriging weights are determined based on minimizing the expected squared difference between the estimated value and the true value at an unsampled location of $\boldsymbol{u}_0$, namely, $\sigma^2_{\mathrm{SK}}(\boldsymbol{u}_0) = E\left\{(Z^*_{SK}(\boldsymbol{u}_0) - Z(\boldsymbol{u}_0))^2\right\}$. Kriging assesses the uncertainty in the estimate by the minimized estimation error variance (kriging variance). The error variance of SK can be calculated by (Deutsch and Journel, 1998):

$$\sigma^2_{SK}(\boldsymbol{u}) = \sigma^2 - \sum_{i=1}^{n} \lambda_i^{SK}(\boldsymbol{u}) \cdot C(\boldsymbol{u}_i, \boldsymbol{u}_0) \quad \forall \boldsymbol{u}, \boldsymbol{u}_0, \boldsymbol{u}_i \in A \tag{2.8}$$

σ^2 is the stationary variance of the data and $\{C(\boldsymbol{u}_i, \boldsymbol{u}_0), i = 1, \cdots, n\}$ represents the covariance between the data location $\{\boldsymbol{u}_i, i = 1, \cdots, n\}$ and the location being estimated $\boldsymbol{u}_0$. Once the variogram model has been established from the observed data, the covariance matrices $\{C(\boldsymbol{u}_i, \boldsymbol{u}_0), i = 1, \cdots, n\}$ are calculated and determined.

Ordinary Kriging (OK) is another kriging estimator where the mean of the data is implicitly estimated as a constant value within each search window (Deutsch and Journel, 1998). The consideration of a moving search window allows OK to balance the decision on the stationarity and the number of observations to consider (Chiles and Delfiner, 2012). The OK estimate at unsampled

location $\boldsymbol{u}_0$ is given as:

$$\begin{aligned} z^*_{OK}(\boldsymbol{u}_0) - m &= \sum_{i=1}^{n} \lambda_i^{OK}(\boldsymbol{u}) \cdot [z(\boldsymbol{u}_i) - m] \\ \Longrightarrow z^*_{OK}(\boldsymbol{u}_0) &= \sum_{i=1}^{n} \lambda_i^{OK}(\boldsymbol{u}) \cdot z(\boldsymbol{u}_i) + [1 - \sum_{i=1}^{n} \lambda_i^{OK}(\boldsymbol{u})] \cdot m \quad \forall \boldsymbol{u}, \boldsymbol{u}_0, \boldsymbol{u}_i \in A \end{aligned} \tag{2.9}$$

OK constrains the sum of the kriging weights so that the mean m in Equation 2.9 is not used in the estimate. The kriging weights of OK are constrained to be unity:

$$1 - \sum_{i=1}^{n} \lambda_i^{OK}(\boldsymbol{u}) = 0 \quad \forall \boldsymbol{u} \in A \tag{2.10}$$

The unknown mean m does not influence the estimation. The error variance of OK estimate is calculated by (Deutsch and Journel, 1998):

$$\sigma^2_{OK} = \sigma^2 - \mu^{OK}(\boldsymbol{u}) - \sum_{i=1}^{n} \lambda_i^{OK}(\boldsymbol{u}) \cdot C(\boldsymbol{u}_i, \boldsymbol{u}_0) \quad \forall \boldsymbol{u}, \boldsymbol{u}_i \in A \tag{2.11}$$

here $\mu^{OK}(\boldsymbol{u})$ is the Lagrange parameter associated with the constraint in Equation 2.10. The minimization of OK estimation variance with the constraint on the weights requires a stationary spatial dependence structure in each local search neighborhood.

Estimation methods provide unique and smooth models. The conditioning data are reproduced; however, the estimation does not produce the variability of the regionalized variable and does not assess the uncertainty.

2.2.2 Simulation

Geostatistical conditional simulation has become popular to quantify uncertainty at multiple locations simultaneously. Simulation provides an expected value at each location, as well as alternative realizations to quantify the variability and uncertainty of the regionalized variable. The generated realizations depict the

statistical characteristics of the observed data and unsampled locations (Rossi and Deutsch, 2014). The scheme of Monte Carlo Simulation (MCS) is utilized in simulation methods (Metropolis and Ulam, 1949).

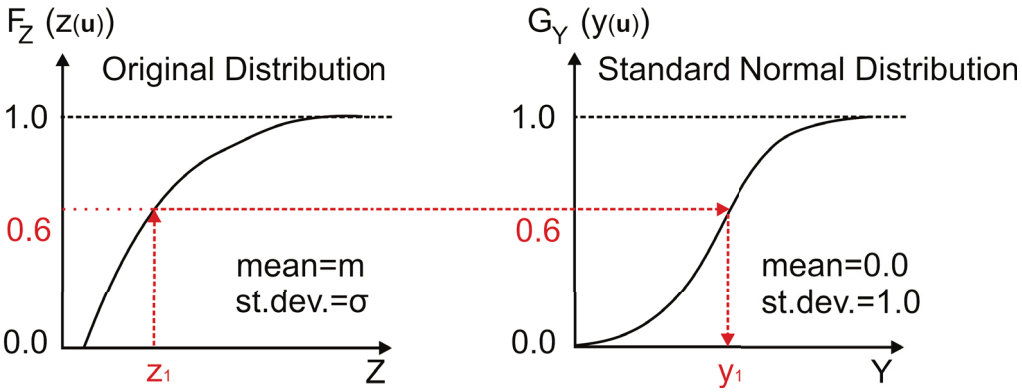


Figure 2.1: Schematic illustration of the NS transformation

Simulation of continuous variables is almost always done in Gaussian units due to the simplicity and tractability of the multivariate Gaussian distribution. Consider a continuous variable z with a univariate CDF $\{F_Z(z(\boldsymbol{u}_i)), i = 1, \cdots, n\}$ that represents the entire study area A. The Normal Score (NS) transformation is applied to convert the original variable Z that follows a distribution $\{F_Z(z(\boldsymbol{u}_i)), i = 1, \cdots, n\}$ with a mean of m and a standard deviation of σ to another variable Y that follows a standard normal distribution $\{G_Y(y(\boldsymbol{u}_i)), i = 1, \cdots, n\}$ with a mean of zero and a standard deviation of one. The normal distribution is a parametric distribution and is schematically illustrated in Figure 2.1 that is give as (Deutsch and Journel, 1998):

$$y(\boldsymbol{u}_i) = G_Y^{-1}(F_Z(z(\boldsymbol{u}_i))) \quad \forall \boldsymbol{u}_i \in A \text{ and } i = 1, \cdots, n \tag{2.12}$$

here $\{y(\boldsymbol{u}_i), i = 1, \cdots, n\}$ is the transformed Gaussian variable at n data locations. $G_Y^{-1}(\cdot)$ represents the inverse Gaussian CDF of the random variable Y and $F_Z(\cdot)$ indicates the CDF of the original random variable Z. This transformation

is a nonlinear quantile-to-quantile transform. The unique property of the multivariate normal distribution is that all conditional and marginal distributions are normal distributed (Deutsch, 2010a). After the normal score transformation, a simulation path is defined randomly. Gaussian realizations are back transformed to the original distribution according to (Deutsch and Journel, 1998)

$$z(\boldsymbol{u}_i) = F_Z^{-1}(G_Y(y(\boldsymbol{u}_i))) \quad \forall \boldsymbol{u}_i \in A \text{ and } i = 1, \cdots, N \tag{2.13}$$

once all simulation nodes N have been visited.

An MCS method called Sequential Gaussian Simulation (SGS) is widely used for modeling the continuous variables. SGS is conducted in a sequence of conditional distributions. Consider N dependent events $\{A_i, i = 1, \cdots, N\}$, the simulation proceeds (Deutsch and Journel, 1998; Pyrcz and Deutsch, 2014):

$$\begin{aligned}
\text{Prob}\{A_1, \cdots, A_N\} &= \text{Prob}\{A_N \mid A_1, \cdots, A_{N-1}\} \cdot \text{Prob}\{A_1, \cdots, A_{N-1}\} \\
&= \text{Prob}\{A_N \mid A_1, \cdots, A_{N-1}\} \cdot \{A_{N-1} \mid A_1, \cdots, A_{N-2}\} \cdot \text{Prob}\{A_1, \cdots, A_{N-2}\} \\
&\cdots \\
&= \text{Prob}\{A_N \mid A_1, \cdots, A_{N-1}\} \cdot \{A_{N-1} \mid A_1, \cdots, A_{N-2}\} \cdots \text{Prob}\{A_2 \mid A_1\} \cdots \text{Prob}\{A_1\}
\end{aligned} \tag{2.14}$$

Where $\{A_i, i = 1, \cdots, N\}$ symbolically represents a Gaussian value in the sequence. This algorithm is performed in Gaussian units. The central step in the simulation is to randomly draw simulated values from conditional distributions. These conditional distributions are derived from the simple kriging mean and variance that are calculated by the data values and previously simulated values using Equations 2.7 and 2.8. SGS assumes a multivariate Gaussian distribution. Each variable is transformed to be a univariate Gaussian distribution using normal score transformation. The results are restored using the associated normal

score back transformation. The application is well established and straightforward to execute (Deutsch and Journel, 1998; Chiles and Delfiner, 2012; Rossi and Deutsch, 2014).

Turning Bands Simulation (TBS) originated by Matheron (1973) and developed by Journel (1974) provides 3-D unconditional realizations. The data are transformed into Gaussian units to ensure the reproduced statistics. The first step in TBS algorithm is to generate unconditional simulated values $\left\{y^l_{UC}(\boldsymbol{u}), l = 1, \cdots, L\right\}$ with the total number of realizations L based on the covariance models from the regionalized variables. In this case, the data are not reproduced in realizations. The conditional kriging is then carried out to ensure the data are reproduced, namely, $\{y_{KC}(\boldsymbol{u})\}$. The kriging with the unconditional simulated values $\left\{y^l_{KU}(\boldsymbol{u}), l = 1, \cdots, L\right\}$ at conditioning data locations is then conducted. The final conditional simulated values $\left\{y^l_{CS}(\boldsymbol{u}), l = 1, \cdots, L\right\}$ are calculated by the unconditional simulated values and the difference between the kriged values that are given as (Rossi and Deutsch, 2014):

$$y^l_{CS}(\boldsymbol{u}) = y^l_{UC}(\boldsymbol{u}) + (y_{KC}(\boldsymbol{u}) - y^l_{KU}(\boldsymbol{u})) \quad \forall \boldsymbol{u} \in A \text{ and } l = 1, \cdots, L \tag{2.15}$$

This algorithm is fast; however, some artifacts may be shown in realizations due to the limitation of the partitioned lines in the 3-D space (Deutsch and Journel, 1998; Ren et al., 2004; Ren, 2005).

Lower - Upper (LU) triangular matrix simulation is often considered in the presence of few conditioning data, small grid nodes and high demanding in the number of realizations (Goovaerts, 1997; Deutsch and Journel, 1998; Rossi and Deutsch, 2014). The large and positive definite covariance matrix $\boldsymbol{C}$ is solved by Choleski LU decomposition (Deutsch and Journel, 1998; Villalba and Deutsch,

2011):

$$\begin{aligned} \boldsymbol{C} &= \begin{bmatrix} \left[C_Y(\boldsymbol{u}_i, \boldsymbol{u}_j)\right]_{n\cdot n} & \left[C_Y(\boldsymbol{u}_i, \boldsymbol{u}_\beta^{'})\right]_{n\cdot N} \\ \left[C_Y(\boldsymbol{u}_\alpha^{'}, \boldsymbol{u}_j)\right]_{N\cdot n} & \left[C_Y(\boldsymbol{u}_\alpha^{'}, \boldsymbol{u}_\beta^{'})\right]_{N\cdot N} \end{bmatrix} \\ &= \boldsymbol{L}\cdot\boldsymbol{U} \qquad \forall \boldsymbol{u}_i, \boldsymbol{u}_j, \boldsymbol{u}_\alpha^{'}, \boldsymbol{u}_\beta^{'} \in A \\ &\qquad\qquad i, j = 1, \cdots, n \text{ and } \alpha, \beta = 1, \cdots, N \end{aligned} \tag{2.16}$$

here the subscripts $\{\boldsymbol{u}_i, \boldsymbol{u}_j, i/j = 1, \cdots, n\}$ represent the conditioning data locations and $\left\{\boldsymbol{u}_\alpha^{'}, \boldsymbol{u}_\beta^{'}, \alpha/\beta = 1, \cdots, N\right\}$ represent the simulated grid nodes. $\boldsymbol{L}$ represents the lower triangular matrix, while $\boldsymbol{U}$ represents the upper triangular matrix. The realization $\left\{\boldsymbol{y}^{(l)}, l = 1, \cdots, L\right\}$ in the conditional simulation is calculated by the lower triangular matrix $\boldsymbol{L}$ and a normally distributed random matrix $\boldsymbol{w}^{(l)}$ (Deutsch and Journel, 1998). More realizations could be obtained by drawing a new normal random matrix $\boldsymbol{w}$. The LU formalism provides a fast solution once the decomposition is complete. All grid nodes and data locations are considered simultaneously in a single covariance matrix in LU decomposition algorithm. The drawback of such approach is memory requirements due to a potentially large covariance in Choleski LU decomposition. This method is not suitable for a large amount of conditioning data or for a large number of simulated locations.

The simulation method reproduces the conditioning data, the histogram and the spatial variability. Multiple realizations provide an assessment of the uncertainty. The simulation algorithm is popular because of straightforward implementation and robustness; however, all simulation algorithms make a strong assumption of stationarity. The local accuracy of the predicted uncertainty may be unreliable if this assumption is not satisfied.

2.3 Trend Description, Importance and Detection

Most mineral deposits have high valued locations and lower grade locations, see Figures 1.1a and 1.1b. These examples suggest some deterministic variation to the spatial features. Such features violate the assumption of stationarity and the application of geostatistical modeling is no longer straightforward.

2.3.1 Trend Description

The trend is understood as large-scale spatial varying features of the property (Watson, 1972; Journel and Huijbregts, 1978; Myers, 1989; Deutsch and Journel, 1998; Lloyd, 2011). One common approach of representing a regionalized variable in the presence of a trend is to consider two components, namely a deterministic trend and a random fluctuation (Krumbein, 1959; Watson, 1972; Delfiner, 1976), that is,

An observed geologic data value =

A smoothly varying deterministic trend

+ A rapidly varying stochastic fluctuation

In geological terms, the deterministic trend represents the regional behavior of the geological phenomenon at a large-scale, whereas the stochastic fluctuation reflects local changes at a short-scale. In common geostatistical notation (Delfiner, 1976; Journel and Rossi, 1989; Wackernagel, 2003):

$$Z(\boldsymbol{u}) = \underbrace{m(\boldsymbol{u})}_{\text{Deterministic trend}} + \underbrace{R(\boldsymbol{u})}_{\text{Stochastic fluctuation}} \qquad \forall \boldsymbol{u} \in A \tag{2.17}$$

here $Z(\boldsymbol{u})$ is a random variable under consideration. The random variable $R(\boldsymbol{u})$ is a spatially correlated fluctuation plus some actually random differences where $E\{R(\boldsymbol{u})\} = 0$. The fluctuation is often understood as a stationary regionalized

variable with some spatial correlation. $m(\boldsymbol{u})$ is a deterministic variable and it is considered to be a predicable locally varying mean. The trend values are often assumed uncorrelated with the fluctuations, that is, $\rho\{m(\boldsymbol{u}), R(\boldsymbol{u})\} \cong 0$. The trend values equal the expected value of $Z(\boldsymbol{u})$ at locations $m(\boldsymbol{u}) = E\{Z(\boldsymbol{u})\}$.

2.3.2 Trend Importance

Rivoirard (1987) discussed the importance of the trend which could be understood by the kriging weights applied to the local mean. Consider the trend values $\{m(\boldsymbol{u}_i), i = 0, 1, \cdots, n\}$ are known throughout the domain A. Assuming second-order stationarity and kriging equation is written as:

$$z^*(\boldsymbol{u}_0) - m(\boldsymbol{u}_0) = \sum_{i=1}^{n} \lambda_i(\boldsymbol{u}) \cdot [z(\boldsymbol{u}_i) - m(\boldsymbol{u}_i)] \quad \forall \boldsymbol{u}, \boldsymbol{u}_0, \boldsymbol{u}_i \in A \tag{2.18}$$

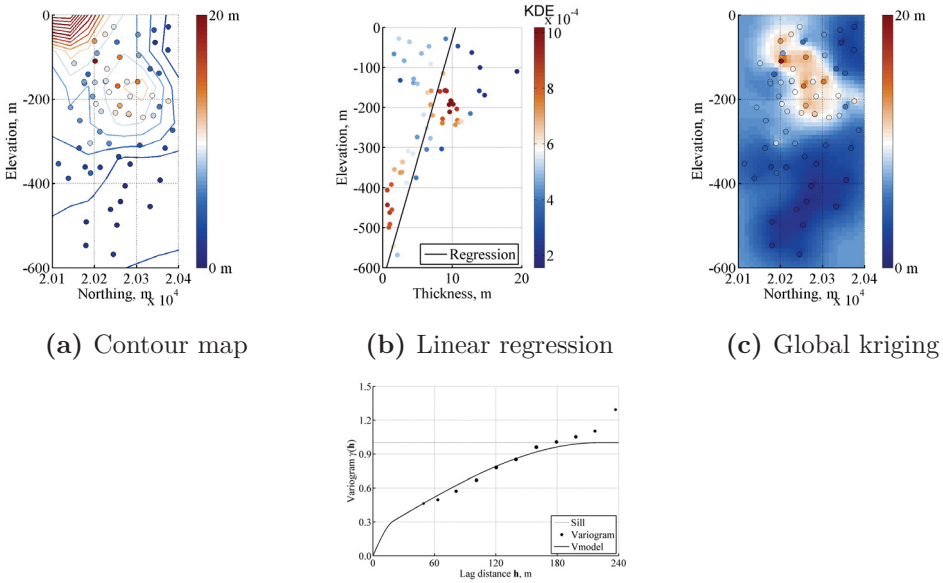
here $z^*(\boldsymbol{u}_0)$ is an estimated value, while $\{z(\boldsymbol{u}_i), i = 1, \cdots, n\}$ are the surrounding data values. $\{\lambda_i(\boldsymbol{u}), i = 1, \cdots, n\}$ denote the kriging weights assigned to each data Let us suppose the trend is location-independent that means $\{m(\boldsymbol{u}_i) = m(\boldsymbol{u}_0), i = 0, 1, \cdots, n\}$. Then, the estimate could be re-written as:

$$z^*(\boldsymbol{u}_0) = \underbrace{[1 - \sum_{i=1}^{n} \lambda_i(\boldsymbol{u})]}_{\text{weights given to trend}} \cdot m(\boldsymbol{u}_0) + \sum_{i=1}^{n} \lambda_i(\boldsymbol{u}) \cdot z(\boldsymbol{u}_i) \quad \forall \boldsymbol{u}, \boldsymbol{u}_0, \boldsymbol{u}_i \in A \tag{2.19}$$

here $\{1 - \sum_{i=1}^{n} \lambda_i(\boldsymbol{u})\}$ represents the weight given to the mean or the spatial trend. This is an indicator of how the trend affects the estimation. The importance of the trend from Equation 2.19 is understood as $|1 - \sum_{i=1}^{n} \lambda_i(\boldsymbol{u})|$. The less weight given to the data, the more important the spatial trend.

2.3.3 Trend Detection

The large-scale trend features in the data must be detected before geostatistical modeling. Some statistical tools to help with this include a location map, a scatter plot applying regression analysis (Davis, 2002; Chatfield, 2004), a neutral model of the regionalized variable (Deutsch and Journel, 1998) and the variogram (Leuangthong, 2003). The practitioner should decide if there is a trend based on large-scale average values of the regionalized variable.



(a) Contour map **(b)** Linear regression **(c)** Global kriging

(d) Variogram

Figure 2.2: Common methods for trend detection

The most obvious way of identifying the local variations is simply to map the data values with a standard automatic contouring algorithm. This approach provides the first hint of large-scale trends. The location map of thickness with contour lines is shown in Figure 2.2a. High values are shown in the North and the values are reduced to the South.

Regression analysis could be considered to describe the regularity of the change. Correlation between the data and the coordinates are summarized by a regression function providing a measure of the trend. Figure 2.2b shows the location map of the reducing thickness values with the depth. The relationship between the data values and the depth is fitted by a linear regression model showing a reducing value with increasing depth. The regression analysis is used for confirming the existence of a trend.

A neutral model analysis provides an indication of the relative importance of large- and short-scale variabilities. Kriging strongly depends on the decision of stationarity; however, it can still be used for mapping large-scale trends in some cases. Almost all kriged maps show some artifacts. These artifacts exist due to the limited search neighborhood typically used in kriging. Global kriging uses all data values for estimation at each location and the kriged map smoothly varies without artifacts. Figure 2.2c shows a kriged map with global kriging where high values are concentrated in the North.

The spatial trends may also be revealed from the variogram model. The experimental variogram is calculated by Equation 2.5 and it may show trend features in one or more directions. The trend can be identified when the variogram increases above the expected sill (the data variance) for large lag distance $\boldsymbol{h}$ in some directions, see Figure 2.2d.

2.4 Techniques for Trend Modeling

The aim of modeling a trend is to understand the large-scale regionalized variable and to remove the trend properly in order to satisfy the stationarity assumption in geostatistics.

2.4.1 Trend Modeling Techniques

The spatial trend must be inferred from the data or other geological information (McLennan, 2007). Over the years, trend modeling algorithms have been well developed and many different trend methods are found in the literature. Hand mapping and computer mapping are two commonly used methods to construct a trend (McLennan et al., 2005; Pyrcz and Deutsch, 2014). It is hard to decide which method would provide the best trend model. The decision for trend modeling function should be made by users.

A traditional approach for trend modeling is hand mapping. The geologist delineates the trend model based on the available data and their understandings. Contour mapping is one of conventional hand-made trend modeling methods and could be applied on many different types of maps such as rock attribute maps (Krumbein, 1959). This process is a flexible human interpretation technique; however, it is time-consuming and it could only be applied in fairly simple cases (Krumbein, 1959; Pyrcz and Deutsch, 2014).

Some computer mappings could be applied. Polynomial fitting is a multiple linear regression function that is a widely used method for fitting a trend. The mathematical and statistical theories associated with polynomial regressions are represented by a number of authors (Miller, 1956; Krumbein, 1959; Journel and Rossi, 1989; Goovaerts, 1997; Howarth, 2001; Rossi and Deutsch, 2014). Values at unsampled locations are computed by the polynomial function (Rossi and Deutsch, 2014). One common form of the polynomial functions for trend modeling is given as (Journel and Rossi, 1989; Goovaerts, 1997):

$$m^*(\boldsymbol{u}) = \sum_{l=0}^{L} a_l(\boldsymbol{u}) \cdot f_l(\boldsymbol{u}) \quad \forall \boldsymbol{u} \in A \tag{2.20}$$

where $\{f_l(\boldsymbol{u}), l = 0, \cdots, L\}$ are functions of the data coordinates $\boldsymbol{u}$, by convention, $f_0(\boldsymbol{u}) = 1$. $\{a_l(\boldsymbol{u}), l = 0, \cdots, L\}$ are unknown parameters and

$\{l, l = 0, \cdots, L\}$ represent the number of functions. The advantages of the polynomial function are flexibility and computational savings (Grant, 1957). However, the polynomial methods are not generally recommended in trend modeling due to unstable extrapolation properties (Stoer and Bulirsch, 2002).

Gonzalez et al. (2006) and Hong and Deutsch (2009) discussed the idea of constructing a trend with lower order trends. A full three-dimensional trend $m^*(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ is calculated from a 2-D areal trend $m^*(\boldsymbol{x}, \boldsymbol{y})$ and a 1-D vertical trend $m(\boldsymbol{z})$:

$$m^*(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = \frac{m^*(\boldsymbol{z}) \cdot m^*(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})}{m_{\text{gobal}}} \quad \forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in A \tag{2.21}$$

here m_{gobal} is the global mean from the distribution. This equation scales the vertical trend by the areal trend. The disadvantage of this method is that the trend in areal and vertical directions must be independent which is unrealistic in real geological phenomena (Deutsch, 2010a).

Kriging is a straightforward, effective and robust trend modeling approach (Journel and Huijbregts, 1978; Deutsch and Journel, 1998). Kriging assumes stationarity over the domain, but it still provides an estimate that may have the target properties of a trend. The kriged trend model may not be smooth with a low nugget effect. Kriging can be used to model a trend but it is relatively computationally demanding. Additionally, kriging is an exact estimation and block kriging must be used to avoid the exact reproduction of the data values.

Moving window averages are straightforward for fitting a set of observed values. This method does not make strong assumptions about the statistical properties of the data. Moving window averages are usually computed on a regular grid that will be used for the subsequent geostatistical modeling. Krige (1976) introduced moving averages to avoid systematic overestimation of reserves in the field of mining. The benefit of this function is that all estimated values are bounded by the minimum and maximum data values and no extreme values are

interpolated (Crain and Bhattacharyya, 1967). The difficulty of this approach exists due to many mathematical functions available to define a weighting function. The general prediction formula is given as:

$$m^*(\boldsymbol{u}) = \sum_{i=1}^{n} w(\boldsymbol{u}) \cdot z(\boldsymbol{u}_i) \quad \forall \boldsymbol{u}, \boldsymbol{u}_i \in A \tag{2.22}$$

here $m^*(\boldsymbol{u})$ is the estimated trend value, $\{z(\boldsymbol{u}_i), i = 1, \cdots, n\}$ are the data values and $w(\boldsymbol{u})$ denote the moving average weights assigned to the data in the domain A. The trend models depend on the choice of the weighting function. The trend would be modeled differently with nearby data according to different weighting schemes.

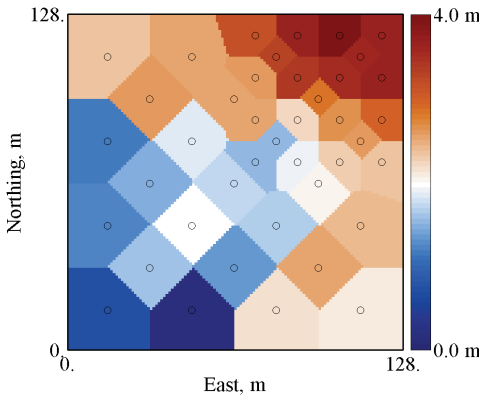


Figure 2.3: Modeled trend using Voronoi algorithm

The nearest neighbor is the simplest method that assigns known data values at unsampled locations with the same values from the nearest data based on polygons. The boundary of each polygon could be determined by Voronoi or Thiessen algorithm. The known values $\{z(\boldsymbol{u}_i), i = 1, \cdots, n\}$ are located in the centers of the polygons $\{V(\boldsymbol{u}_i), i = 1, \cdots, N\}$. All gridded cells within the same polygon are assigned with the same value (Webster and Oliver, 2008; Lloyd,

2011). Figure 2.3 shows the modeled trend with Voronoi algorithm and polygons are dyed with data values. The generated map produces abrupt jumps in the surface at the polygon edges. Nearest neighbor methods are not suitable for trend modeling. Alternative polygonal vector schemes, such as Delaunay triangulation (Burrough et al., 2015) and natural neighbors (Tily and Brace, 2006; Webster and Oliver, 2008), are also implemented. However, this method only produces a smooth trend model within the boundary of the available data and cannot extrapolate outside the convex hull of the data locations. The trend values beyond the data locations are invalid.

Moving averages with a distance-related weighting function are naturally adapted to the geological variables. More weights to the nearby data and assign less weights to more distant data. The spatial importance of geologic data rarely changes linearly (Brunsdon et al., 1998). A linearly decreasing weighting function that ignores the complex spatial relationships between geologic data is not considered. Some non-linear weighting functions are applied for trend modeling.

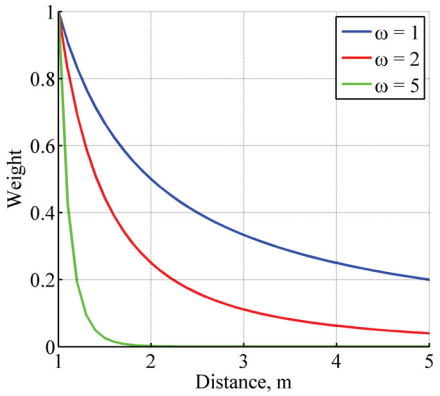
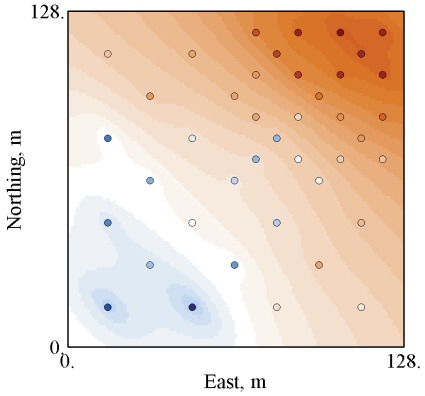
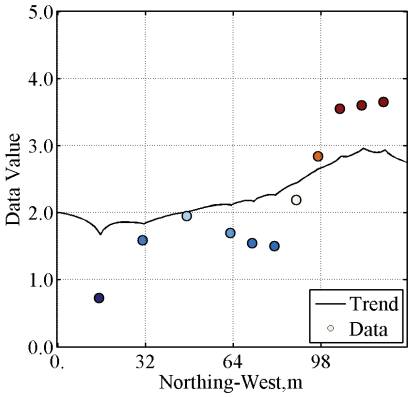


Figure 2.4: IDW weights with different exponents ω

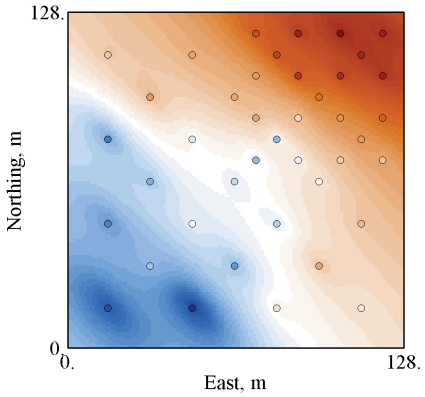
Moving averages algorithm with an Inverse Distance Weighting (IDW) function is a simple spatial prediction technique that considers the similarity



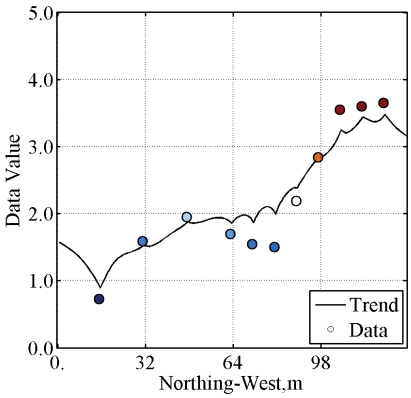
(a) Trend model with $\omega = 1$



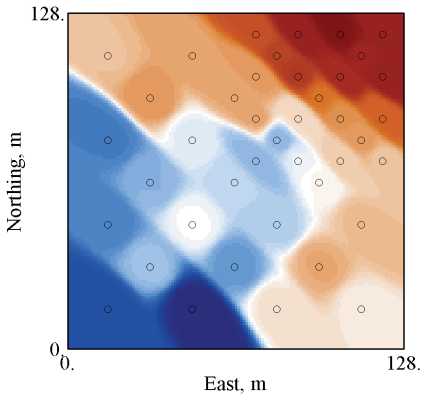
(b) The diagonal section of the trend model with $\omega = 1$



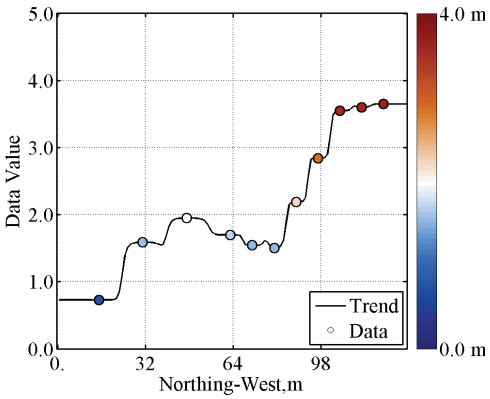
(c) Trend model with $\omega = 2$



(d) The diagonal section of the trend model with $\omega = 2$



(e) Trend model with $\omega = 10$

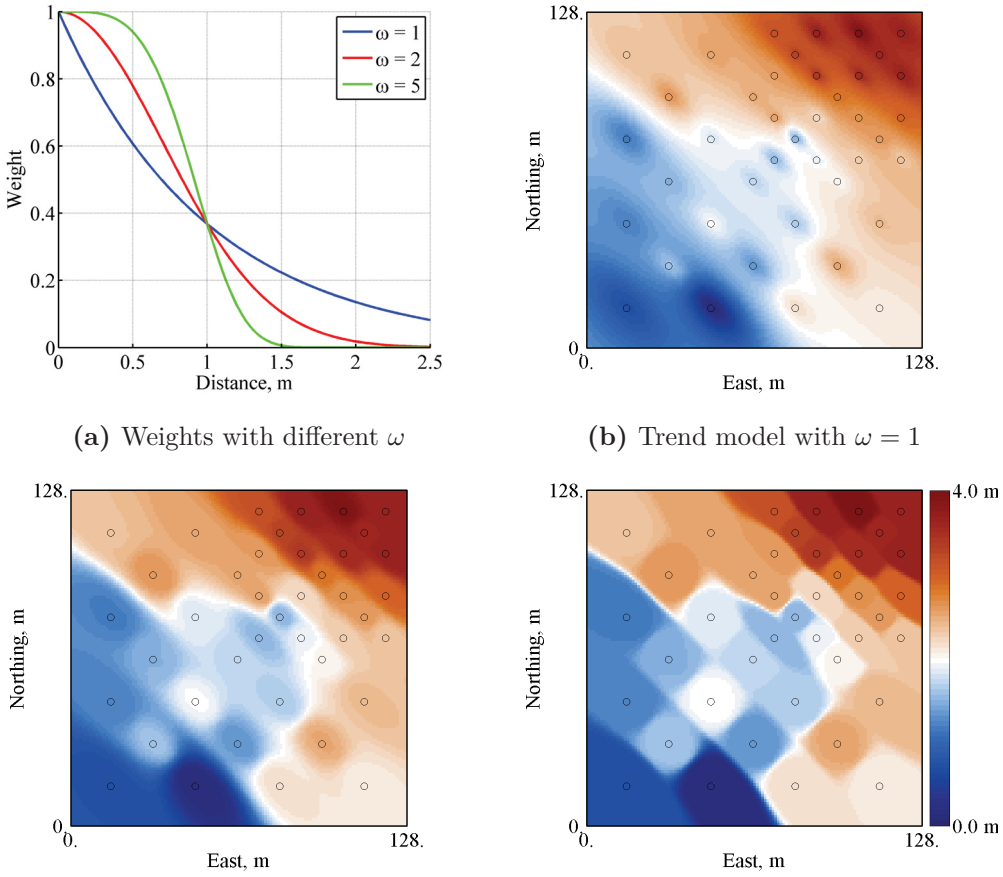


(f) The diagonal section of the trend model with $\omega = 10$

Figure 2.5: Modeled trend using moving averages with an IDW function

between data (Bartier and Keller, 1996). Figure 2.4 shows the weighting schemes with different exponents ω and the weights are scaled to unity. The estimated location receives more weight than those further away. IDW weighting allows a non-linear calculation and controls weights by the exponent and the distance. Figure 2.5 shows the trend models with different exponent weights. The trend model using lower exponent weights (no more than 2) tends to make a smooth surface with gradual gradients but obvious changes near data locations, while the model with higher exponent weights (greater than 2) produces a relatively flat trend surface near data locations but very steep changes between data locations (Shepard, 1968; Isaaks and Srivastava, 1989).A satisfactory empirical exponent is 2 for trend surface mapping, see Figures 2.5c and 2.5d, although the trend surface is not smooth near data locations. The artifacts can be found due to large weights near data locations.

Most mineral grades show a nearly positively skewed distribution (Krumbein, 1937; Agterberg, 1970; Sadler, 1981). Moving window averages algorithm with an exponential weighting function becomes popular (Brunsdon et al., 2002; Fotheringham et al., 2002). Figure 2.6a shows the spatial exponential kernels with different exponents ω. The spatial kernel with a small exponent produces a rough trend model, see Figure 2.6b with an exponent of 1; the spatial kernel with a large exponent produces a trend model with patchy artifacts, see Figure 2.6d with an exponent of 5. A reasonable exponent of 2 is a Gaussian-shape weighting function that generates smooth surfaces, see Figure 2.6c. Manchuk and Deutsch (2011) developed this method and included anisotropy in the trend modeling. The problem in this developed method is that the Gaussian weighting function must be specified and there are no objective manners of determining the correct trend parameters.



(a) Weights with different ω

(b) Trend model with $\omega = 1$

(c) Trend model with $\omega = 2$

(d) Trend model with $\omega = 5$

Figure 2.6: Modeled trend using moving averages with different exponential functions

2.4.2 Best Fit of Trend Modeling

It is necessary to consider a measure of the goodness of a trend model and to determine if the trend component is the statistically significant. Some statistical functions could assist in determining the optimum amount of a trend.

The theory of the least squares method is developed by many authors (Mandelbaum, 1963; Harbaugh and Merriam, 1968; Wren, 1973; Opsomer et al., 2001; Kariya and Kurata, 2004). This approach could be applied in the objective

description of a trend model. The goal of the least squared method is to choose the trend values that minimize a measure of samples. The best trend model is measured by residuals that are the offsets between the data values and the trend values. The trend values could be obtained by minimizing the sum of the squared differences between the actual values $\{z(\boldsymbol{u}_i), i = 1, \cdots, n\}$ and the predicted trend values $\{m^*(\boldsymbol{u}_i), i = 1, \cdots, n\}$:

$$mse = \sum_{i=1}^{n} (z(\boldsymbol{u}_i) - m^*(\boldsymbol{u}_i))^2 \quad \forall \boldsymbol{u}_i \in A \tag{2.23}$$

The indicator mse measures the similarity between the modeled trend and the data and it quantifies the goodness of the fit of the trend. The indicator mse should be as small as possible. However, this would cause over-fitting and a discontinuous trend model.

Other statistical approaches are considered for deciding the goodness for trend modeling. McLennan (2007) proposed that the variability between the deterministic trend and the observed data should be no more than 50%, that is,

$$\frac{\sigma^2_{m(\boldsymbol{u})}}{\sigma^2_{Z(\boldsymbol{u})}} \leqslant 50\% \quad \forall \boldsymbol{u} \in A \tag{2.24}$$

however, the guideline of 50% is subjective. The covariance between the trend and the residuals at a lag distance of zero σ^2_{mR} or $C_{mR}(0)$ is a good indicator for the goodness of the fit. This assumes that the trend and the residuals should be independent. This assumption lacks verification.

2.5 Geostatistical Techniques for Non-Stationary Prediction

The process of removing the trend from the observed data is useful in the identification of geological structures. In classic geostatistics, non-stationarity geostatistics refers to locally varying statistics (Journel and Huijbregts, 1978; Goovaerts, 1997; Deutsch and Journel, 1998). A non-stationary geostatistical modeling approach may be considered more appropriate than a stationary modeling approach when trend features exist in the data. Non-stationary models have been applied in geostatistics for many years.

One method of dealing with non-stationarity is subdividing the domain into more stationary sub-regions based on the property of rock types or grade values. Compared with the local varying mean in the whole domain, the mean and variance in each sub-region are assumed constant. Several authors have proposed to divide the spatial domain into several disjoint regions by different partitioning schemes, such as hierarchical clustering of observations, assuming the covariance structure is stationary and the data are location independent within each sub-region (Kim et al., 2005; Heaton et al., 2017). These methods fail to account for the spatial correlation between neighboring locations located in different sub-regions. Meanwhile, the resulting models could show abrupt unreasonable changes at boundaries.

2.5.1 Estimation

OK is discussed in Section 2.2.1 assuming a constant but unknown local mean within each search window. In general, OK estimation is commonly used in the presence of a trend. This algorithm assumes a non-stationary RF model that contains a locally varying mean but stationary covariance (Goovaerts, 1997; Rossi and Deutsch, 2014). The constraint in Equation 2.10 implies the existence of the

non-stationary features. The estimated location-dependent mean $m^*(\boldsymbol{u})$ replaces the constant mean m in Equation 2.9, that is,

$$z^*_{OK}(\boldsymbol{u}_0) = \sum_{i=1}^{n} \lambda_i^{OK}(\boldsymbol{u}) \cdot z(\boldsymbol{u}_i) + \underbrace{[1 - \sum_{i=1}^{n} \lambda_i^{OK}(\boldsymbol{u})]}_{\text{Assumed} = 0} \cdot m^*(\boldsymbol{u}) \quad \forall \boldsymbol{u}, \boldsymbol{u}_0, \boldsymbol{u}_i \in A \quad (2.25)$$

In this case, OK does not require the local mean to be specified over the whole area of interest and the algorithm works in the non-stationary scenario.

Another commonly used technique for non-stationary estimation is to krige with the stationary residuals instead of kriging the data directly. A known and constant mean m is replaced by a smooth and estimated trend $\{m^*(\boldsymbol{u}_i), i = 1, \cdots, N\}$ at each grid cell. This algorithm is called Simple Kriging with Locally Varying Mean (SK-LVM), also known as Kriging the Residuals (KR). Consider $Z(\boldsymbol{u})$ is not a second-order stationarity random function. This non-stationary random function $Z(\boldsymbol{u})$ could be divided by a deterministic mean $m(\boldsymbol{u})$ and a stochastic random function $R(\boldsymbol{u})$ in Equation 2.17. The trend is modeled first so that the estimated trend value at each location is known, that is, $\{m^*(\boldsymbol{u}_i), i = 1, \cdots, N\}$. The estimation variance of the residual data $E\left\{(R^*_{SK-LVM}(\boldsymbol{u}_0) - R(\boldsymbol{u}_0)^2)\right\}$ at estimated location is assumed identical to that of the original data $E\left\{(Z^*_{SK-LVM}(\boldsymbol{u}_0) - Z(\boldsymbol{u}_0)^2)\right\}$. The residual values $\{r^*(\boldsymbol{u}_i) = z(\boldsymbol{u}_i) - m^*(\boldsymbol{u}_i), i = 1, \cdots, n\}$ are obtained from the data values and the prior estimated mean at sampled locations. The kriging system in SK-LVM is similar as for OK. The difference is that sum of weights should be zero. The kriging variance of the original data comes from the variance in the trend model and the kriging variance of residual data (Wackernagel, 2003).

Kriging with a Trend (KT), which is also called Universal Kriging (UK),

is a an extension of OK. The establishment of this method comes from Equation 2.17 that where data are decomposed into a deterministic trend plus correlated residuals (Armstrong, 1984). Unlike OK, the generalization of KT allows the forms of the trend model to be varying in space, such as Equation 2.20 (Matheron, 1971; Davis, 2002). The estimator of KT expands the constant trend from zero order to a polynomial of order l in Equation 2.20. The covariance of residual data rather than the covariance of the original variable should be used in KT system (Deutsch and Journel, 1998). The error variance of the KT estimate is calculated by (Armstrong, 1984; Deutsch, 2000; Chiles and Delfiner, 2012):

$$\sigma_{KT}^2 = \sigma_R^2 - \sum_{i=1}^{n} \lambda_i^{KT} \cdot C_R(\boldsymbol{u}_0, \boldsymbol{u}_i) - \sum_{l=0}^{L} \mu_l(\boldsymbol{u}) \cdot f_l(\boldsymbol{u}) \quad \forall \boldsymbol{u}, \boldsymbol{u}_0, \boldsymbol{u}_i \in A \tag{2.26}$$

here σ_R^2 is the variance of the residuals. $\{C_R(\boldsymbol{u}_0, \boldsymbol{u}_i), i = 1, \cdots, n\}$ represent the covariance matrices of the residual data between the data locations and the locations being estimated. $\{\mu_l(\boldsymbol{u}), l = 0, \cdots, L\}$ are the Lagrange multipliers. Care must be taken with the functions $\{f_l(\boldsymbol{u}), l = 0, \cdots, L\}$ in Equation 2.20. KT may not be stable because of the indeterminancy of the trend model, as well as the underlying variogram model (Olea, 1974; Armstrong, 1984).

Kriging with an external trend (KED) is a variant of KT. The secondary variable is incorporated in the estimation of the primary variable. The function of the spatial coordinates $\{f_l(\boldsymbol{u}), l = 0, \cdots, L\}$ in KT is replaced by a smooth and exhaustive secondary variable $s(\boldsymbol{u})$ into kriging. This single trend function is defined at each location and often considered from an external secondary variable and limited to two terms (Deutsch and Journel, 1998), that is,

$$E\{Z(\boldsymbol{u})\} = m(\boldsymbol{u}) = a_0 + a_1 \cdot s(\boldsymbol{u}) \quad \boldsymbol{u} \in A \tag{2.27}$$

where is $s(\boldsymbol{u})$ is assumed the spatial trends that contains a linear relationship

with the variable $z(\boldsymbol{u})$. a_0 and a_1 are the linear parameters between variables. The external variable $s(\boldsymbol{u})$ should be smooth and known at all locations. The problem with this method is that the assumption of the linear relationship between the primary variable and the secondary variable is may not be realistic and the relation between the two variables should make geological sense (Rossi and Deutsch, 2014).

Intrinsic Random Function with order k (IRF-k) offers an alternative procedure to model the non-stationary features (Matheron, 1973; Delfiner, 1976; Chiles and Delfiner, 2012). This method requires the existence of the generalized covariance functions that overcomes the possible bias from the trend dichotomy calculation in Kriging (Knotters et al., 1995). Consider that $Z(\boldsymbol{u})$ is an IRF with order k, then an isotropic function $K(\boldsymbol{h})$ exists. The estimator of the intrinsic function with order k at unsampled location $\boldsymbol{u}_0$ is given as:

$$z^*_{IRF-k}(\boldsymbol{u}_0) = \sum_{i=1}^{n} \lambda_i^{IRF-k}(\boldsymbol{u}) \cdot z(\boldsymbol{u}_i) \quad \boldsymbol{u}, \boldsymbol{u}_0, \boldsymbol{u}_i \in A \tag{2.28}$$

here $\left\{\lambda_i^{IRF-k}(\boldsymbol{u}), i = 1, \cdots, n\right\}$ are the IRF weights. The estimation error at an unsampled location $\boldsymbol{u}_0$ is given as:

$$Var\left\{Z^*_{IRF-k}(\boldsymbol{u}_0) - Z(\boldsymbol{u}_0)\right\} = \sum_{j=1}^{n}\sum_{i=1}^{n} \lambda_j^{IRF-k}(\boldsymbol{u})\lambda_i^{IRF-k}(\boldsymbol{u})K(\boldsymbol{u}_i, \boldsymbol{u}_j) \quad \forall \boldsymbol{u}, \boldsymbol{u}_0, \boldsymbol{u}_{i/j} \in A \tag{2.29}$$

here $\{K(\boldsymbol{u}_i, \boldsymbol{u}_j), i/j = 1, \cdots, n\}$, short for $K(\boldsymbol{h})$, are the generalized covariances with order k between the locations. $\left\{\lambda_j^{IRF-k}(\boldsymbol{u}), \lambda_i^{IRF-k}(\boldsymbol{u}), i/j = 1, \cdots, n\right\}$ are the estimation weights assigned to data values. The model of the generalized covariance $K(\boldsymbol{h})$ is given as the polynomial function (Matheron, 1973):

$$K(\boldsymbol{h}) = \sigma_R^2\delta(\boldsymbol{h}) + \sum_{p=0}^{k}(-1)^{p+1}b_p|\boldsymbol{h}|^{2p+1} \quad \forall \boldsymbol{h} \in A \tag{2.30}$$

here the coefficients $\{b_p, p = 1, \cdots, k\}$ satisfy $b_p \geq 0$. k represents the order. $\sigma_R^2 \delta(\boldsymbol{h})$ is the nugget effect. The intrinsic kriging variance is given as (Chiles and Delfiner, 2012):

$$\sigma^2_{IRF-k} = \sigma_R^2 - \sum_{i=1}^{n} \lambda_i(\boldsymbol{u}) K(\boldsymbol{u}_0, \boldsymbol{u}_i) - \sum_{l=0}^{L} \mu_l(\boldsymbol{u}) f_l(\boldsymbol{u}) \quad \forall \boldsymbol{u} \in A \tag{2.31}$$

here $\{\mu_l(\boldsymbol{u}), l = 0, \cdots, L\}$ are the Lagrange multipliers. The system of the IRF-k kriging variance is the same as the KT system. The kriging variance only depends on the generalized covariance. IRF theory is based on the stationarity of generalized increments that avoids the estimation of the trend model and the prediction with the spatial structures from the residual data; however, the algorithm increases the cost of statistical inference where the high-order generalized covariance may not be easily defined (Matheron, 1973; Myers, 1989; Deutsch and Journel, 1998).

2.5.2 Simulation

Non-stationary simulation is similar to non-stationary kriging assuming a trend is added to stationary residuals in Equation 2.17. Simulation with a locally varying mean assumes the local mean is informed by a secondary data. This method accounts for the trend model as a secondary data. The trend would be removed from the regionalized variable at the start. All simulation steps are consistent with conventional simulation on the residual data. The trend model is added to each of the residual realizations $\left\{r^{(l)}(\boldsymbol{u}), l = 1, \cdots, L\right\}$ to obtain multiple realizations of the regionalized variable $\left\{z^{(l)}(\boldsymbol{u}), l = 1, \cdots, L\right\}$:

$$z^{(l)}(\boldsymbol{u}) = r^{(l)}(\boldsymbol{u}) + m(\boldsymbol{u}) \quad \forall \boldsymbol{u} \in A \text{ and } l = 1, \cdots, L \tag{2.32}$$

here $\{l, l = 1, \cdots, L\}$ represent the number of realizations. Some unexpected simulated values may be encountered in this algorithm. For example, when simulating the mineral grade values in the presence of a trend, some negative simulated values are found if the simulated residuals are less than the local mean. A correction of the algorithm would be required to avoid such simulated values.

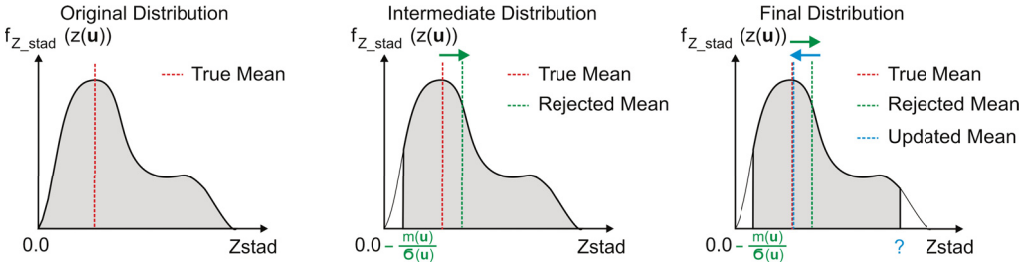


Figure 2.7: Schematic illustration of the reject sampling in simulation

Qu and Deutsch (2014) proposed to use the standardization transformation where the regionalized variable is standardized by a local mean (trend value) and a local standard deviation (trend standard deviation):

$$z_{\text{stand}}(\boldsymbol{u}) = \frac{z(\boldsymbol{u}) - m(\boldsymbol{u})}{\sigma(\boldsymbol{u})} \quad \forall \boldsymbol{u} \in A \tag{2.33}$$

This standardization transformation is simple and flexible. The assumption of stationarity is enforced in standardized units. The standardized residual values $z_{\text{stand}}(\boldsymbol{u})$ are used in the geostatistical modeling. To avoid the unexpected negative simulated values in original units, a constraint is considered as follows:

$$\begin{aligned} z(\boldsymbol{u}) &= z_{\text{stand}}(\boldsymbol{u}) \cdot \sigma(\boldsymbol{u}) + m(\boldsymbol{u}) \geq 0 \\ \Longrightarrow z_{\text{stand}}(\boldsymbol{u}) &\geq -\frac{m(\boldsymbol{u})}{\sigma(\boldsymbol{u})} \quad \forall \boldsymbol{u} \in A \end{aligned} \tag{2.34}$$

Figure 2.7 shows a schematic illustration of a Probability Distribution Function (PDF) $f_{Z\text{stand}}(z)$ of a standardized residual data Z_{stand}. The original true mean

of the standardized variable is shown with a red dash line. Simulated values less than $-\frac{m(\boldsymbol{u})}{\sigma(\boldsymbol{u})}$ should be rejected during the simulation process. The updated mean of each realization increases and is shown with a green dash line. Some high values with an arbitrary threshold in Figure 2.7 are also rejected to keep the local mean unbiased and the final local mean is shown with a blue dash line. This method is motivated by a desire to reproduce the conditioning data, and it modifies the updated local mean (blue dash line) which is very close to the true local mean (red dash line). This method could better reflect the behavior of non-stationary features; however, other features of the regionalized variable are poorly reproduced.

Machuca-Mory and Deutsch (2013) proposed a locally stationary approach that builds a location-dependent probability function under an assumption of local stationarity. The non-stationary features are captured by location-dependent distributions using a Gaussian weighting distance approach. The local spatial distributions are locally transformed to be Gaussian. However, a practical disadvantage of this approach is the computational cost.

2.5.3 Transformation

Geostatistical modeling with a decomposition of a regionalized variable in the presence a trend is not always possible. A conditional transformation could be considered in an attempt to remove the trend-like features from the regionalized variable. The complex relations in the regionalized variable should be restored after the simulation and all features are reproduced.

Projection Pursuit Multivariate Transform (PPMT) is used for transforming complex data to an uncorrelated multivariate Gaussian distribution (Barnett et al., 2014). This transformation can be applied in high dimensions and requires fewer tuning parameters. Figure 2.8 shows the core engine of the algorithm. The conditioning data are normal score transformed independently, and then,

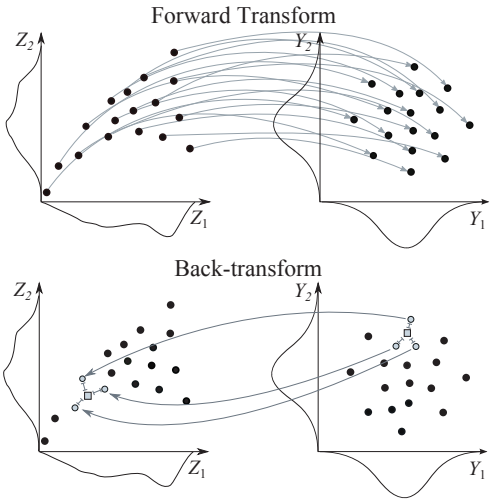
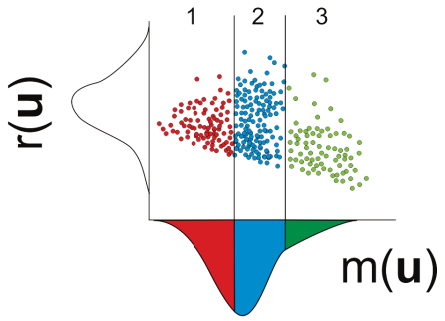


Figure 2.8: Schematic illustration of forward and back Gaussian mapping transformation associated with the PPMT (Barnett et al., 2014)

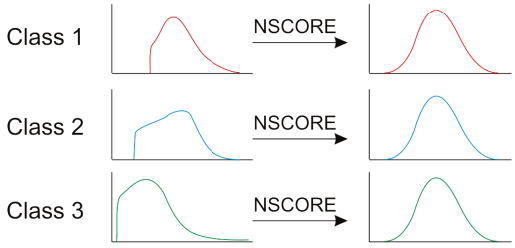
sphered forcing the marginally standard normal distributions and the orthogonal covariance matrices (Friedman, 1987). The values along projection vectors are transformed to be Gaussian. This step is iterated until a multivariate Gaussian distribution is reached. Simulation with transformed stationary data is then conducted. The relative distances between the simulated node and its nearest neighbors are persevered in the back transformation. No binning or griding issues exist. The challenge of this method, in the context of considering a trend, is that the transformation can only be applied to the equally sampled variables, yet the trend is available everywhere.

The Stepwise Conditional (SC) transformation technique was introduced by Rosenblatt (1952) and it is an extension of the normal score transformation. Leuangthong (2003) introduced this technique to geostatistics and developed the practical application (Leuangthong and Deutsch, 2004). The advantage of this transformation is that it addresses multivariate complexities, such as removing the non-linear, heteroscedastic and constraint features from geologic data. The

(a) Partition residual data, r(u), into classes conditional to normal scores of trend component, m(u).



(b) Normal score transform each class of R(**u**).



(c) Crossplot of normal score residuals, Y_R(**u**), and trend, m(**u**).

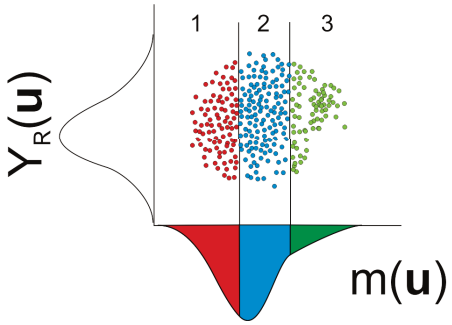


Figure 2.9: Schematic illustration of the Transformation with stepwise conditional transformation: (**a**) partition the residual data $r(\boldsymbol{u})$ into several classes based on the trend values $m(\boldsymbol{u})$; (**b**) normal score transform each residual class; and (**c**) gather all transformed residuals from all classes and plot against the trend to show bivariate distribution with approximately zero correlation and the marginal distribution of $Y_R(\boldsymbol{u})$ is in Gaussian units (Leuangthong and Deutsch, 2004)

idea is to transform the k^{th} variable conditional to the previous $k-1$ variables. The original proposed stepwise conditional transformation algorithm works on the residual data (Leuangthong, 2003). Figure 2.9 shows the schematic illustration of the transform with stepwise conditional transformation algorithm using the residuals. The residual data are normal score transformed conditional to its trend component. The corresponding residual data are conditionally transformed by the probability classes of the trend component (Leuangthong and Deutsch, 2004):

$$Y_R(\boldsymbol{u}) = G^{-1}\left(F_{R|m}(\boldsymbol{u})\right) \quad \forall \boldsymbol{u} \in A \tag{2.35}$$

Conventional univariate simulation can be conducted with this de-trended variable and the trend features would be restored in the back transformation. This method successfully reproduces the complex relationship between variables; however, simulated values should be examined due to some unexpected simulated values. Silva and Deutsch (2016) proposed a Gaussian mixture model that would be combined with the stepwise conditional transformation in geostatistics. The use of Gaussian mixtures allows the easy assessment of any conditional distributions of regionalized variables and removes the binning issue. More discussions of the stepwise conditional transformation and the Gaussian mixture model are found in Chapter 4.

Chapter 3

Trend Modeling

Trend modeling is inevitably subjective and dependent on the practitioner. Modeling a reasonable trend is essential for the best possible geostatistical prediction. The challenge of this chapter is to develop a practical framework to account for the deterministic and stochastic features of regionalized variables. The goal is to model an artifact-free trend model that leads to the best possible future predictions in an objective manner.

This chapter will primarily focus on creating reasonable trend models in the presence of non-stationary variables. Three objectives are summarized as: (1) the construction of a trend model by a mathematical function that realistically represents the deterministic character of the regionalized variable; (2) the calibration of the trend parameters, including the background value, the anisotropy and other specific considerations in an objective manner; and (3) the development of an objective function to assess the trend models in order to improve geostatistical modeling in the presence of a trend. Details of the trend modeling algorithm are described and illustrated with examples. A final section in this chapter focuses on the challenge of trend extrapolation modeling. The formation of a novel weighting function is proposed for trend extrapolation.

3.1 Parameterization of Trend Modeling

Trend modeling aims to use the limited available samples to map some primary large-scale features over a specified domain. The challenges of trend modeling are related to the complexities of the geological phenomena, the scale of the trend model, as well as computational constraints of the trend modeling algorithm. First, the trend model captures an approximation of the real geological process. Many geological characteristics are highly non-linear. A reasonable trend should be inferred in the presence of the best possible understanding of the geological context. Second, the trend appears at different scales. The decision should be made on how much of the variability to put into the trend model. The trend is modeled for large-scale features relative to the data spacing. The choice of the trend function, the spatial weighting function and the anisotropy are discussed for trend modeling.

3.1.1 Choice of Trend Function

Several trend functions are reviewed in Chapter 2. Weighted average statistics are considered reliable to characterize a realistic geological trend. The equation for the weighted trend $m(\boldsymbol{u})$ is expressed by:

$$m(\boldsymbol{u}) = \frac{\sum_{i=1}^{n} w(\boldsymbol{u}_i) \cdot z(\boldsymbol{u}_i)}{\sum_{i=1}^{n} w(\boldsymbol{u}_i)} \quad \forall \boldsymbol{u}, \boldsymbol{u}_i \in A \tag{3.1}$$

where $m(\boldsymbol{u})$ denotes the locally varying trend at the location being estimated $\boldsymbol{u}$. $\{\boldsymbol{u}_i, i = 1, \cdots, n\}$ represent a set of n surrounding observed data. $\{z(\boldsymbol{u}_i), i = 1, 2, \cdots, n\}$ denote the observed values at data locations. The weights are denoted by $\{w(\boldsymbol{u}_i), i = 1, 2, \cdots, n\}$ and they depend on the unsampled location $\boldsymbol{u}$ being considered. All samples are considered within the domain A.

The mathematical formulation of the trend modeling algorithm is understandable and accessible. This trend function averages the influence of extreme values and produces a smooth trend model.

3.1.2 Choice of Spatial Weighting Function

Several requirements should be accounted for in the proposed weighting function, including a smooth decrease with increased distance, positive weights, independence of units and a global consistency for all statistics (Machuca-Mory, 2010). There are alternative spatial interpolation algorithms discussed in Chapter 2 that could be applied, but they have disadvantages relative to the proposed method.

An exponential weighting function that has less numerical difficulties and produces a smooth model consistent with the notion of a trend model is considered. Observed data are connected with a non-linear spatial pattern. More weight is given to close samples and less weight to more distant samples. The proposed Gaussian-like weights $w(\boldsymbol{u})$ are written as:

$$\begin{aligned} w(\boldsymbol{u}) = [\epsilon + (1.0 - \epsilon) \cdot \exp(-3.0 \cdot d^2(\boldsymbol{u}_{i0})] \cdot w_{\text{dec}}(\boldsymbol{u}_i) \\ \forall \boldsymbol{u}, \boldsymbol{u}_i, \boldsymbol{u}_{i0} \in A \text{ and } i = 1, 2, \cdots, n \end{aligned} \tag{3.2}$$

here the total weights will be scaled to a unit sum. ϵ is a background value that avoids computational problems and controls the smoothness of the trend estimate far from the data. $\exp(\cdot)$ is the exponential function. $\{d(\boldsymbol{u}_{i0}), i = 1, 2, \cdots, n\}$ are the standardized anisotropy-corrected distances between the observed locations $\{\boldsymbol{u}_i, i = 1, 2, \cdots, n\}$ and the location $\boldsymbol{u}_0$ where the trend is being estimated. a_{hmax} is a distance factor that ensures a continuous trend. This factor would be determined by a smoothing parameter and is discussed below. The number of the nearest samples involved in the calculation is determined by a threshold distance, beyond which points receive a background value in the calculation.

$\{w_{\text{dec}}(\boldsymbol{u}_i), i = 1, 2, \cdots, n\}$ are the declustering weights assigned to the observed data. Note that declustering is considered if the samples are preferentially sampled.

The weighting function would be tuned so that the trend is smooth and free of artifacts. However, the moving average trend function with such exponential weighting scheme is reasonable. The weights $w(\boldsymbol{u})$ approach to the background value (with declustering) when the distances $\{d(\boldsymbol{u}_{i0}), i = 1, 2, \cdots, n\}$ become large. Furthermore, the trend estimate values approach the global (declustered) mean. This may not be a desirable feature when extrapolating beyond the data in the presence of a large scale trend. An alternative weighting function should be considered. Trend extrapolation will be discussed at the end of this chapter.

3.1.3 Choice of Anisotropy

Geologic data exhibit spatial variability that depends on directions. In geostatistical modeling applications, the experimental variogram is a practical tool to reveal the behavior of the spatial data and quantify the anisotropy of the regionalized variable. The anisotropy must be accounted for to obtain a realistic trend model.

Consider that $\boldsymbol{u}_0$ represents the location where the trend is being estimated and $\{\boldsymbol{u}_i, i = 1, \cdots, n\}$ are the data locations. $\{\boldsymbol{h}_{i0}, i = 1, \cdots, n\}$ represent the distance vectors between the location being considered and the data locations, where

$$\boldsymbol{h}_{i0} = \boldsymbol{u}_i - \boldsymbol{u}_0 = \begin{bmatrix} h_{\text{x,i0}} \\ h_{\text{y,i0}} \\ h_{\text{z,i0}} \end{bmatrix} \quad i = 1, \cdots, n \text{ and } \quad \forall \boldsymbol{h}_{i0}, \boldsymbol{u}_0, \boldsymbol{u}_i \in A \tag{3.3}$$

here $\boldsymbol{h}_{i0}^{T} = \{h_{\mathrm{x,i0}}, h_{\mathrm{y,i0}}, h_{\mathrm{z,i0}}\}$ are the distances in the original X, Y and Z axes, respectively. The vectors $\{\boldsymbol{h}_{i0}, i = 1, \cdots, n\}$ should be rotated to the principal directions of anisotropy. The anisotropy-corrected vectors after the rotation $\left\{\boldsymbol{h}_{i0}^{'}, i = 1, \cdots, n\right\}$ are given as:

$$\boldsymbol{h}_{i0}^{'} = \begin{bmatrix} h_{\mathrm{min,i0}} \\ h_{\mathrm{max,i0}} \\ h_{\mathrm{vert,i0}} \end{bmatrix} = [\mathbf{T}] \cdot \boldsymbol{h}_{i0} = [\mathbf{T}] \cdot \begin{bmatrix} h_{\mathrm{x,i0}} \\ h_{\mathrm{y,i0}} \\ h_{\mathrm{z,i0}} \end{bmatrix} \quad i = 1, \cdots, n \tag{3.4}$$

here $[\mathbf{T}]$ is the rotation matrix following the consistent definitions of the directions and ranges in variogram models, that is,

$$[\mathbf{T}] = \underbrace{\begin{bmatrix} \frac{1}{a_{\mathrm{hmin}}} & 0 & 0 \\ 0 & \frac{1}{a_{\mathrm{hmax}}} & 0 \\ 0 & 0 & \frac{1}{a_{\mathrm{hvert}}} \end{bmatrix}}_{\text{Scaling Range}} \cdot \underbrace{\begin{bmatrix} cos\theta & 0 & sin\theta \\ 0 & 1 & 0 \\ -sin\theta & 0 & cos\theta \end{bmatrix}}_{\text{Dip Rotation}} \cdot \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\beta & sin\beta \\ 0 & -sin\beta & cos\beta \end{bmatrix}}_{\text{Plunge Rotation}} \cdot \underbrace{\begin{bmatrix} cos\alpha & -sin\alpha & 0 \\ sin\alpha & cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{Azimuth Rotation}} \tag{3.5}$$

$\{a_{\mathrm{hmin}}, a_{\mathrm{hmax}}, a_{\mathrm{hvert}}\}$ are the ranges in the minimum and maximum horizontal directions (within the plane of the greatest continuity) and the vertical direction (perpendicular to the plane of the greatest continuity), respectively. The dip rotation denotes a counter-clockwise rotation with a angle of θ where Y axis remains the same. The plunge rotation is a rotation matrix corresponding to a counter-clockwise rotation with a angle of β around X axis. The matrix of the azimuth rotation corresponds to a clockwise rotation of X-Y plane with the angle of α where Z axis remains the same (Goovaerts, 1997; Deutsch and Journel, 1998; Leuangthong et al., 2011; Deutsch, 2015). Figure 3.1 shows the schematic of the anisotropy rotation in the directions. The anisotropy takes the form of an ellipsoid oriented according to the directions and length scales of continuity.

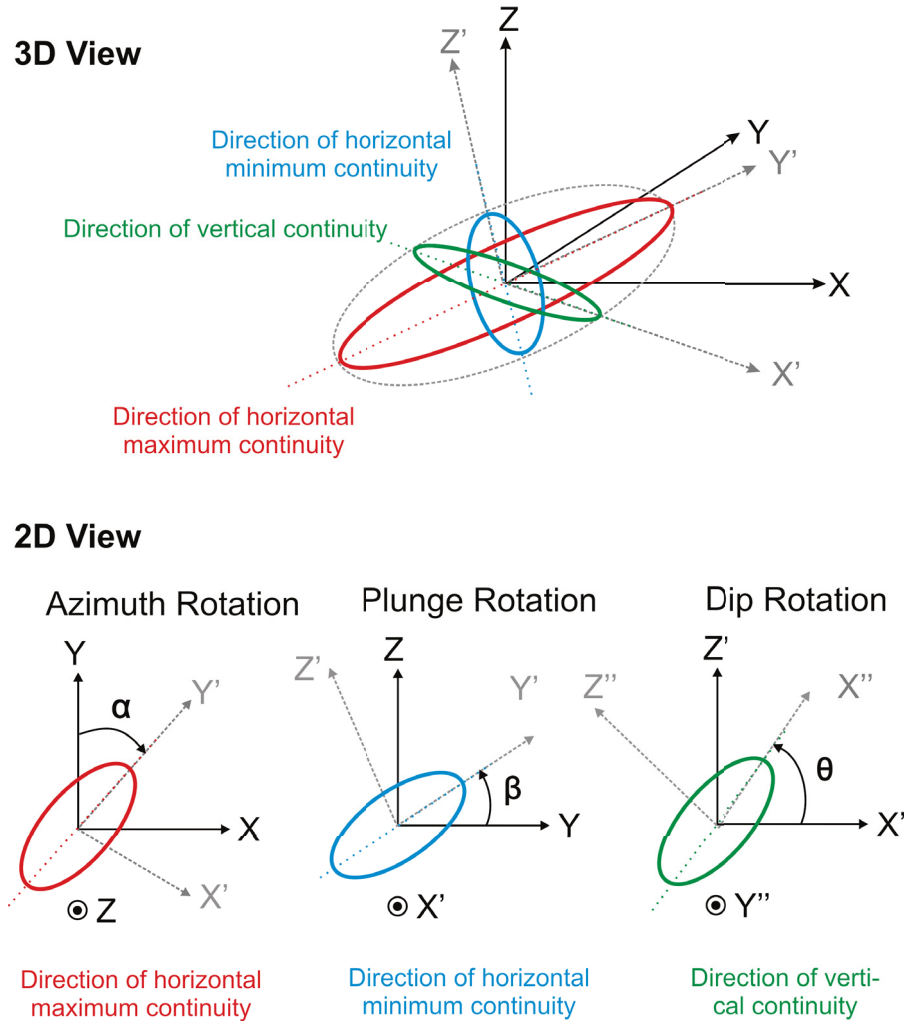


Figure 3.1: Schematic illustration of the anisotropy rotation on 2-D and 3-D views (modified from Deutsch and Journel (1998))

Choice of the appropriate directions and the scaling ranges are important for trend modeling. The directions of the anisotropy $\{\alpha, \beta, \theta\}$ in Equation 3.5 should follow the continuity of the regionalized variable indicated by the directional variogram models. The scaling ranges in Equation 3.5 quantify the anisotropy and are often determined by the variogram ranges. The scalar normalized distances with the anisotropy $d(\boldsymbol{u}_{i0})$ in Equation 3.2 are calculated as:

$$d(\boldsymbol{u}_{i0}) = \sqrt{\frac{h_{\max,i0}^2 + {h_{\min,i0}}^2 + {h_{\text{vert},i0}}^2}{|A| \cdot \text{SP}}} \quad i = 1, \cdots, n \tag{3.6}$$

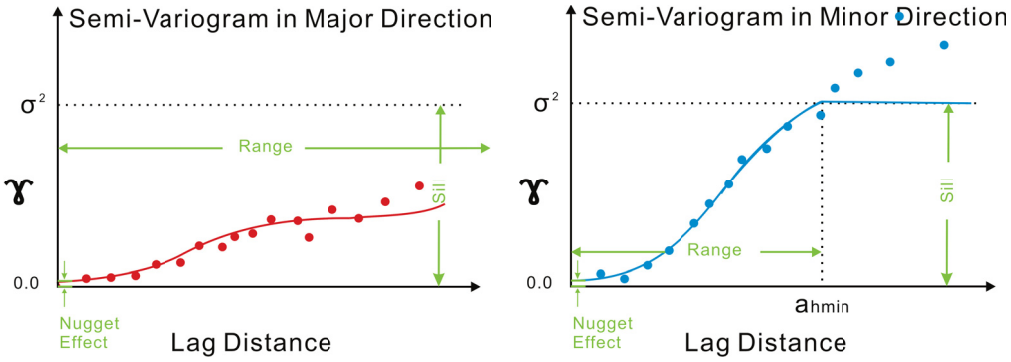
here $|A|$ represents the domain size and SP represents a smoothing parameter. This formalism allows a dimensionless smoothing parameter to be tuned to adjust the trend model with less concern for the modeling units.

Tabular deposits such as coal seams may pose a challenge in that the variogram ranges may not be calculated accurately due to a lack of data in some directions or due to a zonal anisotropy. In this case, the ranges in different directions could be specified independently.

3.1.4 Anisotropy Correction

As mentioned above, the anisotropy ratios are often obtained from the variogram ranges in different directions. Data with an obvious trend often show a strong anisotropy that does not have a straightforward range parameter. For example, a zonal anisotropy makes the range a difficult parameter to estimate. Figure 3.2 shows a schematic illustration of the variogram models with a zonal anisotropy. The left figure indicates a variogram model in the major direction, while the right figure shows the model in the minor direction. The red and blue markers denote the experimental variograms in the major and minor directions, while the red and blue lines represent the fitted stationary variogram models in different directions. The range in the minor direction a_{hmin} (blue line) is estimated as

the distance where the experimental variograms (blue markers) reach to the sill variance σ^2, see Figure 3.2b; however, the range in the major direction a_{hmaj} (red line) is unclear, see Figure 3.2a. The range in the major direction is often set as an arbitrarily large number for the variogram modeling, but this causes the anisotropy ratio to become unstable.



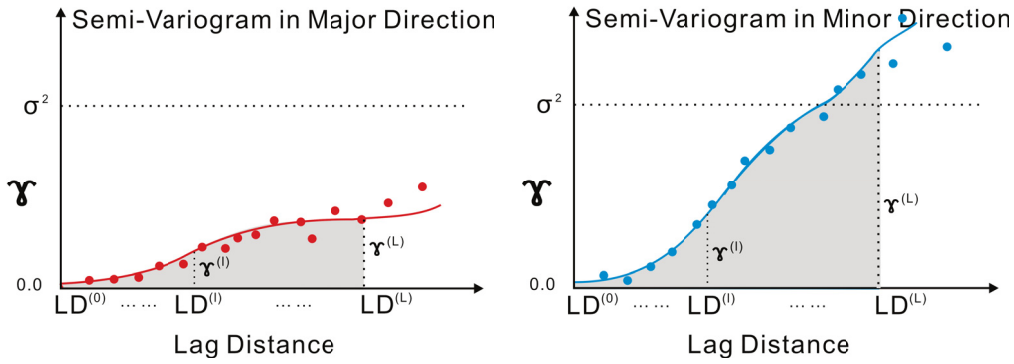
(a) Variogram model in major direction **(b)** Variogram model in minor direction

Figure 3.2: Difficulty on determining the range with a zonal anisotropy

The calculation of the anisotropy ratio should be done in a robust manner. Instead of using the range ratios, the area ratios under the variogram model curves are considered. The rotation matrix $[\mathbf{T}]$ in Equation 3.5 is corrected to:

$$[\mathbf{T}'] = \underbrace{\begin{bmatrix} \text{area}_{\text{hmin}} & 0 & 0 \\ 0 & \text{area}_{\text{hmax}} & 0 \\ 0 & 0 & \text{area}_{\text{hvert}} \end{bmatrix}}_{\text{Corrected Scaling Range}} \cdot \underbrace{\begin{bmatrix} cos\theta & 0 & sin\theta \\ 0 & 1 & 0 \\ -sin\theta & 0 & cos\theta \end{bmatrix}}_{\text{Dip Rotation}} \cdot \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\beta & sin\beta \\ 0 & -sin\beta & cos\beta \end{bmatrix}}_{\text{Plunge Rotation}} \cdot \underbrace{\begin{bmatrix} cos\alpha & -sin\alpha & 0 \\ sin\alpha & cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{Azimuth Rotation}} \tag{3.7}$$

here $\text{area}_{\text{hmax}}$ represents the area under the variogram model curve in the major direction, while $\text{area}_{\text{hmin}}$ and $\text{area}_{\text{hvert}}$ denote the areas in the minor and vertical directions, respectively. The advantage of this approach is that the anisotropy ratios are stable and robust, and not depend on the unstable choice of an arbitrarily range.



(a) Variogram model in major direction **(b)** Variogram model in minor direction

Figure 3.3: Schematic illustration of the proposed anisotropy ratio calculation on variogram models

Two questions need to be considered in calculating the stable anisotropy ratio: how to fit the variogram models with non-stationary features and what is the available lag distance to consider for the area calculation. The experimental variogram in the minor direction (blue markers) in Figure 3.2 continues to climb steadily above the sill variance showing the presence of a spatial trend, while the experimental variogram in the major direction (red markers) stays constant in the long distance due to the regionalized variable having a great continuity in this direction. The proposed variogram fittings should be modeled following the experimental variograms without considering the sill variance, that is, the variogram may be fitted above the sill in one direction and below the sill in another direction, see lines in Figure 3.3.

The proposed anisotropy by area ratios is then calculated. Suppose the lag distances of the non-stationary variogram model in X axis are divided into several segments $\left\{LD^{(l)}, l = 0, 1, \cdots, L\right\}$ and the corresponding variogram values are labeled as $\left\{\gamma^{(l)}, l = 0, 1, \cdots, L\right\}$. The area under the variogram model curve is calculated with these segments that is given as:

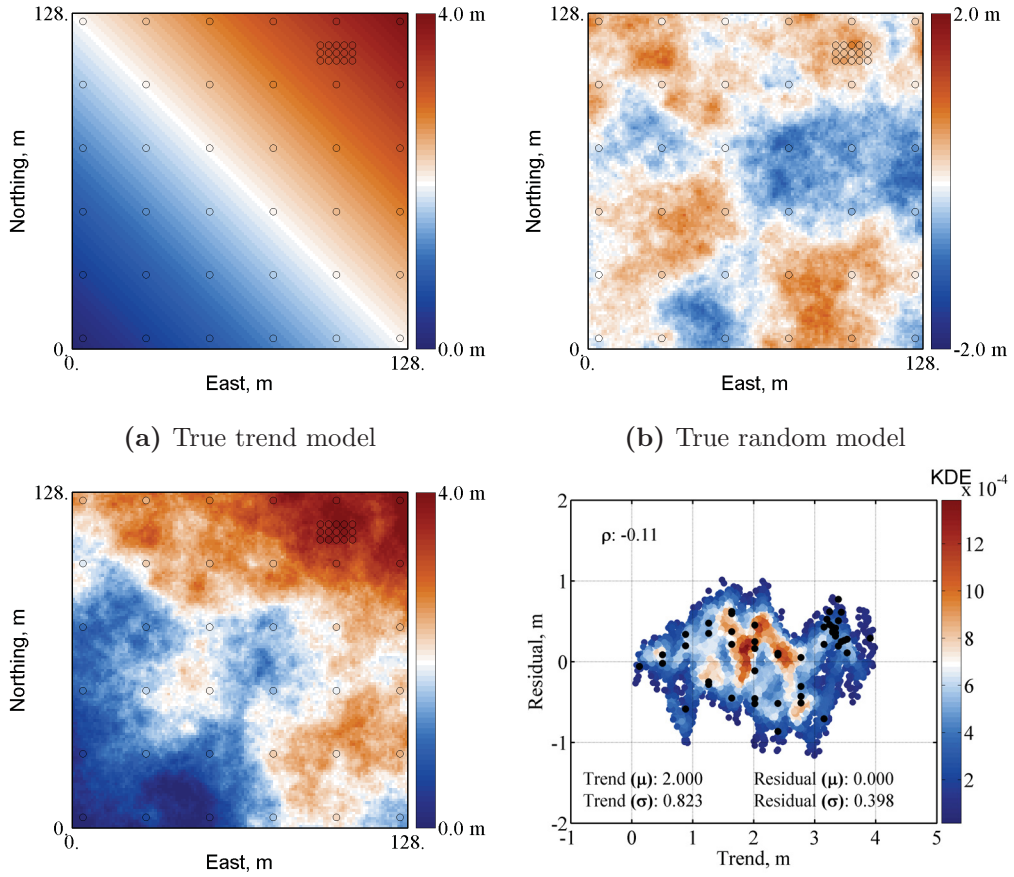
$$\text{area} = \int_{LD^{(0)}}^{LD^{(L)}} \gamma(\boldsymbol{h}) d(\boldsymbol{h}) \tag{3.8}$$

here $\gamma(\boldsymbol{h})$ represent the variogram values and $\left\{LD^{(l)}, l = 0, 1, \cdots, L\right\}$ are the lag distances at a segment of l from the origin. $\gamma^{(0)}$ is the initial variogram value that denotes the nugget effect of the variogram model and $LD^{(0)}$ is the origin. $\gamma^{(L)}$ and $LD^{(L)}$ represent the maximum variogram value and the corresponding maximum lag distance for the area calculation, respectively. The maximum lag distance should be chosen carefully in calculating the area ratios. A value one half the domain size in each direction appears reasonable.

3.2 Implementation of Trend Modeling

The trend modeling function with the exponential weighting in Equation 3.1 and 3.2 has several parameters. The selections of these parameters should be optimized to minimize subjectivity and improve the final geostatistical model to the greatest extent possible. These parameters include: (1) background value ϵ; (2) anisotropy, including directions $\{\alpha, \beta, \theta\}$ and ratios $\{r_1, r_2\} = \left\{\frac{\text{area}_{\text{hmin}}}{\text{area}_{\text{hmax}}}, \frac{\text{area}_{\text{hmin}}}{\text{area}_{\text{hmax}}}\right\}$; (3) declustering weights $\{w_{dec}(\boldsymbol{u}_i), i = 1, \cdots, n\}$ and (4) smoothing parameter SP. These parameters will be investigated and validated by cross validation with examples where the true trend is exhaustively known. The suggestions on the selection of optimal trend parameters would be useful for practitioners.

The trend parameters will be validated with cases where a true trend is known. A small 2-D synthetic example is presented for demonstrating the parameters of trend modeling. Figure 3.4 shows the reference models. Figure 3.4a illustrates the synthetic true trend with a strong anisotropy in the direction of 135°, while Figure 3.4b shows a random residual model on 128×128 regular pixel grids. Figure 3.4c shows a 2-D well-defined true image that is combined with the trend and the residual model. 51 data with black circles are sampled from this exhaustive image. Data are preferentially collected in high grade zones. The



(a) True trend model

(b) True random model

(c) True subsurface model showing a N45E trend

(d) Scatter plot with 51 data showing a correlation of −0.11

Figure 3.4: Synthetic subsurface models for trend parameter selections. The true trend values are shown in Figure 3.4a and the true residual values are shown in Figure 3.4b. The true subsurface model in Figure 3.4c is combined with the true trend model and the true residual model. 51 black circles/markers represent the conditioning data. Figure3.4d shows a 2-D kernel density estimation (KDE) scatter plot

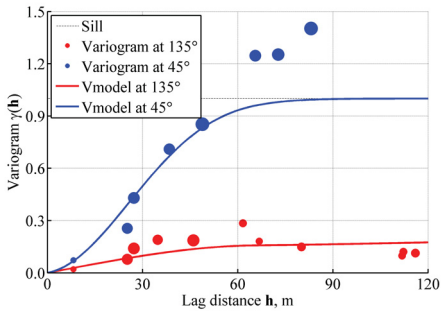


Figure 3.5: Variogram models from 51 data. The sizes of the markers represent the relative pairs of the experimental variogram in each lag distance

trend model could not be appropriately extrapolated beyond the area sampled by the data. Thus, in this example, several control points close to the edge of the study area are included in order to avoid this extrapolation issue. The correlation coefficient between the trend and the residuals at 51 data (black markers) locations is shown on a 2-D density estimation that is around -0.11 in Figure 3.4d. A standardized experimental variogram model from 51 data is calculated and shown with markers in Figure 3.5. The major direction is at 135° and the minor direction is at 45°. The experimental variogram (blue markers) goes above the sill in the minor direction, while the experimental variogram in the major direction (red markers) shows the presence of a zonal anisotropy and stays below the expected sill variance. The variograms are modeled with Equation 3.9 as below:

$$\gamma(\boldsymbol{h}) = 0.15 \cdot \mathrm{Sph}(\boldsymbol{h})_{\substack{\alpha=135^\circ \\ a_{\mathrm{hmax}}=65\ \mathrm{m} \\ a_{\mathrm{hmin}}=65\ \mathrm{m}}} + 0.85 \cdot \mathrm{Gauss}(\boldsymbol{h})_{\substack{\alpha=135^\circ \\ a_{\mathrm{hmax}}=1{,}200\ \mathrm{m} \\ a_{\mathrm{hmin}}=65\ \mathrm{m}}} \tag{3.9}$$

here $\gamma(\cdot)$ represents the variogram model and $\boldsymbol{h}$ denotes the lag vector. $\mathrm{Sph}(\cdot)$ represents the spherical structure of the variogram model and $\mathrm{Gauss}(\cdot)$ represents the Gaussian structure. α is the azimuth rotation in the horizontal direction,

while a_{hmax} and a_{hmin} are the variogram ranges in the major and minor directions. No nugget effect is considered in this synthetic example. The variogram models are isotropic at small distances and anisotropic at long distances. The conventional anisotropy ratio with range calculations is $r_1 = \dfrac{1,200 \text{ m}}{60 \text{ m}} = 18.46$. This anisotropy ratio is unstable due to an ill-defined range in the major direction.

Cross validation is a simple way to check the models. The purpose is to assist in choosing the correct parameters to arrive at the best possible trend. The difference between the true trend value and the modeled trend value at all locations gives a measure of how well the estimated trend approximates the true trend. There are various performance indexes to measure the discrepancy between the true trend and the predicted trend. Commonly used error measurements include the mean error value and the mean squared error value. Mean error determines the bias in the estimates, while mean squared error indicates the precision in the estimates (Sheiner and Beal, 1981). Here, mean squared error values are used to help choose the best parameters for trend modeling. Mean squared error (MSE) value mse is given as:

$$mse = \frac{1}{N}\sum_{i=1}^{N}[m(\boldsymbol{u}_i) - m^*(\boldsymbol{u}_i)]^2 \quad \forall \boldsymbol{u}_i \in A \tag{3.10}$$

where $\{m(\boldsymbol{u}_i), i = 1, \cdots, N\}$ is the known trend values at exhaustive locations $\{\boldsymbol{u}_i, i = 1, 2, \cdots, N\}$ and $\{m^*(\boldsymbol{u}_i), i = 1, \cdots, N\}$ is the exhaustive predicted trend values at all locations within the entire domain A. mse should be minimized with correctly chosen trend modeling parameters.

3.2.1 Background Value

The background value ϵ in Expression 3.2 is designed for computational stability when all of the weights to the samples might go to nearly zero. Figure 3.6 shows

the weighting functions with different background values. The weight reduces gradually as increasing the distance from the estimate location and stabilizes to the background value when it is far away from the estimate location. The background value permits some contributions from distant samples.

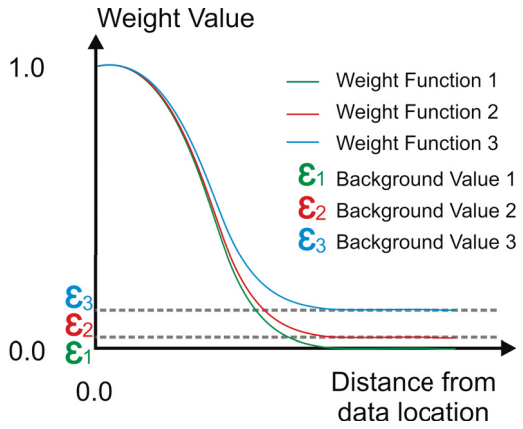
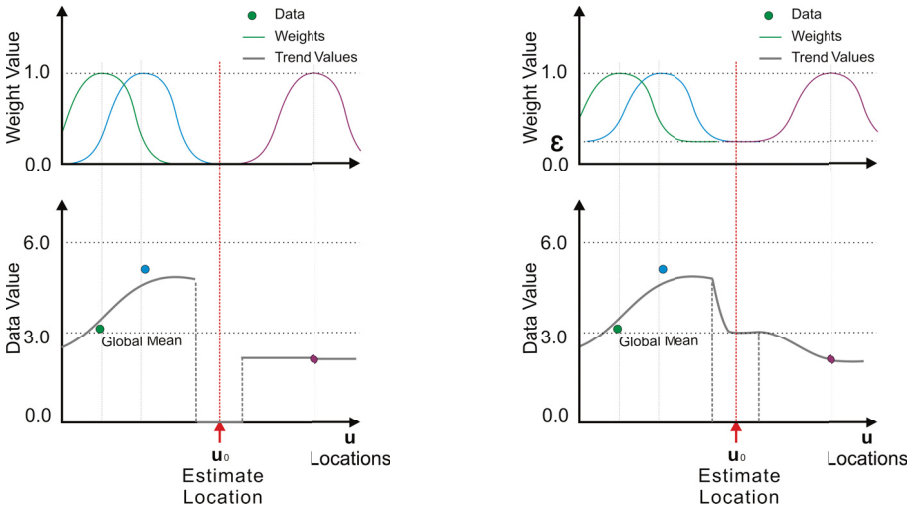


Figure 3.6: Schematic illustration of the Gaussian weighting functions



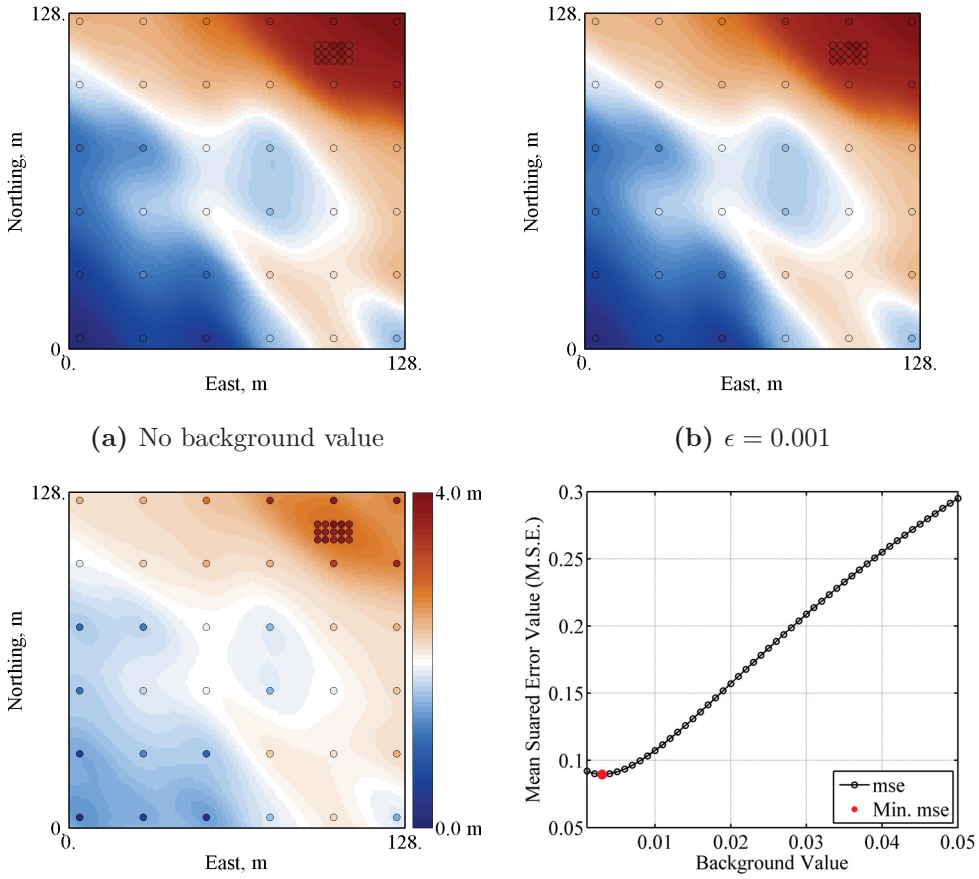
(a) No background value **(b)** A background value ϵ

Figure 3.7: Schematic illustration of the importance of a background value

Figure 3.7 shows the importance of considering a background value into trend modeling. The estimate location receives a gradually reduced weight from samples. The samples that are far away from a location being estimated contribute zero weights. These zero weights lead to an unstable jump in the trend model, see Figure 3.7a. A noncontinuous trend surface could be generated without considering the background value. The trend value approaches the global mean if a background value ϵ is considered in trend modeling. The introduction of a background value ensures a smooth trend, see Figure 3.7b. An appropriate background value should be incorporated into trend modeling.

Several trends are modeled with different background values. In order to show the importance of the incorporation of a background value, the spread of the weighting kernels should be steep, that is, the smoothing parameter is set to a small number 0.3. The rest of the trend parameters stay the same. Figure 3.8 shows modeled trends with different background values of a zero value, 0.001 and 0.05, respectively. The result shows that the trend with a large background value shows a smoothly varying model compared with the one with a small background value. The mean squared error comparison is made between the modeled trend and the true trend and shown in Figure 3.8d. The optimal background value is 0.002 that minimizes the mean squared error value.

The observation that a small background value leads to the best trend leads to the consideration of a local search in the trend calculation. Considering all of the data in the domain may not be required; however, for computational stability and to avoid search artifacts, a global search for data in trend modeling is considered. Thus, setting up a background value with a small number for computational stability and not for the smoothness is recommended. Other trend parameters are tuned for trend smoothness.



(a) No background value **(b)** $\epsilon = 0.001$

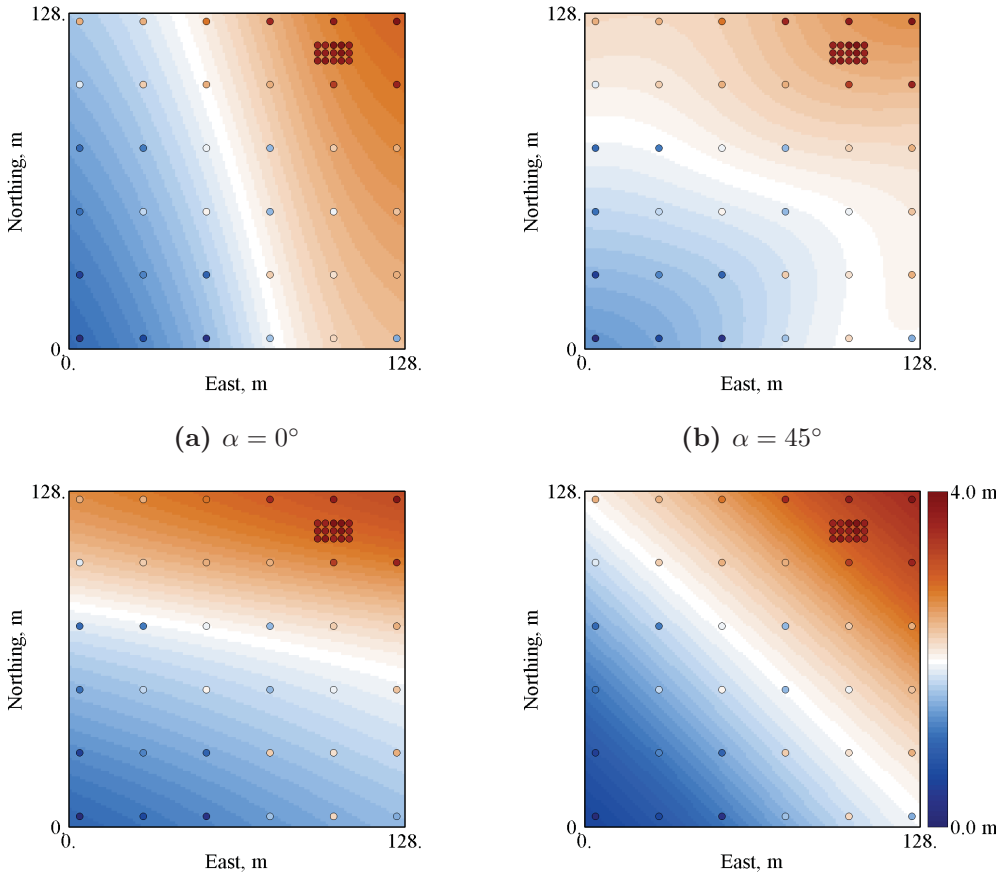
(c) $\epsilon = 0.050$ **(d)** MSE comparison

Figure 3.8: Modeled trends with different background values and MSE comparison

3.2.2 Anisotropy

The anisotropy is considered in the distance values $\{d(\boldsymbol{u}_{i0}), i = 1, 2, \cdots, n\}$ in the exponential weighting function in Equation 3.2. The same 2-D example is used to illustrate the anisotropy (angles and ratios) in trend modeling.

The first application is the choice of the anisotropy directions in trend modeling. Thirty-six trends with different anisotropy directions are modeled and with a 5° increment. Four of them, 0°, 45°, 90° and 135°, are shown in



(a) $\alpha = 0°$ **(b)** $\alpha = 45°$

(c) $\alpha = 90°$ **(d)** $\alpha = 135°$

Figure 3.9: Modeled trends with different anisotropy directions

Figure 3.9. The mean squared error calculation is shown in Figure 3.10. The map that shows the lowest mean squared error is obtained with the direction of 135°. The anisotropy direction of the trends matches the major direction in the variogram models in Figure 3.5. Thus, as expected, the selection of anisotropy directions in trend modeling should follow the global variogram model.

Another test on the anisotropy ratio is developed. The conventional anisotropy ratio in Figure 3.5 is estimated as 18.46 : 1 from the large-scale structure range in Equation 3.9. Figures 3.11a, 3.11b and 3.11c show three trend

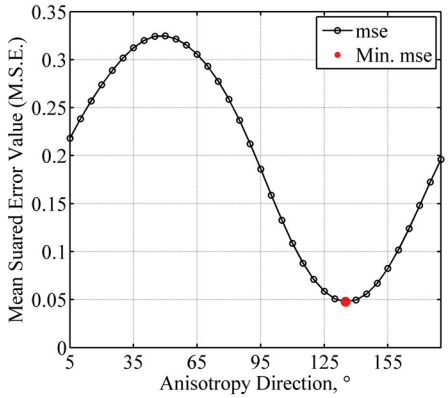


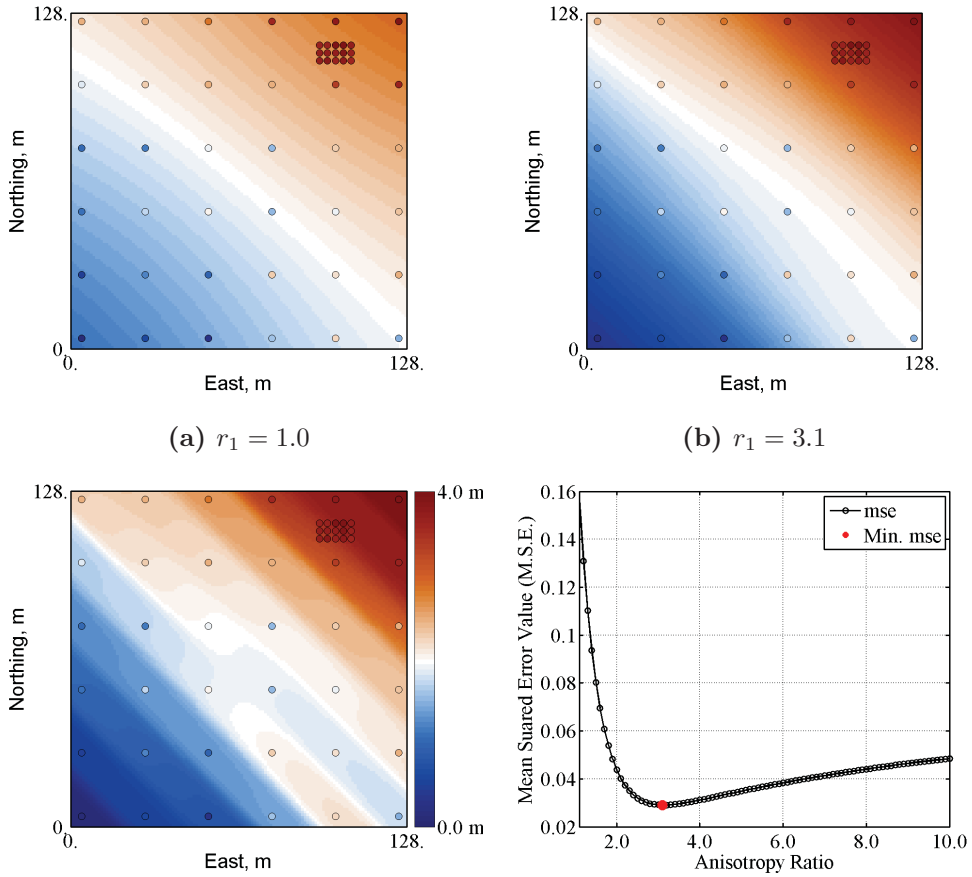
Figure 3.10: MSE comparisons with anisotropy directions

models with anisotropy ratios of 1.0 : 1, 3.1 : 1 and 10.0 : 1, respectively. The trend model is excessively smooth with an isotropic ratio, while it shows abrupt jumps with a large anisotropy ratio. The mean squared error value between the true trend and the modeled trend is calculated and shown in Figure 3.11d. According to the minimization of the mean squared error value, the test indicates that the optimal anisotropy ratio is around 3.1 : 1 for trend modeling. The result shows that the anisotropy in trend modeling is more isotropic than the one suggested by the stationary variogram ranges from the experimental variogram. The anisotropy ratio should be re-considered so that it supports a smooth trend.

The developed anisotropy ratio by the areas under the variogram model curves is considered. The non-stationary variogram model is given as:

$$\gamma(\boldsymbol{h}) = 0.20\cdot\mathrm{Sph}(\boldsymbol{h})\underset{\substack{a_{\mathrm{hmin}}=280\text{ m}}}{\overset{\alpha=135^\circ}{\underset{a_{\mathrm{hmax}}=60\text{ m}}{}}} + 1.00\cdot\mathrm{Sph}(\boldsymbol{h})\underset{\substack{a_{\mathrm{hmin}}=280\text{ m}}}{\overset{\alpha=135^\circ}{\underset{a_{\mathrm{hmax}}=3{,}000\text{ m}}{}}} + 1.80\cdot\mathrm{Sph}(\boldsymbol{h})\underset{\substack{a_{\mathrm{hmin}}=280\text{ m}}}{\overset{\alpha=135^\circ}{\underset{a_{\mathrm{hmax}}=10{,}000\text{ m}}{}}} \tag{3.11}$$

here three spherical-type variograms with a continuity of 135° are considered. The total contribution of the non-stationary variogram model is $0.2+1.0+1.8 = 3.0$. The maximum distance $LD^{(L)}$ for calculating the anisotropy ratios should

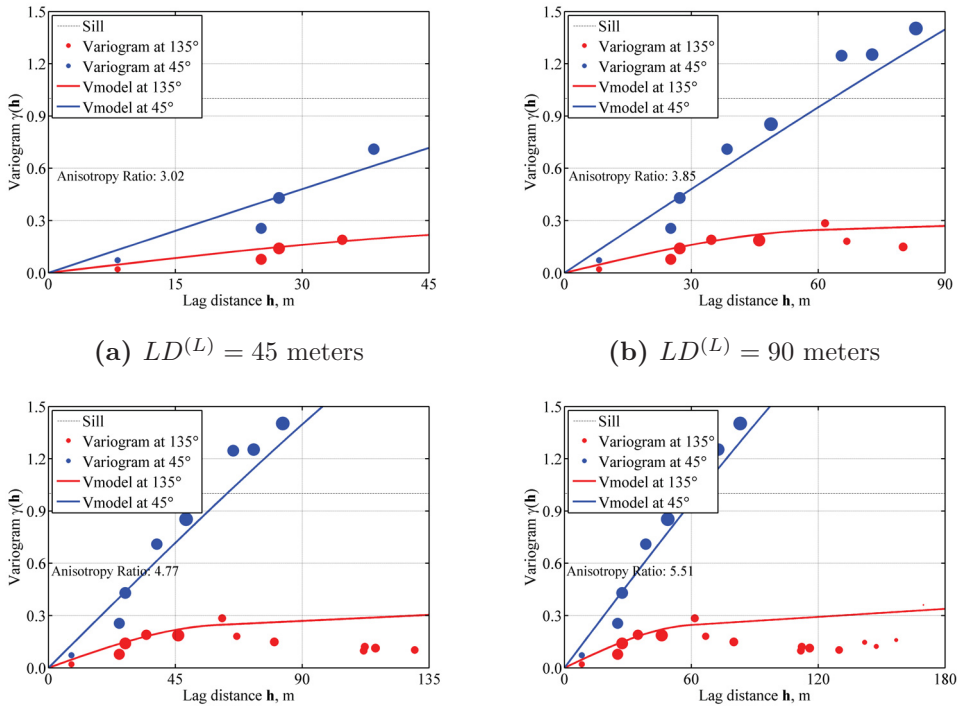


(a) $r_1 = 1.0$ **(b)** $r_1 = 3.1$

(c) $r_1 = 10.0$ **(d)** MSE comparison

Figure 3.11: Modeled trends with different anisotropy ratios

be then determined from the domain size. Figure 3.12 shows the non-stationary variogram models with different maximum distances. These distances that are 45 meters, 90 meters, 135 meters and 180 meters corresponding to a quarter, half, three quarters and whole of the domain size with the rotation are considered. The sizes of the markers represent the relative pairs of the experimental variogram in each lag distance. The anisotropy ratios based on the area calculations under the variogram model curves are 3.02 : 1, 3.85 : 1, 4.77 : 1 and 5.51 : 1, respectively, for these cases. The area ratio with a maximum distance of $LD^{(L)} = 45$ meters closes



(a) $LD^{(L)} = 45$ meters **(b)** $LD^{(L)} = 90$ meters

(c) $LD^{(L)} = 135$ meters **(d)** $LD^{(L)} = 180$ meters

Figure 3.12: Inference of the maximum distance in calculation of proposed anisotropy ratio. The sizes of the markers represent the relative pairs of the experimental variogram in each lag distance.

to the minimized mean squared error value in Figure 3.11d; however, the pairs of the experimental variogram are not reliable due to fewer lags are calculated in pairing samples within such a short distance. The experimental variograms are also unstable beyond half of the domain size. Thus, half of the domain size $LD^{(L)} = \frac{1}{2}|A|$ is considered reliable to model the non-stationary variogram models, and further, calculate the anisotropy ratios which is recommended. The proposed anisotropy ratio in this case is close to $3.85 : 1$ when $LD^{(L)} = 90$ meters.

In order to show the stable proposed anisotropy ratio calculation, a test

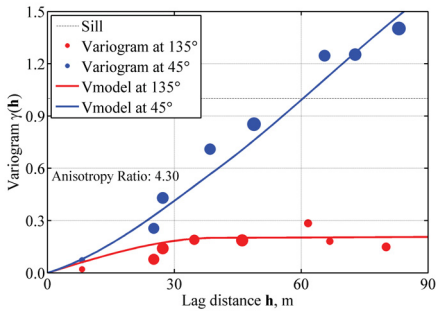


Figure 3.13: Inference of a stable anisotropy ratio in the presence of non-stationary features. The sizes of the markers represent the relative pairs of the experimental variogram in each lag distance

on changing the non-stationary variogram model is considered. Another non-stationary variogram model fitted with the same experimental variograms is given as:

$$\gamma(\boldsymbol{h}) = 0.20 \cdot \mathrm{Sph}(\boldsymbol{h})\underset{\substack{a_{\mathrm{hmax}}=60\text{ m}\\ a_{\mathrm{hmin}}=140\text{ m}}}{{}_{\alpha=135^\circ}} + 2.30 \cdot \mathrm{Gauss}(\boldsymbol{h})\underset{\substack{a_{\mathrm{hmax}}=5{,}000\text{ m}\\ a_{\mathrm{hmin}}=140\text{ m}}}{{}_{\alpha=135^\circ}} \tag{3.12}$$

here the variogram model with two spherical structure in a continuity of 135° is used. The total contribution of the non-stationary variogram model is 0.2+2.3 = 2.5. The anisotropy ratio under the variogram model curves within half of the domain size is 4.30 : 1, see Figure 3.13. The result indicates that the proposed anisotropy ratio does not significantly change with the way of fitting the non-stationary variograms. This result also provides a relatively stable anisotropy ratio.

3.2.3 Declustering Weight

Geologic data are often preferentially sampled at locations with high values because of their economic and technical importance (Davis, 2002). The proposed weighting method is affected by the clustering data. Such clustered sampling

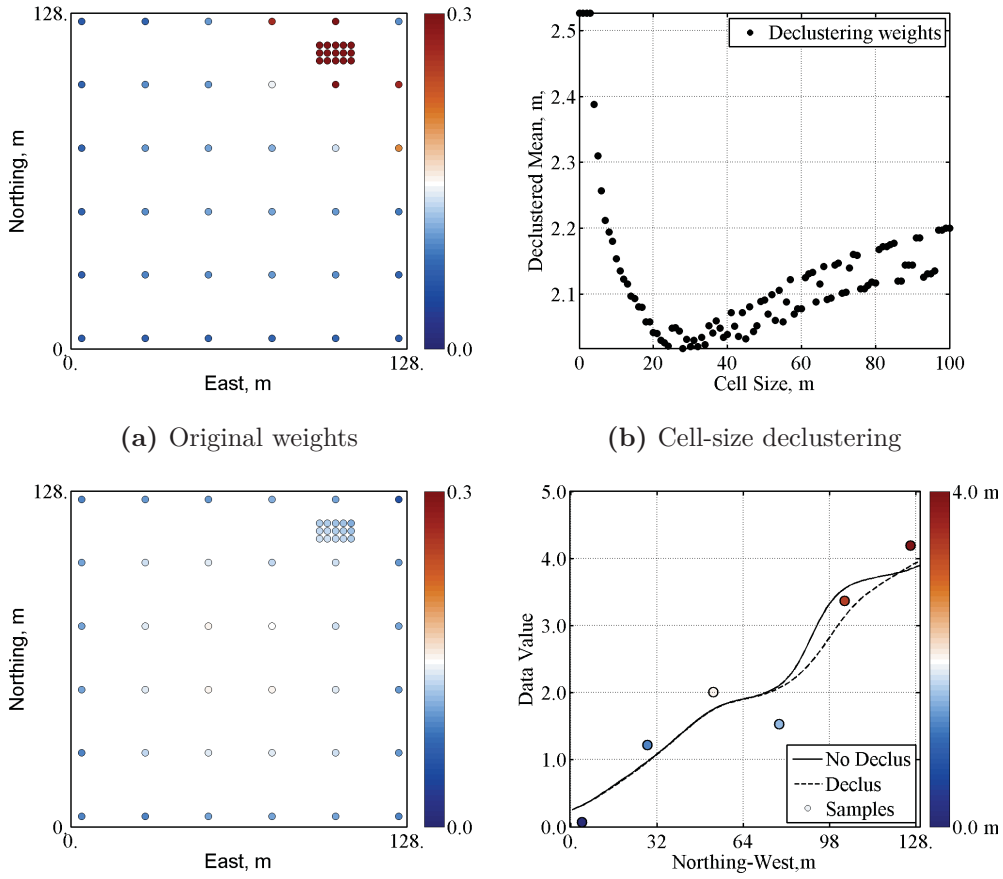
may induce a bias and impact the results. Declustering should be used to correct the weights in trend modeling.

Declustering is applied for correcting the histogram of spatially clustered data. The declustering weights $\{w_{\text{dec}}(\boldsymbol{u}_i), i = 1, \cdots, n\}$ in Equation 3.2 are calculated based on the distance between the data. Polygonal declustering and cell declustering are two common declustering methods. Polygonal declustering is straightforward, but it is considered unstable in 3-D (Pyrcz and Deutsch, 2014). Cell declustering method is considered to be a robust method where the weight is assigned based on the cell size and the number of data in different cells. This method is suitable for a range of data configurations (Pyrcz et al., 2003).

The original weights at the data locations are shown in Figure 3.14a. The declustering with a 24-meter cell size is applied with 51 data and shown in Figure 3.14b to obtain the declustered statistics. Figure 3.14c shows the corrected weight where the samples in the high-density zones is given lower weight than that in the low-density zones. Figure 3.14d highlights the use of declustering weights. After the correction, the trend is modeled with a better fit to the grades where the closer spaced samples are collected. The use of declustering weights ensures that the final trend model is unbiased. Thus, it might be appropriate to consider declustering in situations when data are clustered together.

3.2.4 Smoothing Parameter

The smoothing parameter SP simplifies the length scale settings for the major direction a_{hmax} in Equation 3.6. The smoothing parameter is an alternative for controlling the length scales, and further, controlling the spread of the weighting kernel. A smaller weighting kernel matches the data better, while a larger weighting kernel leads to a smoother trend model. The smoothing parameter is introduced to more easily understand and simplify the length scale settings for practitioners.

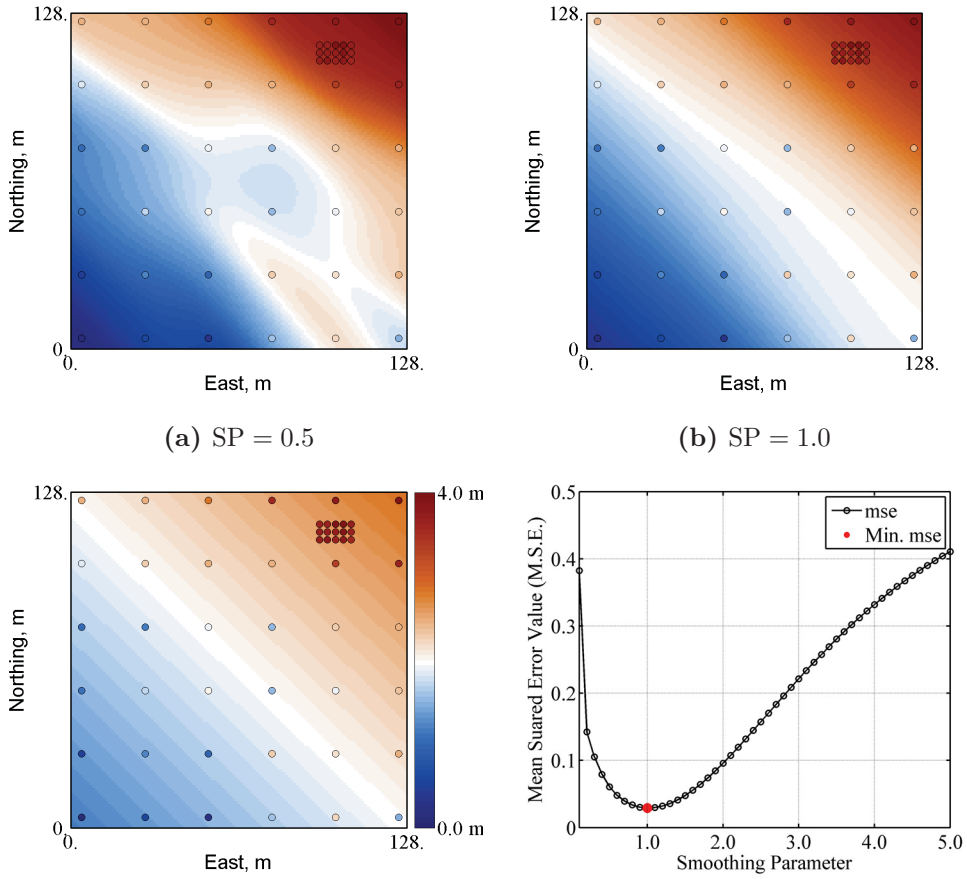


(a) Original weights **(b)** Cell-size declustering

(c) Weights correction **(d)** Trend models in diagonal section

Figure 3.14: Modeled trends with the declustering correction

The smoothing parameter changes the length scales of smoothing in a continuous way. Figures 3.15a, 3.15b and 3.15c show modeled trends with different smoothing parameter values of 0.5, 1.0 and 3.0, respectively. Other trend parameters are held constant. The trend model becomes smoother with increasing smoothing parameter. Figure 3.15d compares the true trend with the modeled trend based on the mean squared error value. In this case, a value of 1.0 appears optimal when the mean squared error value reaches the minimum. A user would set reasonable length scales and then fine tune the results with the smoothing



(a) SP = 0.5 **(b)** SP = 1.0

(c) SP = 3.0 **(d)** MSE comparison

Figure 3.15: Modeled trends with different smoothing parameters

parameter. The objective function for trend modeling will be inferred from this parameter and discussed in the next section.

3.3 Trend Optimization

The trend-like features in real geological phenomena are usually more complex than a simple model can capture. The ultimate aim of trend modeling is to

provide accurate and precise predictions for improved engineering designs. Visualization can help decide whether a trend model is appropriate for the data; however, this is subjective and could lead to an over- or under-fit trend model. An over-fitted trend leaves too little variation and too little uncertainty in the prediction and an under-fitted trend does not reproduce large-scale features (Pyrcz and Deutsch, 2014). A goal of this research is to develop an objective function to optimize the trend for the subsequent geostatistical modeling.

The proposed objective function should be calculable and straightforward to implement. Here, the objective function is to minimize the mean squared error in Equation 3.10 when the true trend is known. Then, the objective function can be applied where the true trend is not known.

3.3.1 Relations between Observations, Trend Values and Residual Values

The analysis of trend modeling is designed to separate the observed data $Z(\boldsymbol{u})$ into two components: a large-scale trend $m(\boldsymbol{u})$ influencing the entire domain and the residuals $R(\boldsymbol{u})$ that are the difference between the observed data and the trend values expressing the local effects. The trend values are modeled by neighboring observed data with the moving average method resulting in calculation of the residuals. The residuals, with respect to the neighboring observed data, are either positive or negative. A zero mean value of residuals $E\{R(\boldsymbol{u})\} = 0$ is assumed in Equation 2.17, in which the amount of the positive and negative values in residuals are equal likely. If the mean value of residuals $R(\boldsymbol{u})$ equals to zero, then it follows that $E\{Z(\boldsymbol{u})\} = m(\boldsymbol{u})$, that is, the local mean of the regionalized variable $Z(\boldsymbol{u})$ equals to the trend value $m(\boldsymbol{u})$. If these assumptions are made, the relationships of the decomposed geologic data could be explained as: (1) the trend values should match the observed data values as much as possible; and (2) the trend values and the residuals should be uncorrelated.

The variance of the random variable $Z(\boldsymbol{u})$ is given as:

$$\begin{aligned} Var\{Z(\boldsymbol{u})\} &= Var\{m(\boldsymbol{u}) + R(\boldsymbol{u})\} \\ \Rightarrow \sigma_Z^2 &= \sigma_m^2 + \sigma_R^2 + 2 \cdot C_{mR}(0) \quad \forall \boldsymbol{u} \in A \end{aligned} \tag{3.13}$$

here σ_Z^2, σ_m^2 and σ_R^2 represent the variance of the original data, the trend model and the residual model. The covariance $C_{mR}(0)$ is an important part of deciding on the trend at zero lag distances (McLennan, 2007). The variogram of the random variable $Z(\boldsymbol{u})$ can be decomposed to:

$$\gamma_Z(\boldsymbol{h}) = \gamma_m(\boldsymbol{h}) + \gamma_R(\boldsymbol{h}) + 2 \cdot \gamma_{mR}(\boldsymbol{h}) \quad \forall \boldsymbol{u}, \boldsymbol{h} \in A \tag{3.14}$$

here $\gamma_Z(\boldsymbol{h})$, $\gamma_m(\boldsymbol{h})$ and $\gamma_R(\boldsymbol{h})$ represent the direct variograms of the original data, the trend model and the residual model. $\gamma_{mR}(\boldsymbol{h})$ is the cross variogram between the trend values and the residual values at a lag distance of $\boldsymbol{h}$ and is related to the covariance $C_{mR}(\boldsymbol{h})$. Thus, the trend values and the residuals are assumed uncorrelated at all lag distances so that the minimized correlation between the trend $m(\boldsymbol{u})$ and residuals $R(\boldsymbol{u})$ is a promising solution.

Figure 3.16 shows the correlations among the observed data, the trend values and the residual values with increasing smoothing parameter. There are two concave functions, $\rho\{m(\boldsymbol{u}), R(\boldsymbol{u})\}$ and $\rho\{Z(\boldsymbol{u}), R(\boldsymbol{u})\}$, and one monotonic function, $\rho\{m(\boldsymbol{u}), Z(\boldsymbol{u})\}$. The possible numerical optimization would be selected from these two concave functions such that a simple maximization/minimization might be obtained.

The relationships between these components by the geological context are analyzed here: (1) the correlation between the trend values and the observed values should be maximized, that is, max. $\rho\{m(\boldsymbol{u}), Z(\boldsymbol{u})\}$; and (2) the absolute value of the correlation between the trend values and the residual values should be minimized, that is, min. $|\rho\{m(\boldsymbol{u}), R(\boldsymbol{u})\}|$. These correlations

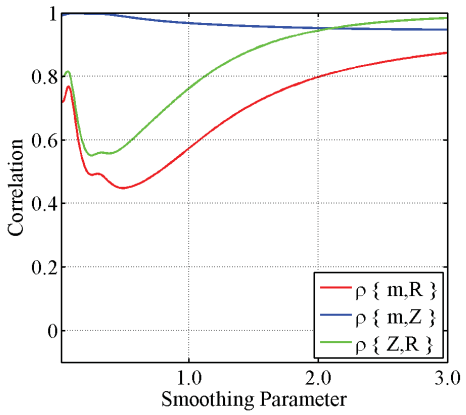


Figure 3.16: Relationships on observations, trend values and residual values

could be considered; however, some of them may have no clear geological explanations, for example, the absolute correlation between the observed values and the residual values, that is $|\rho\{Z(\boldsymbol{u}), R(\boldsymbol{u})\}|$, does not have obvious geological meanings. Moreover, although a combination of the relations, such as $\rho\{m(\boldsymbol{u}), Z(\boldsymbol{u})\} - |\rho\{m(\boldsymbol{u}), R(\boldsymbol{u})\}|$, may have a clear geological context, this combination is abandoned because the weight to each factor is difficult to determine. Thus, one promising correlation that could be considered for the trend objective function is that the trend and the residuals should be absolutely and minimally correlated, that is, min. $|\rho\{m(\boldsymbol{u}), R(\boldsymbol{u})\}|$.

The same 2-D example in Figure 3.4 is considered. The smoothing parameter SP is the only factor that will be varied with an increment of 0.01. Figure 3.17 shows the original and the absolute correlations between the trend and the residuals $\rho\{m(\boldsymbol{u}), R(\boldsymbol{u})\}$ and $|\rho\{m(\boldsymbol{u}), R(\boldsymbol{u})|\}$ that are overlapping. The best smoothing parameter corresponding to a minimized absolute correlation (red marker) is 0.52.

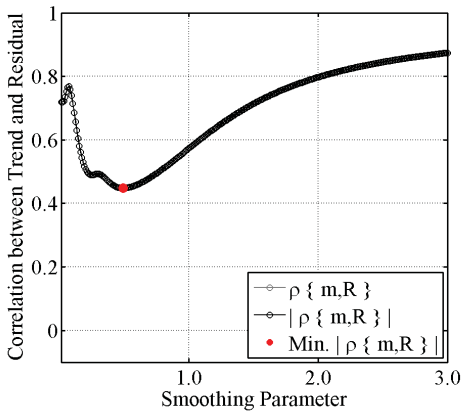


Figure 3.17: Absolute correlation between the trend values and the residual values

3.3.2 Inference of Trend Objective Function

The goal is to develop a general objective function that leads to an optimal trend. The smoothing parameter that minimizes the absolute correlation between the trend and residuals appears to provide a reasonable estimate; however, it may not reach to the final decision of the objective function. In this validation mode, the true trend is assumed known. An adjustment factor f is introduced to tune the smoothing parameter so that it determines how the factor would lead to a trend estimate close to the known trend in a wide variety of circumstances. In the future, the optimal trend could be determined with this factor when the true trend is not known.

Cross validation is considered to check the model. Figure 3.18 shows a schematic illustration for the determination of the factor f. A variety of trends are modeled with increasing smoothing parameter. The absolute correlations between the trend and residuals $|\rho\{m(\boldsymbol{u}), R(\boldsymbol{u})\}|$ are calculated and shown with the black line. The chosen smoothing parameter that minimizes this correlation is labeled as $\hat{\text{SP}}$. The mean squared errors between the trend where they

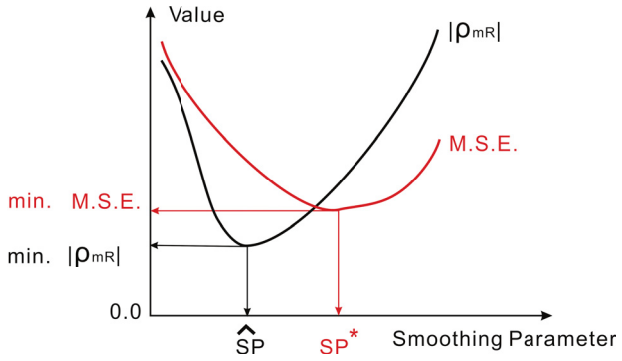
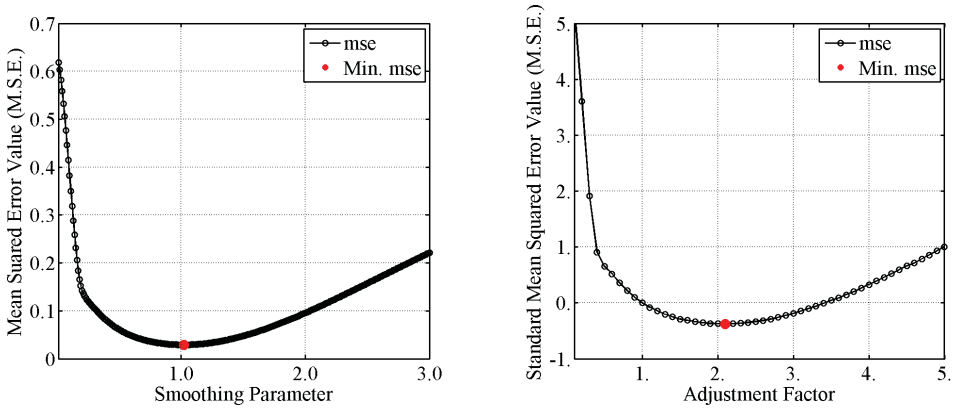


Figure 3.18: Schematic illustration of the adjustment factor in determination of a trend objective function

are available and the trend being estimated are also calculated and shown with the red line. The corresponding smoothing parameter with the minimized mean squared error comparison is labeled as SP^*. The smoothing parameter that minimizes the correlation would be, ideally, closely related to the smoothing parameter that minimizes mean squared error value. The smoothing factor for a trend model that minimizes the mean squared error has been observed somewhat larger than the one that minimizes the correlation between the trend and residuals. The adjustment factor should match the two smoothing parameters, that is, $\mathrm{SP}^* = f \times \hat{\mathrm{SP}}$.

The previous 2-D synthetic example is expanded to determine the adjustment factor f. A set of trends are modeled with increasing smoothing parameter using an increment of 0.01. Figure 3.17 shows the correlation function $|\rho\{m(\boldsymbol{u}), R(\boldsymbol{u})\}|$ and the smoothing parameter corresponding to the minimized correlation $\hat{\mathrm{SP}}$ in a reasonable condition equals to 0.52. The difference between the modeled trend and the known trend is compared by the mean squared error in Figure 3.19a and the smoothing parameter with a minimized comparison value is $\mathrm{SP}^* = 1.14$. Figure 3.19b compares the difference between these two smoothing parameters, namely $\hat{\mathrm{SP}}$ and SP^*. In order to easily compare the results, the

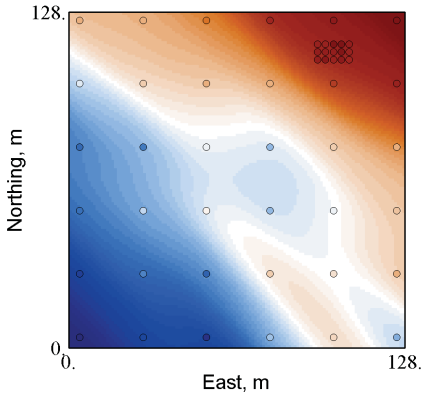


(a) MSE comparison **(b)** Adjustment factor f

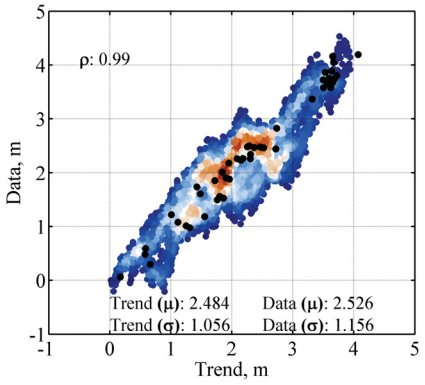
Figure 3.19: Determination of a trend objective function with smoothing parameters

mean squared error value between these smoothing parameters is standardized showing an adjustment factor close to 2.0. Figure 3.20 shows the trend models and scatter plots with different factors. The trend model appears to show too much variability with a factor of 1.0, while it appears excessive smooth and conditional biasedness with a factor of 3.0. The optimal trend with a factor of 2.0 is shown in Figure 3.20c that performs a fairly smooth and variable trend model.

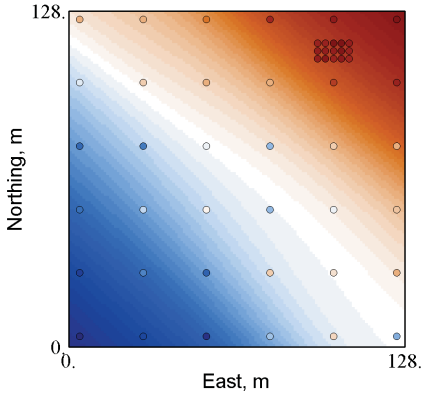
Another example with different data is also checked in Figure 3.21. The 2-D synthetic trend and residual models stay the same and shown in Figures 3.4a and 3.4b. 41 data are sampled in this exhaustive subsurface showing a N45E trend, see Figure 3.21a. The scatter plot of 41 data shows a correlation of -0.06 in Figure 3.21b. A variety of trend models are constructed with increasing smoothing parameter using an increment of 0.01. Figure 3.21c shows that a smoothing parameter of 0.56 reaches to the minimized absolute correlation, while Figure 3.21d shows the calculation based on the mean squared error comparison indicating an optimal smoothing parameter of 1.18. Figure 3.21e shows that the adjustment factor of 2.1 is the optimal. The result also confirms that



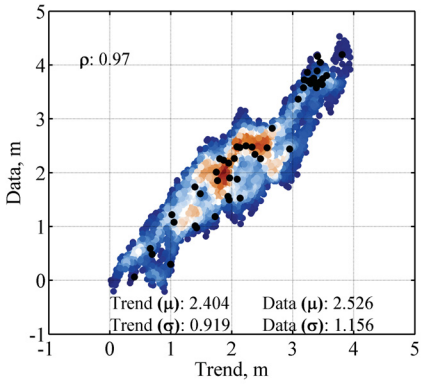
(a) Trend model with $f = 1.0$



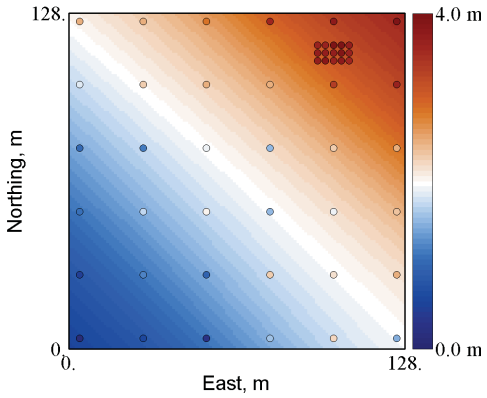
(b) Scatter plot with $f = 1.0$



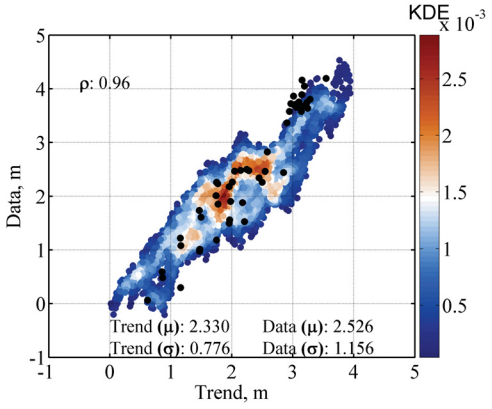
(c) Trend model with $f = 2.0$



(d) Scatter plot with $f = 2.0$

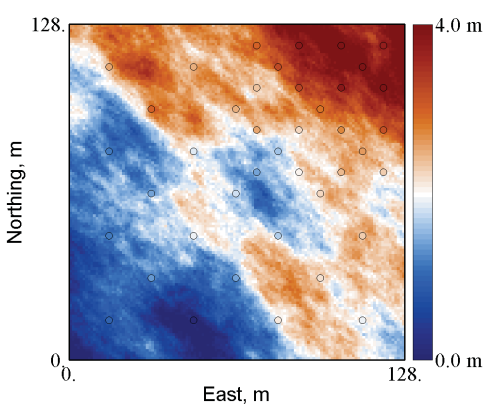


(e) Trend model with $f = 3.0$

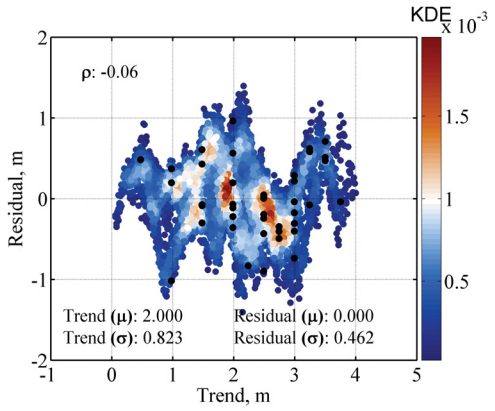


(f) Scatter plot with $f = 3.0$

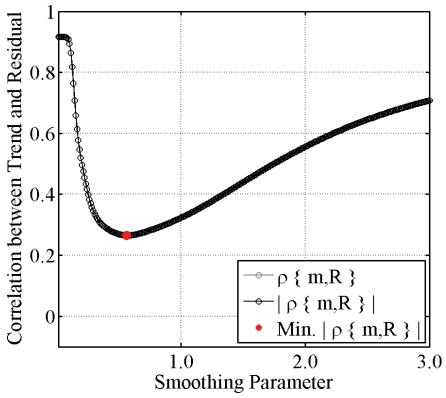
Figure 3.20: Modeled Trends and Crossplot with different adjustment factors. The scatter plots are shown in 2-D KDE units



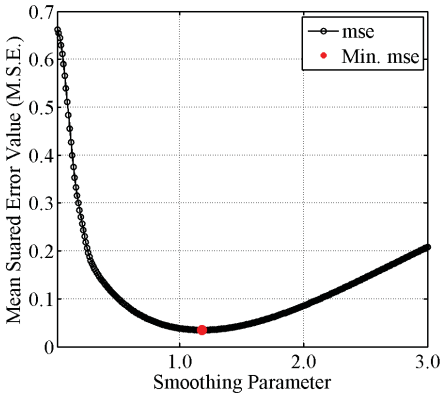
(a) True exhaustive model showing a N45E trend



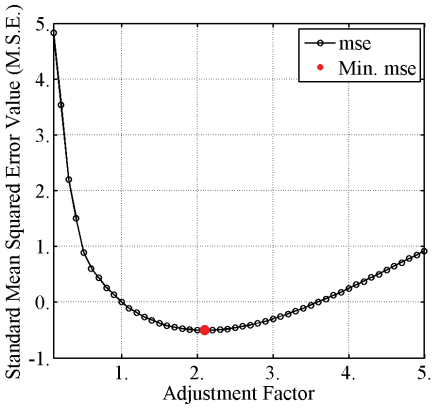
(b) Scatter plot with 41 data showing a correlation of -0.06



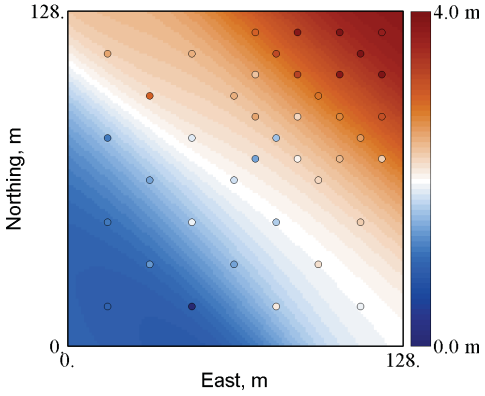
(c) Absolute correlation $|\rho_{m,R}|$



(d) MSE comparison



(e) Adjustment factor f



(f) Optimal trend model

Figure 3.21: An example for determining the adjustment factor from the correlation function. The scatter plot is shown in 2-D KDE units

an adjustment factor of 2.0 provides very good results in this example. The optimal trend model with a smoothing parameter of $2 \times 0.56 = 1.12$ is shown in Figure 3.21f.

In conclusion, the absolute correlation between the trend values and the residual values provides a good indicator of an optimal trend model. The recommended adjustment factor f of 2.0 works for a wide variety of trend modeling cases. The final objective function for trend modeling is given as:

$$\text{Objective Function} = 2.0 \cdot \text{SP when } |\rho\{m(\boldsymbol{u}), R(\boldsymbol{u})\}| \text{ is minimized} \tag{3.15}$$

Note that the adjustment factor is optional in the software. The users may change the adjustment factor based on site specific information.

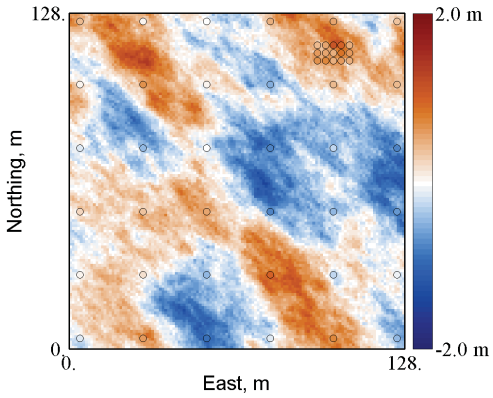
3.3.3 Considerations for the Trend Objective Function

Experience shows that, in most cases, the function $|\rho\{m(\boldsymbol{u}), R(\boldsymbol{u})\}|$ has only one minima; however, difficulties may arise when the absolute correlation between the trend and the residuals has more than one local minimum value. Several studies are tested for the assessment of the optimal trend function in Equation 3.15. Three situations are considered below.

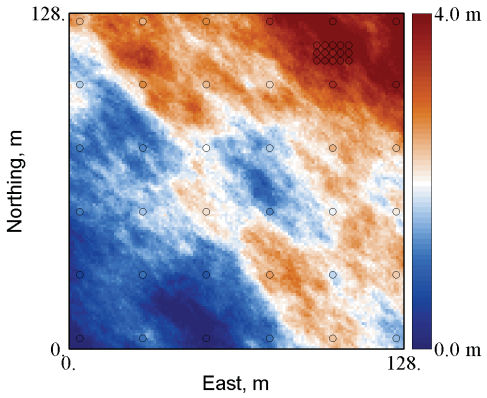
***Situation 1**: Multiple Minima in Correlation Function*

Sometimes, the absolute correlation function has more than one minima. The multiple minima might exist if the smoothing parameter is chosen in an unexpected range. Constraining the smoothing parameter in a reasonable range leads to a well behaved function.

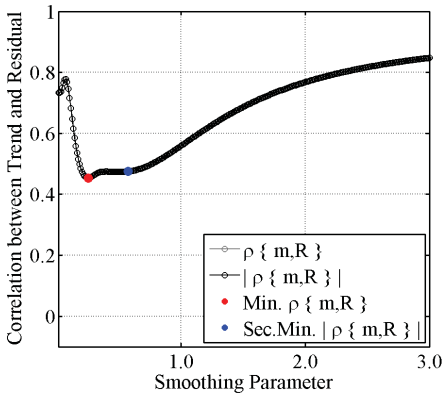
The synthetic trend model in Figures 3.4a stays the same, while a new synthetic residual model in Figure 3.22a is considered. 51 data are sampled in this exhaustive subsurface in Figure 3.22b indicating Northeast trend features. A variety of trend models are constructed with increasing smoothing parameter.



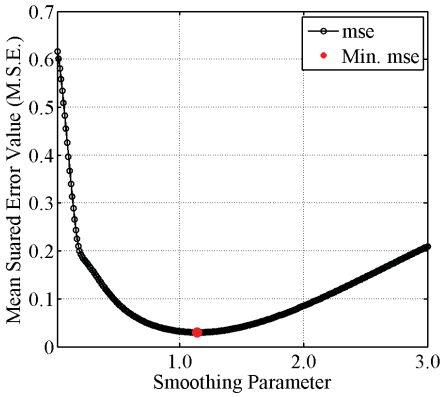
(a) True random model



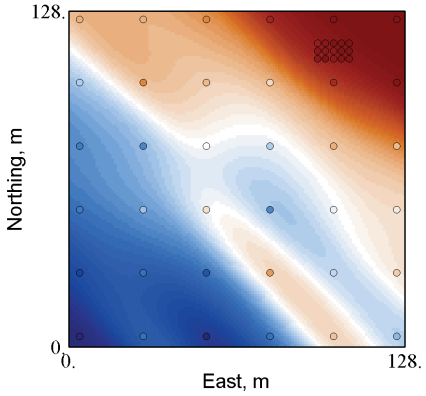
(b) True subsurface model



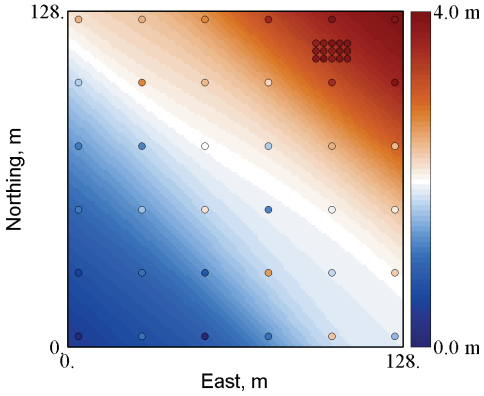
(c) Absolute correlation $|\rho_{m,R}|$



(d) MSE comparison



(e) Trend model with SP = 0.50



(f) Trend model with SP = 1.04

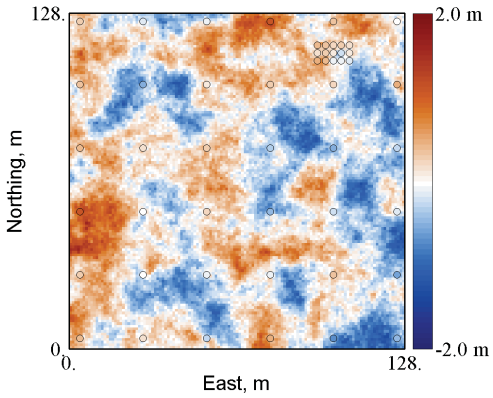
Figure 3.22: Synthetic subsurface models for determining an optimal SP in the presence of multiple minima in trend objective function. The true trend values are shown in Figure 3.4a and the true residual values are shown in Figure 3.22a. The true subsurface model in Figure 3.22b is combined with the true trend model and the true residual model. 51 black circles/markers represent the conditioning data

The absolute correlation function $|\rho\{m(\boldsymbol{u}), R(\boldsymbol{u})\}|$ is shown in Figure 3.22c. The smoothing parameter equals to SP = 0.25 when it reaches to the first minimized correlation value (red marker), and SP = 0.52 when it reaches to the second minimized value (blue marker). Figure 3.22d shows the minimized mean squared error value between the true trend and the modeled trend showing a smoothing parameter of 1.14. The optimal trend model that is obtained from Equation 3.15 is the one that doubles the smoothing parameter when the absolute correlation function first reaches to a minimum. Figure 3.22e shows the trend model with a smoothing parameter of $2 \times 0.25 = 0.50$. The trend model is not that continuous and smooth as expected through the visualization. SP = 0.50 is not a reasonable value due to the trend model appears over-fit. Figure 3.22f shows the trend model with a smoothing parameter of $2 \times 0.52 = 1.04$ showing gradually increasing values from Southwest to Northeast. The inferred trend model appears more reasonable. This result indicates that, in a circumstance of a concave absolute correlation function, the optimal smoothing parameter should be determined by the last minimized value from the correlation function.

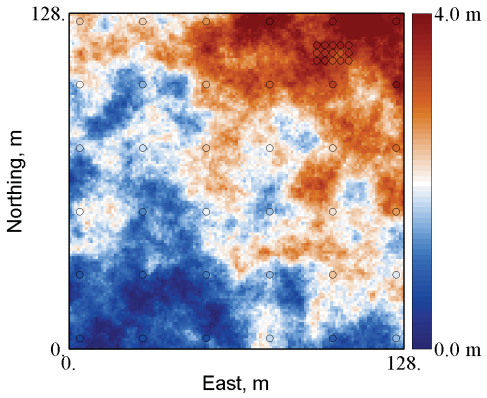
***Situation 2**: Negative Values in the Correlation Function*

The correlation between the trend and the residuals is not always positive, so an absolute correlation function $|\rho\{m(\boldsymbol{u}), R(\boldsymbol{u})\}|$ is considered.

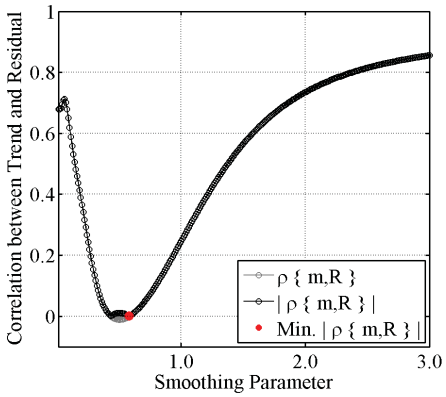
The synthetic trend in Figure 3.4a stays constant and residual models are shown in Figures 3.23a. 51 data are sampled in the exhaustive subsurface in Figure 3.23b. Northeast trend features exist. A variety of trends are modeled with increasing smoothing parameter using an increment of 0.01. Figure 3.23c shows the original correlation between the trend and the residuals with a gray line and the absolute correlation with a black line. The minimum value of the absolute correlation is the one shown with a red marker, SP = 0.58. Figure 3.23d shows the calculation based on the minimized mean squared error comparison



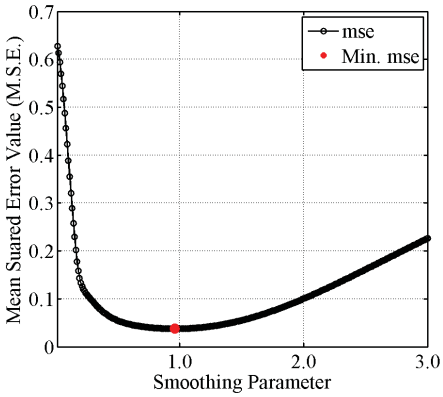
(a) True random model



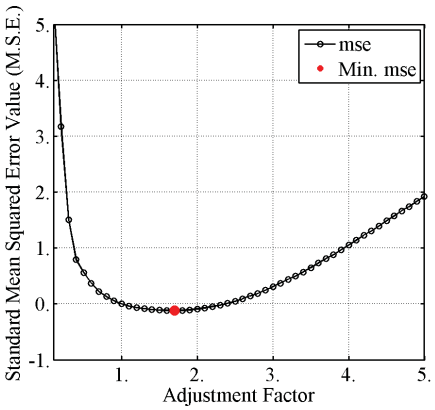
(b) True subsurface model



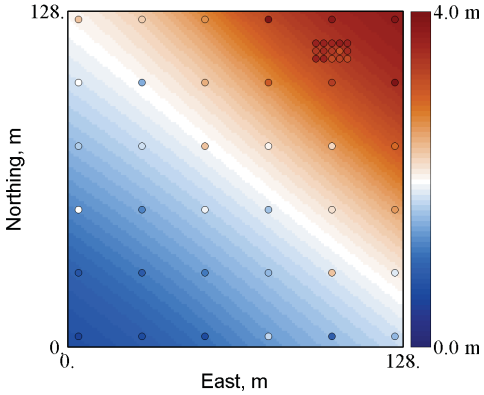
(c) Absolute correlation $|\rho_{m,R}|$



(d) MSE comparison



(e) Adjustment factor f



(f) Optimal trend model SP = 1.16

Figure 3.23: Synthetic subsurface models for determining an optimal SP in the presence of negative values in trend objective function. The true trend values are shown in Figure 3.4a and the true residual values are shown in Figure 3.23a. The true subsurface model in Figure 3.23b is combined with the true trend model and the true residual model. 51 black circles/markers represent the conditioning data

indicating an smoothing parameter of 0.96. Figure 3.23e shows that the adjustment factor of 1.70 reaches to the optimal. Still, an adjustment factor of 2.0 provides very good results in this example. The trend model with a smoothing parameter of $2 \times 0.58 = 1.16$ is shown in Figure 3.23f.

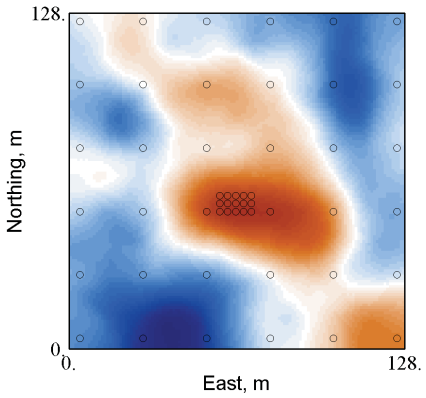
Situation 3*: Complex in the Correlation Functions*

The absolute correlation function is often shown with a single concave function; however, sometimes, it may be a combination of a convex and concave functions. Similar to the multiple minima values in ***Situation 1***, the optimal smoothing parameter should be inferred in a reasonable range.

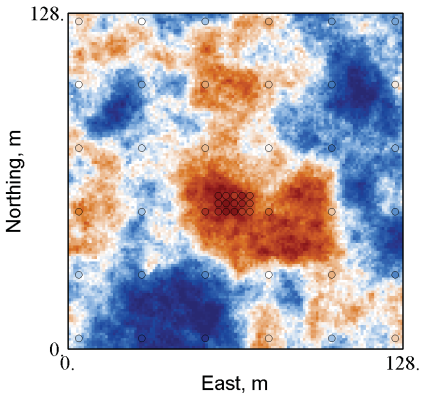
Another synthetic trend model is established in Figure 3.24a and a random model is shown in Figure 3.23a. Figure 3.24b shows the exhaustive true image and 51 samples are collected from it. More high values are concentrated in the center. The trends are modeled with increasing smoothing parameter. Figure 3.24c shows the absolute correlation between the trend and the residuals. The minimum value from this function could not be decided due to the complex function. In the inference of the absolute correlation function, only the values in the concave ranges are considered. A smoothing parameter of 0.13 reaches a minimized concave function. Figure 3.24d shows the calculation based on the mean squared error comparison indicating an optimal smoothing parameter of 0.25. Figure 3.24e shows that the adjustment factor of 1.9 reaches to the optimal. The result also confirms that an adjustment factor of 2.0 works well. The optimal trend model with a smoothing parameter of $2 \times 0.13 = 0.26$ is shown in Figure 3.24f.

3.4 Special Topic of Trend Modeling

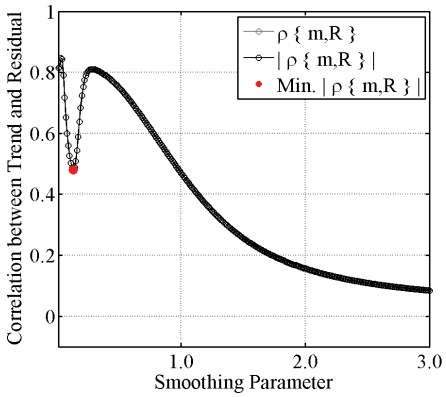
Extrapolation in trend modeling, that is, the trend predictions outside the range of the available data is a challenge and must be considered carefully. In this



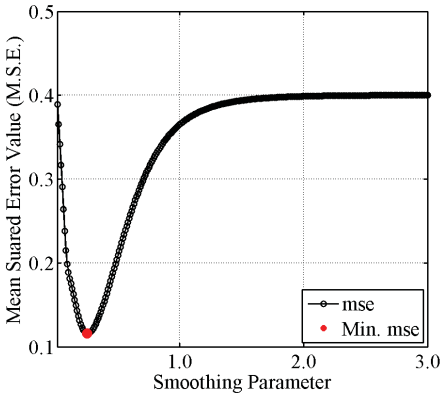
(a) True trend model



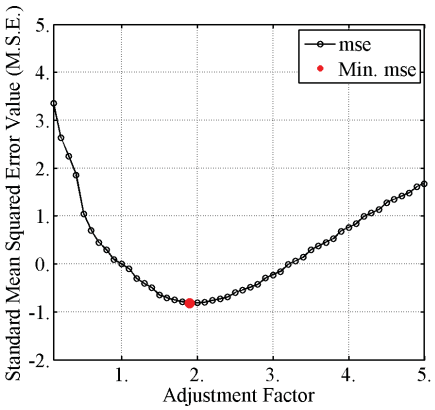
(b) True subsurface model



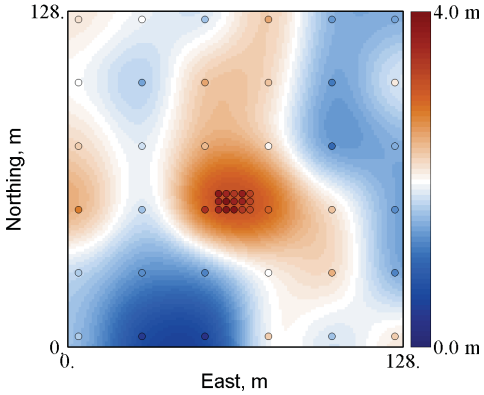
(c) Absolute correlation $|\rho_{m,R}|$



(d) MSE comparison



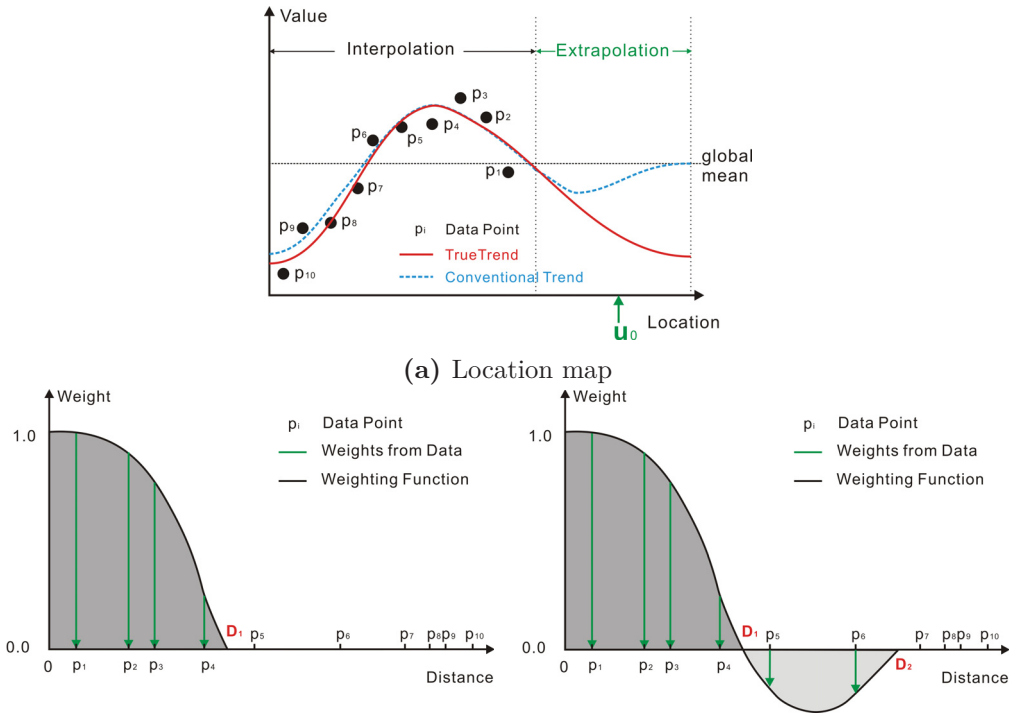
(e) Adjustment factor f



(f) Optimal trend model SP = 0.26

Figure 3.24: Synthetic subsurface models for determining the optimal SP in the presence of mix types in trend objective function. The true trend values are shown in Figure 3.24a and the true residual values are shown in Figure 3.23a. The true subsurface model in Figure 3.24b is combined with the true trend model and the true residual model. 51 black circles/markers represent the conditioning data

section, trend extrapolation is explored with a novel weighting function.



(a) Location map

(b) Conventional weighting function

(c) Proposed weighting function

Figure 3.25: Schematic illustration of the weighting functions for trend extrapolation: (***a***) location map of samples indicating some trend-like features. $\boldsymbol{u}_0$ is located outside of available data location ranges and the trend value at location $\boldsymbol{u}_0$ is required to estimate; (***b***) conventional non-negative weighting function; and (***c***) proposed weighting function suited for trend extrapolation purpose

An illustrative ease example is considered with the proposed approach. Figure 3.25a shows a 1-D schematic illustration. The true trend is shown with a red line. 10 black markers are known data and labeled with numbers from p_1 to p_{10}. These data perform a strong trend in the interpolation region, while the trend value at the location of $\boldsymbol{u}_0$ is being estimated in the extrapolation region. Figure 3.25b shows the conventional weighting function in Equation 3.2 when four data (p_1 to p_4) are incorporated into the weight calculation. More weight is given to the closest data and the less weight is given to the further data.

The weight becomes almost zero or a small background value when the sampled location is far away from the modeled location. Then, the weight is combined with the data values to construct a trend model. The modeled trend is shown with a blue line in Figure 3.25a. The trend value goes to a constant value where the estimate is away from the given samples.

Care should be taken when modeling a trend in a region beyond the range of the data values. Three basic steps should be followed when considering the trend extrapolation. The first is to assemble the available samples, such as the grade values and define the extrapolation boundary for the regionalized variable. The second step is to choose a proper weighting function and the appropriate parameters for the trend model. The choice of the weighting function provides a basis for constructing a reasonable trend model in the extrapolation region. The final step is to use the proposed mathematical function and the estimated parameters to infer the trend model.

A novel weighting function is proposed where nearby data are given more positive weights and the distant data are given progressively negative weight. Figure 3.25c shows a sketch of the developed weighting function suited for the trend extrapolation. Four data from p_1 to p_4 are assigned with the positive weights; two following data p_5 and p_6 are given with the negative weights; and rest of data, from p_7 to p_{10}, are useless. Distance D_1 is a transition point from the positive weight to the negative weight, and D_2 is the distance between the data with the last negative weight and the next unused data. Two quadratic functions are considered to define the new weighting scheme. $\{f(\boldsymbol{u}_i), i = 0, 1, \cdots, np\}$ represent the positive weights and $\{g(\boldsymbol{u}_j), j = 0, 1, \cdots, nn\}$ represent the negative weights that are given as:

$$\begin{cases} f(\boldsymbol{u}_i) & \text{if } 0 \leqslant \boldsymbol{u}_i \leqslant D_1 \\ g(\boldsymbol{u}_j) & \text{if } D_1 \leqslant \boldsymbol{u}_j \leqslant D_2 \end{cases} \quad \boldsymbol{u}_i, \boldsymbol{u}_j \in A \tag{3.16}$$

In a condition of

$$np + nn \leqslant n$$

here n is the total number of available samples within the domain A. np represents the number of data that are given with the positive weights, while nn represents the number of data that are given with the negative weights. Distances D_1 and D_2 are the transition points from the positive weight to the negative weight and from the last negative weight to the following zero weight, respectively. The derivations of the weighting functions are appended in Appendix A. The final weighting functions are written as:

$$\begin{cases} f(\boldsymbol{u}_i) & = 1 - \frac{1}{D_1^2} \cdot \boldsymbol{u}_i^2 \\ g(\boldsymbol{u}_j) & = \frac{6}{(D_2 - D_1)^3} \cdot \Delta \cdot D_1 \cdot D_2 - \frac{6}{(D_2 - D_1)^3} \cdot \Delta \cdot (D_2 + D_1) \cdot \boldsymbol{u}_j + \frac{6}{(D_2 - D_1)^3} \cdot \Delta \cdot \boldsymbol{u}_j^2 \end{cases}$$
$$\boldsymbol{u}_i, \boldsymbol{u}_j \in A \tag{3.17}$$

here the transition point D_1 can be expressed by a simple average distance between point 4 and point 5 in Figure 3.25c, that is, $\frac{D_{\mathrm{p}_4} + D_{\mathrm{p}_5}}{2}$. The distance D_2 could also be easily calculated between point 6 and point 7, that is, $\frac{D_{\mathrm{p}_6} + D_{\mathrm{p}_7}}{2}$. Δ is a factor that accounts for the sum of the negative weights. The equation is given as:

$$\sum_{j=1}^{nn} g(\boldsymbol{u}_j) = -\Delta \tag{3.18}$$

The factor Δ can be inferred by different functions. Here, a linear increment is considered. The sums of the positive weights and the negative weights should be normalized to 1.0, and then, the total positive weights can be given as:

$$\sum_{i=1}^{np} f(\boldsymbol{u}_i) = 1 + \Delta \tag{3.19}$$

Thus, the equation for the trend value $m(\boldsymbol{u})$ using the moving average statistics is expressed by:

$$m^*(\boldsymbol{u}) = \frac{\sum_{i=1}^{\mathrm{np}} f(\boldsymbol{u}_i) \cdot z(\boldsymbol{u}_i)}{\sum_{i=1}^{\mathrm{np}} f(\boldsymbol{u}_i)} + \frac{\sum_{j=1}^{\mathrm{nn}} g(\boldsymbol{u}_j) \cdot z(\boldsymbol{u}_j)}{\sum_{j=1}^{\mathrm{nn}} g(\boldsymbol{u}_j)} \quad \boldsymbol{u}, \boldsymbol{u}_i, \boldsymbol{u}_j \in A \tag{3.20}$$

where $\{z(\boldsymbol{u}_i), i = 1, 2, \cdots, np\}$ and $\{z(\boldsymbol{u}_j), j = 1, 2, \cdots, nn\}$ denote the observed value at the $\boldsymbol{u}_i$ and $\boldsymbol{u}_j$ location within the domain A, respectably. Weights are denoted by the functions $\{f(\boldsymbol{u}_i), a = 1, 2, \cdots, np\}$ and $\{g(\boldsymbol{u}_j), b = 1, 2, \cdots, nn\}$. np represents a set of data that are given with the positive weights, and nn represents a set of data that are given with the negative weights. Note that the used number of data could not exceed the total available samples n, that is, $np + nn \leqslant n$.

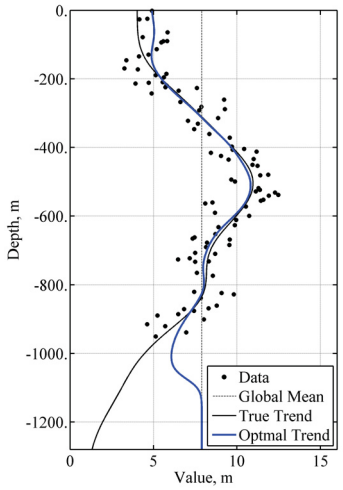
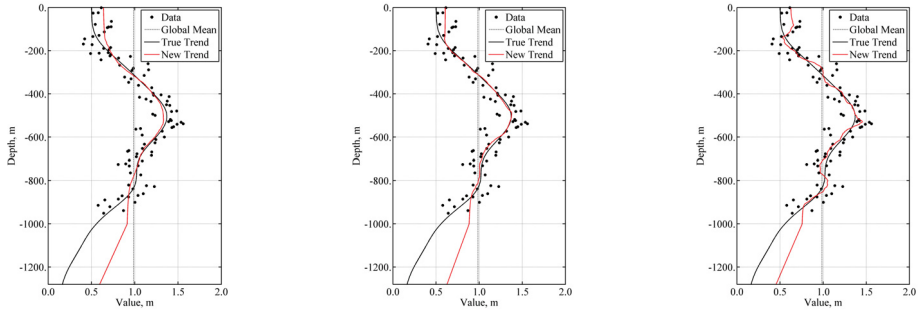


Figure 3.26: Problem with current proposed trend modeling function in Equation 3.15 indicating the importance of an appropriate trend function suited with trend extrapolation. Several data are collected in a shallow depth. The true trend (black line) is considered as a reference indicating a reducing value as the depth lower than 1000 meter. The optimal trend (blue line) is modeled with corrected trend parameters in a objective fashion. The trend values increases as the depth showing the importance of the trend extrapolation

A 1-D synthetic example is shown in Figure 3.26 for demonstrating the proposed weighting function in trend extrapolation. 96 samples with black markers are collected on 1×1280 grid nodes and they are preferably sampled in the shallow depth from 0 meters to 1000 meters and no data are sampled from 1000 meters to 1280 meters. The true trend is shown as a black line. The gray dashed line represents the global mean from 96 data with declustering weights. Trends are modeled with moving averages in Equation 3.1 with the corrected parameters mentioned in previous sections. The optimal trend is chosen by Equation 3.15 and the optimal trend is shown with a blue line in Figure 3.26. This modeled trend looks reasonable at the shallow depth where there are lots of data; however, the trend model goes back to the global mean at the depth in extrapolation; this does not appear reasonable.

1000 meters is considered as a transition point from the interpolation estimate to the extrapolation estimate. All trends are modeled with the proposed weighting function using the closest 40 data. Figure 3.27a represents the modeled trend using the nearest 10 data with positive weights and the following 30 data with negative weights. Figure 3.27b shows the modeled trend using the closest 20 data with positive weights and the following 20 data with negative weights. Figure 3.27c represents the modeled trend using the nearest 30 data with positive weights and the following 10 data with negative weights. These modeled trends show reasonable extrapolations; however, the models contain some variabilities in the interpolation region due to the unsmoothness proposed weighting function in Equation 3.17 and obvious artifacts at the transition point at the location of 1000 meters due to the linear function of the indicator Δ in Equation 3.18.

Weights at different locations should be checked. The weights at -500.50 meters (trend interpolation) and -1100.50 meters (trend extrapolation) are shown in Figure 3.28. The left figure shows the probability distributions of the weights with 40 data, and the right figure shows the cumulative distributions

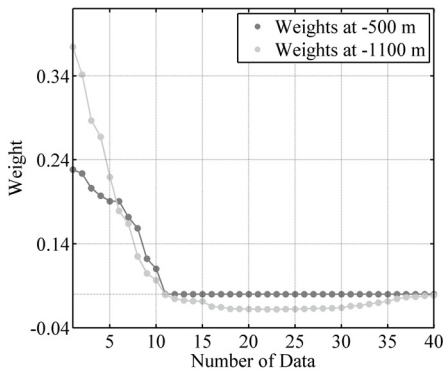


(a) $np = 10, nn = 30$ **(b)** $np = 20, nn = 20$ **(c)** $np = 30, nn = 10$

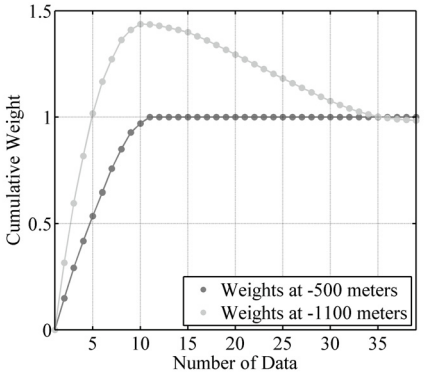
Figure 3.27: Modeled trends with proposed trend extrapolation function with different number of Data

of the weights with these data. Similar to the conventional weighting function, there are no negative weights at the location of -500.50 meters since data are more informed in this location. Unlike the conventional weighting function, the positive weights and negative weights are both considered and used at the location of -1100.50 meters. The total weights are then scaled to a unit sum.

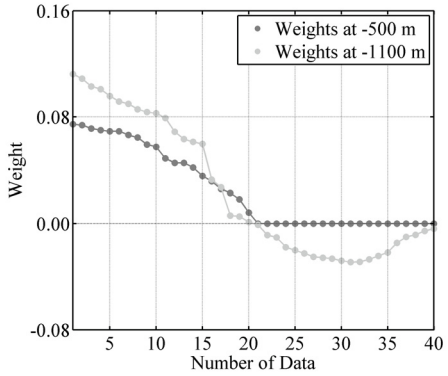
This trend extrapolation method suffers from several shortcomings. The proposed weighting algorithm does not account for the differences in the geological processes. Consequently, the model with the trend extrapolation may not be stable and realistic. Additional resources, such as geological mapping and seismic data, may be helpful to model a realistic trend model. Additionally, the proposed objective function with the correlation does not work for the extrapolation. Visualization by practitioners may be the only choice to determine the optimal trend model. Further, the discontinuities in the weighting function lead to discontinuities in the trend predictions, so does the indicator. A more smooth and continuous weighting function and other types of the indicator should be considered in future research.



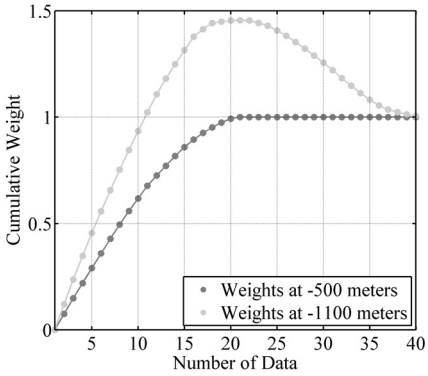
(a) $np = 10, nn = 30$



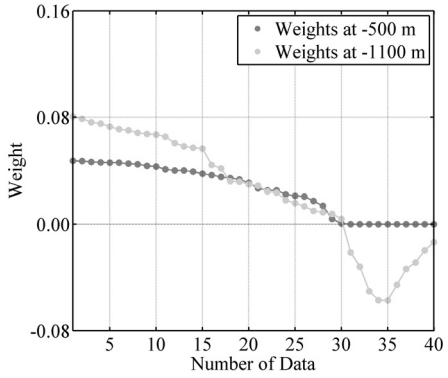
(b) $np = 10, nn = 30$



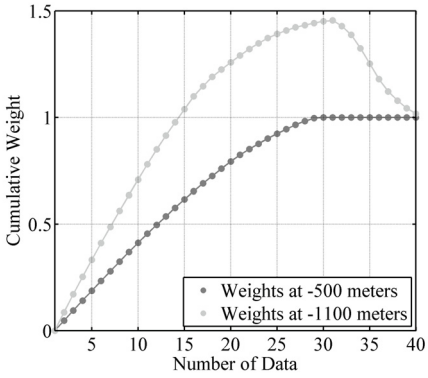
(c) $np = 20, nn = 20$



(d) $np = 20, nn = 20$



(e) $np = 30, nn = 10$



(f) $np = 30, nn = 10$

Figure 3.28: PDFs and CDFs of the weights using the closest 40 Data at different estimate locations.

3.5 Remarks

This chapter emphasizes the model of a deterministic trend and stochastic fluctuation of regionalized variables. The trend model characterizes the large-scale physical features and the residual model captures the short-scale variations with limited data. The optimal selections of parameters for trend modeling including the background value, anisotropy, declustering weights and smoothing parameter are demonstrated with examples. The deterministic trend can be calculated in an objective and repeatable manner.

The smoothing parameter has been calibrated with synthetic examples where a reference trend model is available and where cross validation could be used to validate the objective function. Considering a smoothing parameter that is double that obtained by minimizing the absolute correlation between the trend and residuals ensures a reasonable trend model. This objective function is typically simple to understand and easy to implement.

Although there is a gap between the theory and the practice in the inference of the optimal trend function, the practical success of the objective function is shown. In practice, this objective trend function is well behaved; however, there is a lack of the theoretical understanding why the algorithm on trend modeling works well. The mathematical proof for the algorithm should be considered in future work.

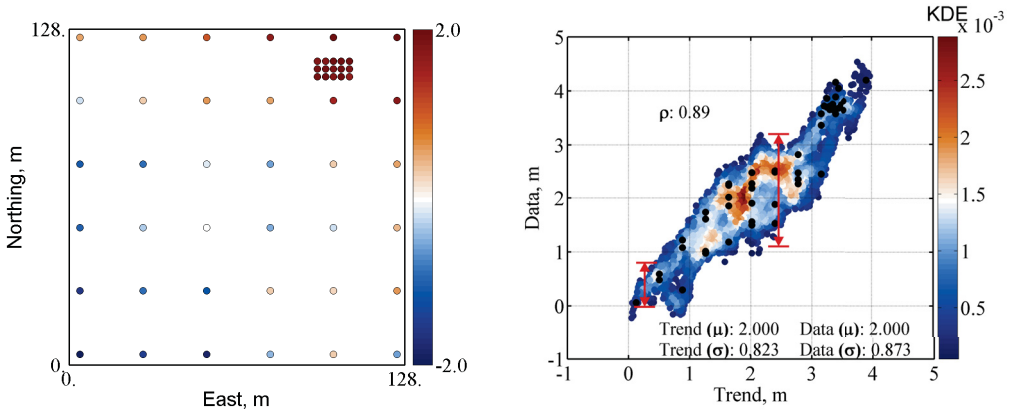
Trend extrapolation is always a challenge and must be considered in 2-D and 3-D cases. A preliminary approach is proposed to infer the trend model in extrapolation. The proposed weighting function considers positive and negative weights. The result provides a useful starting point for future research in trend extrapolation.

Chapter 4

Prediction with a Trend

The conventional procedure for simulation is to transform the regionalized variable into a Gaussian distribution. The regionalized variable is forced to be univariate Gaussian after the normal score transformation; however, the trend-like features are not removed in the normal score transformation, see Figure 4.1a. The relationship between the original values and the trend values, see Figure 4.1b, will not likely be preserved in simulation. Trend features must be considered and enforced properly in geostatistical simulation.

This chapter develops a technique for transforming the values and the trend model to be multivariate Gaussian. The proposed method will account for the trend features in the regionalized variable, reduce the uncertainty of the models and improve the characterization of the variability. The original stepwise conditional transformation proposed by Leuangthong and Deutsch (2003) and the Gaussian mixture model proposed by McLachlan and Peel (2000) are reviewed. A methodology similar to this non-parametric stepwise conditional transformation is proposed. The conditional distribution is calculated by Gaussian mixture models fitted to the deterministic trend and the data. The trend-like features in the regionalized variable are removed by the conditional transformation. Details



(a) Location Map from 51 transformed data in NS units

(b) Crossplot of the trend model with 51 data in original units

Figure 4.1: Location map from 51 transformed data in NS units and crossplot of the trend model with 51 data in original units. The location map shows obvious trend features that high values are located in the Northeast and low values are in the Southwest. The crossplot is shown on 2-D KDE units. Data in the presence of a trend show heteroscedasticity features.

of the proposed geostatistical modeling algorithm are described and illustrated with examples where the values show an apparent trend.

4.1 Review of Conditional Transformation

The original stepwise conditional transformation and Gaussian mixture models are outlined. The advantages and drawbacks are pointed out.

4.1.1 Stepwise Conditional Transformation

The stepwise conditional transformation technique proposed by Leuangthong (2003) provides an alternative to the conventional normal score transformation. This transformation captures non-linear features in the multivariate distributions. The stepwise conditional transformation is identical to the normal score

transformation for the first variable; subsequent variables are normal score transformed based on the previous variables. All variables are uncorrelated and independent at a zero lag distances after the transformation. The back transformation brings all non-linear features back to the original distribution and all correlations between multivariate variables are preserved.

Consider $\{Z(\boldsymbol{u}_{i,k}), i = 1, \cdots, n \text{ and } k = 1, \cdots, K\}$ is a set of K stationary random functions at a set of n data locations in the domain A. The observations of the random function at the location $\{\boldsymbol{u}_i, i = 1, \cdots, n\}$ are denoted as $\{\boldsymbol{z}_{i,1}, \cdots, \boldsymbol{z}_{i,K}\}$. The first variable $\{\boldsymbol{z}_{i,1}, i = 1, \cdots, n\}$ is transformed independently to Gaussian units. The second variable $\{\boldsymbol{z}_{i,2}, i = 1, \cdots, n\}$ is transformed conditionally to the probability classes of the first variable $\{\boldsymbol{z}_{i,1}, i = 1, \cdots, n\}$. The $\left\{k^{\text{th}}, k = 1, \cdots, K\right\}$ variable $\{\boldsymbol{z}_{i,k}, i = 1, \cdots, n\}$ is partitioned to the probability classes of the previous $(k-1)$ variables, and so on. The process is given as:

$$
\begin{aligned}
\boldsymbol{y}_{i,1} &= G^{-1}\left(F_1\left(\boldsymbol{z}_{i,1}\right)\right) \\
\boldsymbol{y}_{i,2} &= G^{-1}\left(F_{2|1}\left(\boldsymbol{z}_{i,2} \mid \boldsymbol{z}_{i,1}\right)\right) \\
&\vdots \\
\boldsymbol{y}_{i,k} &= G^{-1}\left(F_{k|1,\cdots,k-1}\left(\boldsymbol{z}_{i,k} \mid \boldsymbol{z}_{i,1}, \cdots, \boldsymbol{z}_{i,k-1}\right)\right) \\
&\vdots \\
\boldsymbol{y}_{i,K} &= G^{-1}\left(F_{K|1,\cdots,K-1}\left(\boldsymbol{z}_{i,K} \mid \boldsymbol{z}_{i,1}, \cdots, \boldsymbol{z}_{i,K-1}\right)\right) \quad i = 1, \cdots, n
\end{aligned}
\tag{4.1}
$$

here $\{\boldsymbol{y}_{i,k}, i = 1, \cdots, n \text{ and } k = 1, \cdots, K\}$ are the transformed multivariate Gaussian variables that are uncorrelated and independent. $G^{-1}(\cdot)$ represents the inverse Gaussian CDF and $F(\cdot)$ indicates a CDF derived from the original data. The co-located complex features are removed in this forward transformation.

All variables are in the standard Gaussian units after the stepwise conditional transformation. The transformed variables are simulated independently. The stepwise back transformation brings all the complexities back into original units by reversing the order of the forward transformation, that is,

$$\begin{aligned}
\boldsymbol{z}_{i,K} &= F^{-1}\left(G_{K|1,\cdots,K-1}\left(\boldsymbol{y}_{i,K} \mid \boldsymbol{y}_{i,1},\cdots,\boldsymbol{y}_{i,K-1}\right)\right) \\
&\vdots \\
\boldsymbol{z}_{i,k} &= F^{-1}\left(G_{k|1,\cdots,k-1}\left(\boldsymbol{y}_{i,k} \mid \boldsymbol{y}_{i,1},\cdots,\boldsymbol{y}_{i,k-1}\right)\right) \\
&\vdots \\
\boldsymbol{z}_{i,2} &= F^{-1}\left(G_{2|1}\left(\boldsymbol{y}_{i,2} \mid \boldsymbol{y}_{i,1}\right)\right) \\
\boldsymbol{z}_{i,1} &= F^{-1}\left(G_1\left(\boldsymbol{y}_{i,1}\right)\right) \quad i = 1,\cdots,N
\end{aligned} \tag{4.2}$$

where N represents the number of the grid nodes where typically $n < N$. The complex features are restored after the back transformation and the primary correlations between variables are re-introduced.

The original stepwise proposal assumes that the correlations for all lag distances $\boldsymbol{h}$ are removed; however, only features at $\boldsymbol{h} = 0$ are removed and all multivariate distributions are in the Gaussian shapes at $\boldsymbol{h} = 0$. There is no guarantee of de-correlation at non-zero lag distances.

The order of the stepwise transformation is important. The derivation of the conditional distribution becomes increasingly difficult as k increases. The conditional distribution for the first variable is well established; and the accuracy will decrease with the following variables due to a decreasing number of data will be used to infer the conditional distribution. Choosing the most continuous variable to simulate first works well in practice.

The original stepwise proposal considers the non-parametric conditional distributions for the transformation. This approach, however, suffers from artifacts due to the bins used for the conditional distributions. There must be

enough data to inform the conditional probabilities for the transformation in each bin. Each bin requires a minimum of 10 observations for each variable. The required total number of data would be 10^K where K is the number of variables. There are rarely sufficient data to reliably identify the conditional distributions with $K > 3$ variables. Leuangthong (2003) proposed a nested application to reduce the SCT data requirement; however, the binning nature is still shown.

4.1.2 Gaussian Mixture Model Fitting

The Gaussian distribution is fully parameterized by a mean vector and a covariance matrix. One single Gaussian model cannot capture all the complex features of geologic data, while one Gaussian kernel per observation is computationally expensive with a large number of data (Parzen, 1962; Rosenblatt, 1956; Silverman, 1986; Gray and Moore, 2003). Mixture models with a small number of Gaussian kernels could be considered for simplifying the calculation.

Pearson (1894) proposed the underlying theory of the mixture models. A number of authors including Gilardi et al. (2002) and Silva and Deutsch (2016) have used them into geostatistics. The benefits of the Gaussian mixture model (GMM) are that the complex features can be captured by Gaussian components and any conditional distribution can be easily calculated.

The Gaussian mixture model is a probabilistic model and assumes all data can be generated from a mixture of a finite number of Gaussian components with unknown parameters. Consider a set of K dimensional variables at n data locations $\boldsymbol{z} = \{\boldsymbol{z}_1^T, \cdots, \boldsymbol{z}_n^T\}$ in the domain A where the superscript T denotes the vector transpose. $\boldsymbol{y} = \{\boldsymbol{y}_1^T, \cdots, \boldsymbol{y}_n^T\}$ are the set of the normal score transformed

variables where each variable is transformed independently. The Gaussian mixture model is a multivariate probability density function. Such probability density function is written as a sum of g components or mixtures:

$$f^{'}\left(\boldsymbol{y}_i \mid \boldsymbol{\Psi}\right) = \sum_{j=1}^{g} \pi_j \phi\left(\boldsymbol{y}_i \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\right) \quad i = 1, \cdots, n \tag{4.3}$$

here $f^{'}(\cdot)$ is the estimated probability density distribution. $\boldsymbol{\Psi}$ is a set of unknown parameters $\{\pi_1, \cdots, \pi_g; \boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_g; \boldsymbol{\Sigma}_1, \cdots, \boldsymbol{\Sigma}_g\}$ in the mixture model and would be determined. $\{\pi_1, \cdots, \pi_g\}$ are the non-negative mixing proportions/weights assigned to each mixture:

$$0 \leqslant \pi_j \leqslant 1 \quad j = 1, \cdots, g$$

and

$$\sum_{j=1}^{g} \pi_j = 1 \tag{4.4}$$

$\phi(\cdot)$ is the probability density function. $\{\phi\left(\boldsymbol{y}_i \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\right), i = 1, \cdots, n \text{ and } j = 1 \cdots, g\}$ represent the component densities of the normal mixtures. $\{\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_g\}$ indicate the mean vectors of all variables and $\{\boldsymbol{\Sigma}_1, \cdots, \boldsymbol{\Sigma}_g\}$ refer to a set of covariance matrices between variables for each mixture.

4.1.2.1 Expectation Maximization

The Expectation Maximization (EM) algorithm is often considered to determine the parameters of the mixture models. The approach of the EM technique was well explained by Dempster et al. (1977). The EM technique involves two steps that are the Expectation step (E-step) and the Maximization step (M-step). The expectation of the log likelihood with the current estimate parameters is built in the E-step. In the M-step, these parameters are re-estimated by maximizing

the generated expectation of the log likelihood in the E-step. EM is an iterative algorithm. The log likelihood of the data increases with each iteration and approaches a local maximum.

The log likelihood $\log\{L(\boldsymbol{\Psi})\}$ for a set of unknown parameters $\boldsymbol{\Psi}$ can be formed from the observed data $\{\boldsymbol{y}_i, i = 1, \cdots, n\}$ and given as:

$$\begin{aligned} \log\{L(\boldsymbol{\Psi})\} &= \sum_{i=1}^{n} \log f^{'}(\boldsymbol{y}_i \mid \boldsymbol{\Psi}) \\ &= \sum_{i=1}^{n} \log \left\{ \sum_{j=1}^{g} \pi_j \phi(\boldsymbol{y}_i \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \right\} \end{aligned} \tag{4.5}$$

The parameters of the mixtures would be iteratively fitted so that $f^{'}(\cdot)$ closely fits the experimental data. In this context, any conditional distribution can be easily obtained once the function $f^{'}(\cdot)$ is fit. The means, covariance matrices and the mixing proportions for each component should be determined so that the log likelihood $\log\{L(\boldsymbol{\Psi})\}$ in Equation 4.5 is solved in a reasonable way. The procedure in EM algorithm is described as follow.

Consider a set of K-dimensional n observed Gaussian data $\boldsymbol{y} = \left\{\boldsymbol{y}_1^T, \cdots, \boldsymbol{y}_n^T\right\}^T$. A set of n unobserved vectors $\boldsymbol{x} = \left\{\boldsymbol{x}_1^T, \cdots, \boldsymbol{x}_n^T\right\}^T$ is introduced and considered as component-label variables which contain the additional information. If $\{\boldsymbol{y}_i, i = 1, \cdots, n\}$ belongs to the $\left\{j^{\text{th}}, j = 1, \cdots, g\right\}$ component of the mixtures, then the component $\{x_{j,i} = (\boldsymbol{x}_i)_j, i = 1, \cdots, n \text{ and } j = 1, \cdots, g\}$ equals to 1. Otherwise, $\{x_{j,i} = (\boldsymbol{x}_i)_j = 0, i = 1, \cdots, n \text{ and } j = 1, \cdots, g\}$ is satisfied.

E-Step

The E-step initializes the unknown parameters $\boldsymbol{\Psi}^{(0)}$ and calculates the unknown labels from the unobserved data $\{x_{j,i}, j = 1, \cdots, g \text{ and } i = 1, \cdots, n\}$ on the l^{th} iteration.

The initializations of the mixing proportions, means and covariance matrices $\left\{\boldsymbol{\Psi}^{(0)} = \left(\pi_j^{(0)}, \boldsymbol{\mu}_j^{(0)}, \boldsymbol{\Sigma}_j^{(0)}\right), j = 1, \cdots, g\right\}$ for each component are required before the iteration. The initial parameters are chosen randomly. K-means++ algorithm proposed by Arthur and Vassilvitskii (2007) is considered for the initialization. The observed data $\{\boldsymbol{y}_i, i = 1, \cdots, n\}$ are initially partitioned into g components.

The unobserved data $\left\{x_{j,i}^{(0)}, j = 1, \cdots, g \text{ and } i = 1, \cdots, n\right\}$ represents the probability if the observed samples $\{\boldsymbol{y}_i, i = 1, \cdots, n\}$ belonging to the j^{th} component of the mixtures on the initial iteration. The initial conditional expectation of the random variable given the observed data $x_{j,i}^{(0)} = E_{\boldsymbol{\Psi}^{(0)}}\left(X_{j,i} \mid \boldsymbol{y}_i\right)$ using the current the parameters $\left\{\boldsymbol{\Psi}^{(0)}\right\}$ is estimated. The unobserved data continues to be calculated given the current mixing proportions, means and covariance matrices estimation for each component.

M-Step

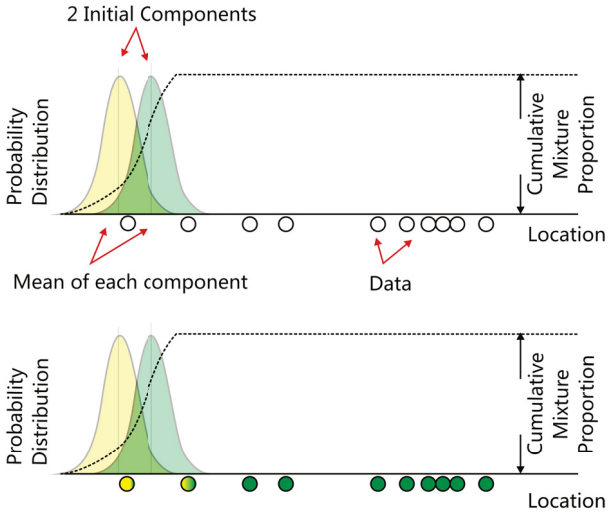
The process in the M-step requires the global maximization of the log likelihood with respect to the parameter $\boldsymbol{\Psi}$ on the $(l+1)^{\text{th}}$ iteration.

The set of parameters $\left\{\Psi^{(l+1)}, l = 0, 1, 2, \cdots, \right\}$ for each component are re-estimated based on the just-computed unobserved data by maximizing the log likelihood in Equation 4.5 on the $(l+1)^{\text{th}}$ iteration. The mixing weights, mean vectors ad covariance matrices in this set of parameters could be calculated by collecting the observed data according to the unobserved data and proportionally averaging the observed values.

The convergence of Gaussian mixtures is considered by the calculation of the log likelihood after each iteration. The log likelihood value is a non-decreasing function after the E- and M-iterations (Dempster et al., 1977). The E- and M-steps are repeated until

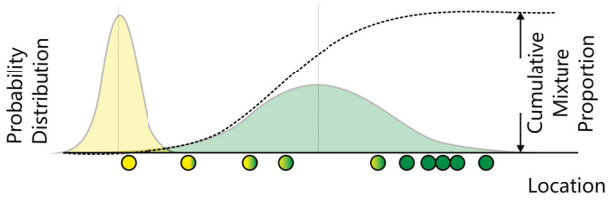
$$L\left(\boldsymbol{\Psi}^{(l+1)}\right) - L\left(\boldsymbol{\Psi}^{(l)}\right) \quad l = 0, 1, \cdots$$

(a) Initial a set parameters Ψ (mixture proportions, means and covariance matrics) and partition the observed data in the E-step.



(b) Calculate the probability of each unobserved data given the current set of parameters in the E-step.

(c) Calculate a better parameters Ψ based on just-computed unobserved data by maximizing the log likelihood function over all possible unobserved data in the M-step.



(d) Iterate E-step and M-step until convergence.

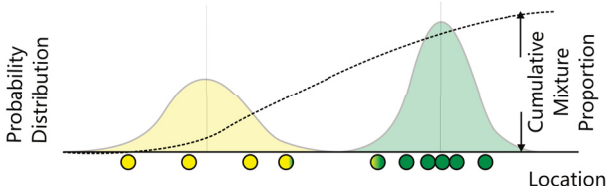


Figure 4.2: Schematic illustration of the EM algorithm: (***a***) initial the unknown set of parameters $\boldsymbol{\Psi}$ to some random values; (***b***) calculate the probability for each unobserved data given the unknown parameters $\boldsymbol{\Psi}$; (***c***) re-calculate the parameters $\boldsymbol{\Psi}$ given the unobserved data in step ***b***; and (***d***) iterate the steps ***b*** and ***c*** until convergence (Modified from Lavrenko (2014))

is arbitrarily small indicating convergence of the sequence of the log likelihood value reaches a local optimum. Figure 4.2 shows a schematic illustration of the EM algorithm.

4.1.2.2 Expectation Maximization with Exhaustive Values

Data should be conditionally transformed by a representative trend distribution. This conditional transformation requires the use of a bivariate distribution of the data and its trend model. Missing values in the trend model make the EM algorithm questionable to apply.

Figure 4.3 illustrates the bivariate distribution between the data and it trend model in original units. The black lines represent the distributions at sampled locations, while the red lines represent the true/exhaustive distributions. Note that the exhaustive trend model is not the same as the true trend model. The exhaustive trend model is constructed based on observed data and selected in an objective manner, while the true trend model is only available in synthetic examples and it is hard to know in real cases. The distribution of the co-located trend (black line) is different with that of the exhaustive trend (red line). If the co-located trend values are used for inferring the conditional transformation, some important features at unsampled locations are ignored and it may introduce some artifacts. Missing trend values should be accounted for in Gaussian mixture models.

EM algorithm can handle the missing values accounting for the missing values in an imputation mechanism (Dempster et al., 1977). The multivariate probabilities should be marginalized by the observed data and the exhaustive trend model in the E-step. The conditional expectations of the missing components given the observed data should be included in each Gaussian component. In the M-step, the means and the covariance matrices in a set of parameters $\mathbf{\Psi}$

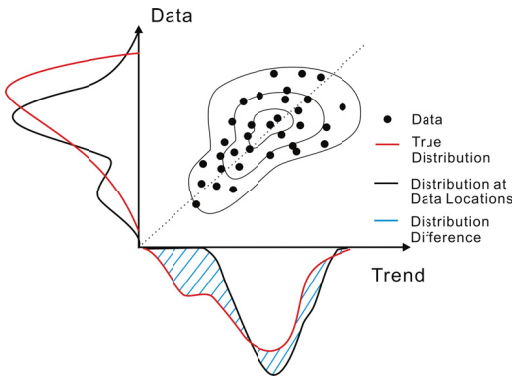


Figure 4.3: Schematic illustration of the missing trend values in the conditional transformation

are re-estimated from the imputed data values (Roberts, 2010; Delalleau et al., 2012; Little and Rubin, 2014).

4.2 Parametric Stepwise Conditional Transformation

The original stepwise conditional transformation proposed by Leuangthong and Deutsch (2003) transforms the residuals conditional to the trend. This approach has binning artifacts due to the non-parametric conditional distributions and creates some negative estimates due to variations within the bins. A revised methodology is proposed. The first change is to transform the variable conditionally to the trend; not the residuals conditioned to the trend. The second change is to use a Gaussian mixture model instead of arbitrary classes to avoid the binning artifacts. The objective of this transformation is to remove the trend-like features from data in a bin-free manner that accounts for the spatial structure and multivariate relationship between the data and the trend.

Consider a set of n observations, $\{\boldsymbol{z}_i, i = 1, \cdots, n\}$. The trend is assumed exhaustive and known. It is represented by $\{\boldsymbol{m}_i, i = 1, \cdots, N\}$. Figure 4.4 shows a schematic illustration of the proposed transformation sequence. The steps for

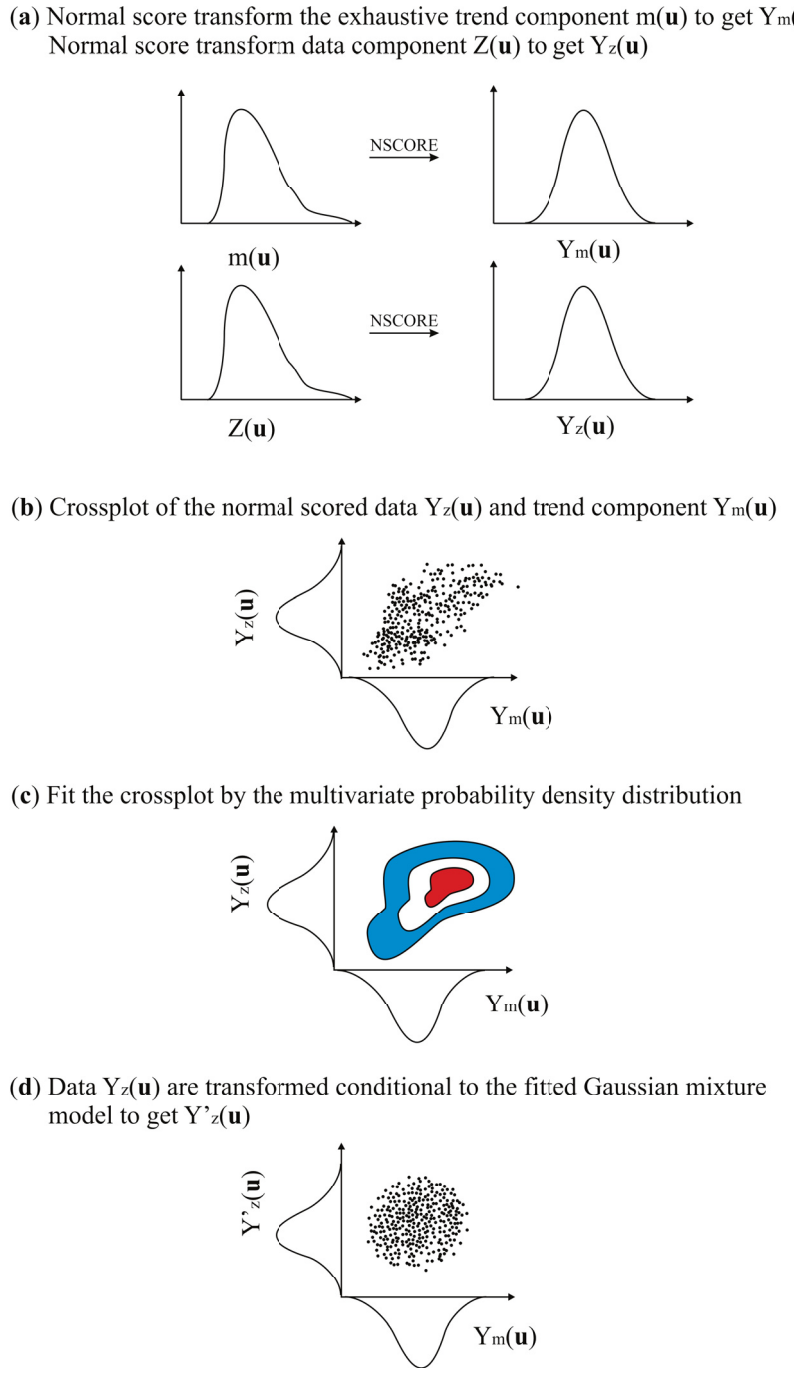


Figure 4.4: Schematic illustration of the proposed parametric conditional transformation procedure: (***a***) normal score transform the trend model and data; (***b***) crossplot of the transformed data and the trend values; (***c***) fit with Gaussian mixtures; and (***d***) transform the data with the Gaussian mixtures

the stepwise conditional transformation using Gaussian mixture models are as follows:

1. Normal score transformation: the trend and the data are transformed into standard normal score units through the normal score transformation individually. The trend model is exhaustive and there is no need to consider the declustering, while the data should be transformed with declustering weights if they are unequally sampled. The normal score transformations are written as:

$$\begin{aligned} \boldsymbol{y}_{m_i} &= G^{-1}\left(F_m(\boldsymbol{m}_i)\right) \quad i = 1, \cdots, N \\ \boldsymbol{y}_{z_i} &= G^{-1}\left(F_Z(\boldsymbol{z}_i)\right) \quad i = 1, \cdots, n \end{aligned} \tag{4.6}$$

   here $\{\boldsymbol{y}_{m_i}, i = 1, \cdots, N\}$ denote the Gaussian transformed trend values. Such trend values are known everywhere. N is the number of grid nodes from the exhaustively sampled trend. $F_m(\cdot)$ represents the cumulative distribution function of the exhaustive trend values. $\{\boldsymbol{y}_{z_i}, i = 1, \cdots, n\}$ is the Gaussian transformed data, while $F_Z(\cdot)$ represents its cumulative distribution function of data values. n represents the number of data where typically $n \leq N$.

2. Review the transformed variables: the transformed data and the co-located transformed trend are cross plotted. This crossplot is used to help choose the number of Gaussian mixture components, g, for the bivariate fitting and the conditional transformation. Too many mixture models will overfit the complexity of the data, while too few mixture models would fail to reproduce important complexity. The number of the Gaussian mixture models should be reasonable, such that it gives reliable conditional distributions from the bivariate distribution of the data and the trend in an artifact-free fashion. It is common to choose between 2 to 5.

3. Multivariate density estimation: consider Gaussian mixture models using a reasonable fitting function to fit the bivariate distribution of the transformed variables. The estimated multivariate density function is calculated by Equation 4.3.

4. Conditional transform the normal score data: the normal score data, $\{\boldsymbol{y}_{z_i}, i = 1, \cdots, n\}$, are transformed by the conditional distribution of $\boldsymbol{y}_{z_i}$ given $\{\boldsymbol{y}_{m_i}, i = 1, \cdots, n\}$. The equation is given as:

$$\boldsymbol{y}'_{z_i} = G^{-1}\left(F_{Z|m}(\boldsymbol{y}_{z_i} \mid \boldsymbol{y}_{m_i})\right) \quad i = 1, \cdots, n \tag{4.7}$$

   where the random variable $\boldsymbol{y}'_{z_i}$ indicates the transformed data by the Gaussian mixture models. $F_{Z|m}(\cdot)$ represents the cumulative distribution function of the data given the exhaustive trend. The cumulative distribution function of the trend facilitates fitting of the mixture model. The transformation considers the trend at each location. The transformed data and the co-located normal score transformed trend has almost no correlation.

The proposed parametric conditional transformation removes the trend-like features that may be problematic in the modeling of the raw data directly. Gaussian simulation can now be used to generate realizations with the transformed data. The back transformation will ensure that the trend model is used everywhere. The trend features will be reproduced in original units.

4.3 Implementation of Modeling with a Trend

In this research, the stepwise conditional transformation with Gaussian mixture models is used to generate multivariate Gaussian variables with synthetic data. The same 2-D example in Chapter 3 is considered and applied. The dataset contains 51 observations. The optimal trend is shown in Figure 3.20 and assumed

known without uncertainty. The data are conditionally transformed to the trend model and applied with sequential Gaussian simulation.

The steps for the simulation using stepwise transformation with Gaussian mixture models are as follows: (1) decluster the data in order to obtain a representative distribution; (2) normal score transform the exhaustive trend and the declustered data; (3) fit the bivariate distribution of the trend and the data with Gaussian mixture models; (4) transform the data with Gaussian mixture models; (5) calculate the variogram models of the transformed data; (6) simulate the transformed data and generate realizations; (7) back transform the realizations by reversing the stepwise conditional transformation and the normal score transformation; and (8) validate the simulated results. All steps are demonstrated with 2-D example.

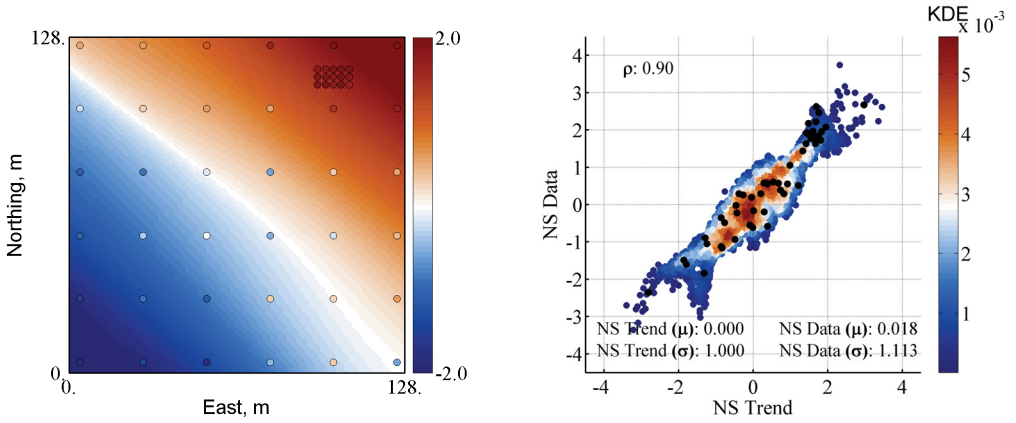
4.3.1 Data Pre-Processing

The data consist of 51 values. The crossplot of the optimal trend with 51 samples is shown with a correlation coefficient $\rho = 0.97$ in Figure 3.20d. The Gaussian mixture models should be fitted in normal score units, thus each variable should be transformed to be a standard normal distribution prior to fitting.

The declustering weights in Section 3.2.3 are considered. The data are normal score transformed with a reference distribution in Figure 3.4c showing slight differences with the assumed mean and standard deviation values. The values of the exhaustive trend model and 51 data are mapped in normal score units in Figures 4.5a and 4.1a. The crossplot in Figure 4.5b shows an almost linear bivariate distribution in density estimation units with a correlation of 0.90.

4.3.2 Gaussian Mixture Model Fitting

Gaussian mixture models are required for the stepwise conditional transformation. Gaussian mixture models incorporate the structures of data that can be



(a) Exhaustive trend model **(b)** Scatter plot

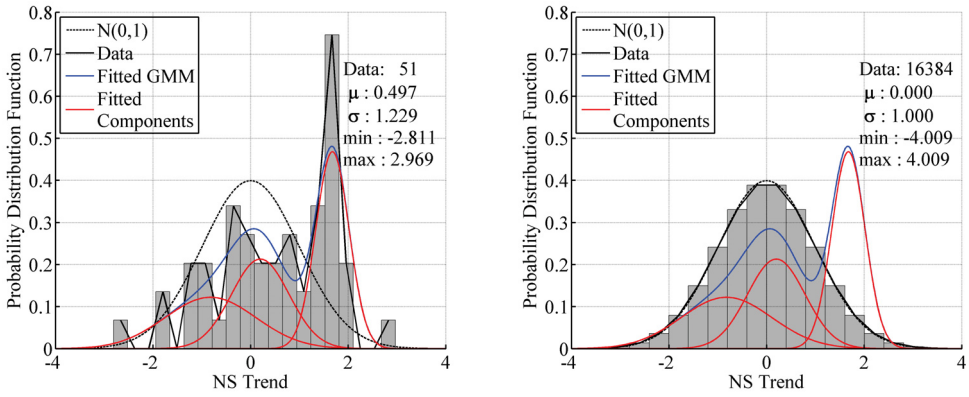
Figure 4.5: Location maps and crossplot of the trend model with 51 data (circles/black markers) in NS units

seen in the crossplot in Figure 4.5b. The EM algorithm is considered for fitting the bivariate distribution of the data values and the trend values for Gaussian mixture models.

Two conditional transformations are constructed based on different types of trend model. Gaussian mixture models with the co-located trend values and Gaussian mixture models with the exhaustive trend values are applied.

4.3.2.1 Fitting with Co-Located Trend Values

Gaussian mixtures with the co-located trend model is considered here for comparison. The mean and covariance of each component from the mixture models are calculated according to the co-located trend values. The fitted Gaussian mixture model, for example with three components, closely matches the probability distribution of the co-located normal score transformed trend values in Figure 4.6a showing a high frequency in high values. However, it does not match with the exhaustive trend distribution in Figure 4.6b. The mismatch of the marginal distribution of the exhaustive trend values would lead to a bias in the



(a) Univariate distribution of the co-located trend

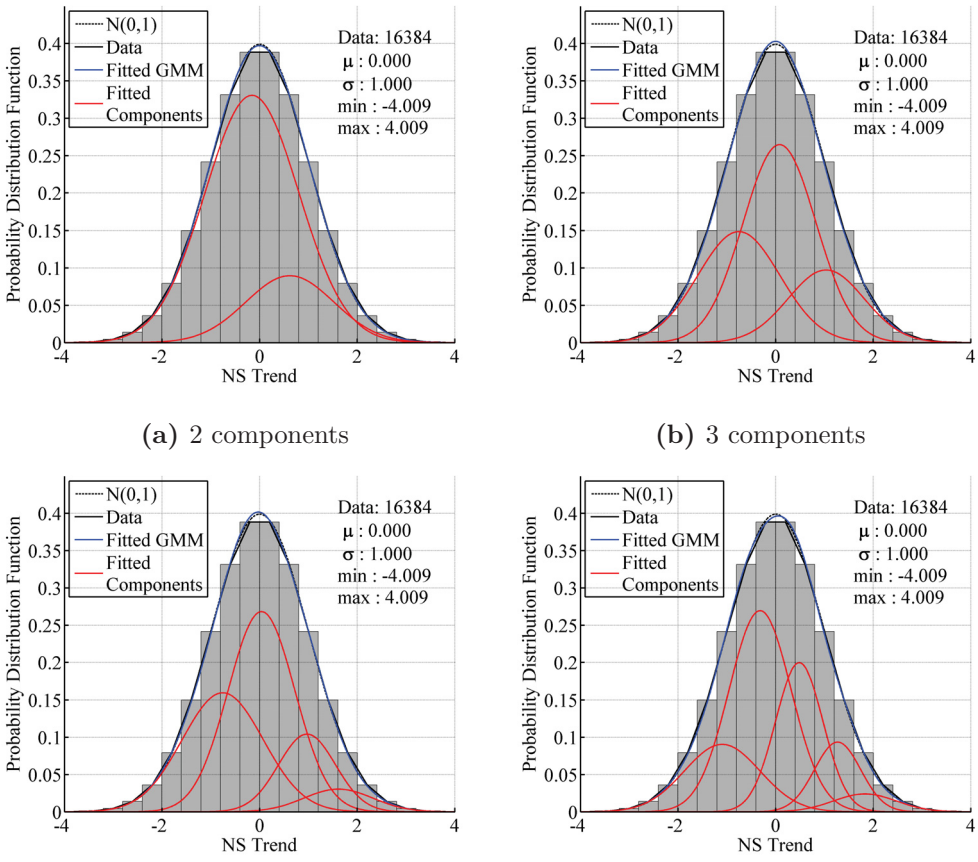
(b) Univariate distribution of the exhaustive trend

Figure 4.6: Univariate distributions of the GMM fitting with co-located trend values using 3 Gaussian components in NS units

final models. Missing values in the trend model should be considered in mixture fittings.

4.3.2.2 Fitting with Exhaustive Trend Values

The EM algorithm that accounts for the missing values from the trend is applied. The fitted Gaussian mixture models using the exhaustive trend with different Gaussian components are shown in Figure 4.7. The deviation appears to be very small. However, the marginal distributions from the Gaussian mixture models of the data values are not exactly normal in Figure 4.8 and small deviations are shown. The bivariate distributions with different Gaussian components are displayed on a 2-D density estimation plot in Figure 4.9. The implementation shows that the Gaussian mixtures with exhaustive trend values work better in fitting with the variables. These Gaussian mixtures using different components are considered to conditionally transform the data.



(a) 2 components

(b) 3 components

(c) 4 components

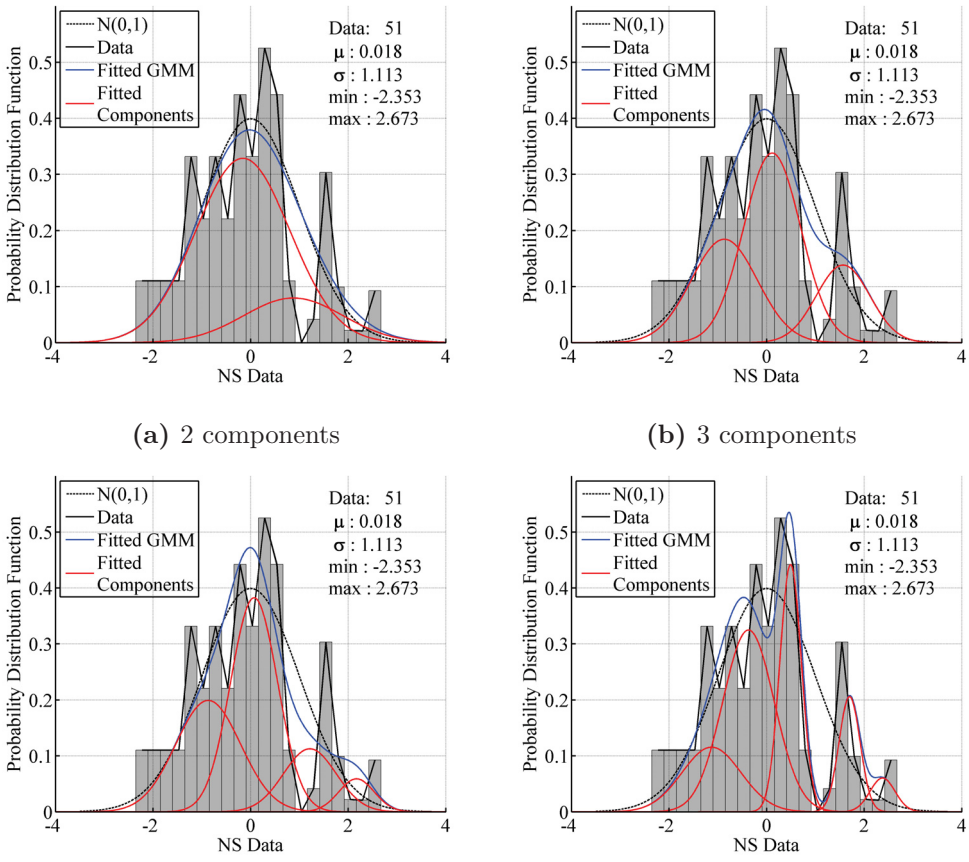
(d) 5 components

Figure 4.7: Univariate distributions of the GMM fitting with exhaustive trend values using different Gaussian components

4.3.3 Parametric Conditional Simulation

Data are conditionally transformed to standard Gaussian distributions according to the chosen Gaussian mixtures. The conditional transformations in Figure4.9 are considered. The order of the stepwise conditional transformation is important. The trend model inferred from the variable must be the primary variable and the the data values are transformed conditionally to the trend model.

Data after applying the stepwise conditional transformation becomes independent from the trend. The location maps of the transformed variables with



(a) 2 components

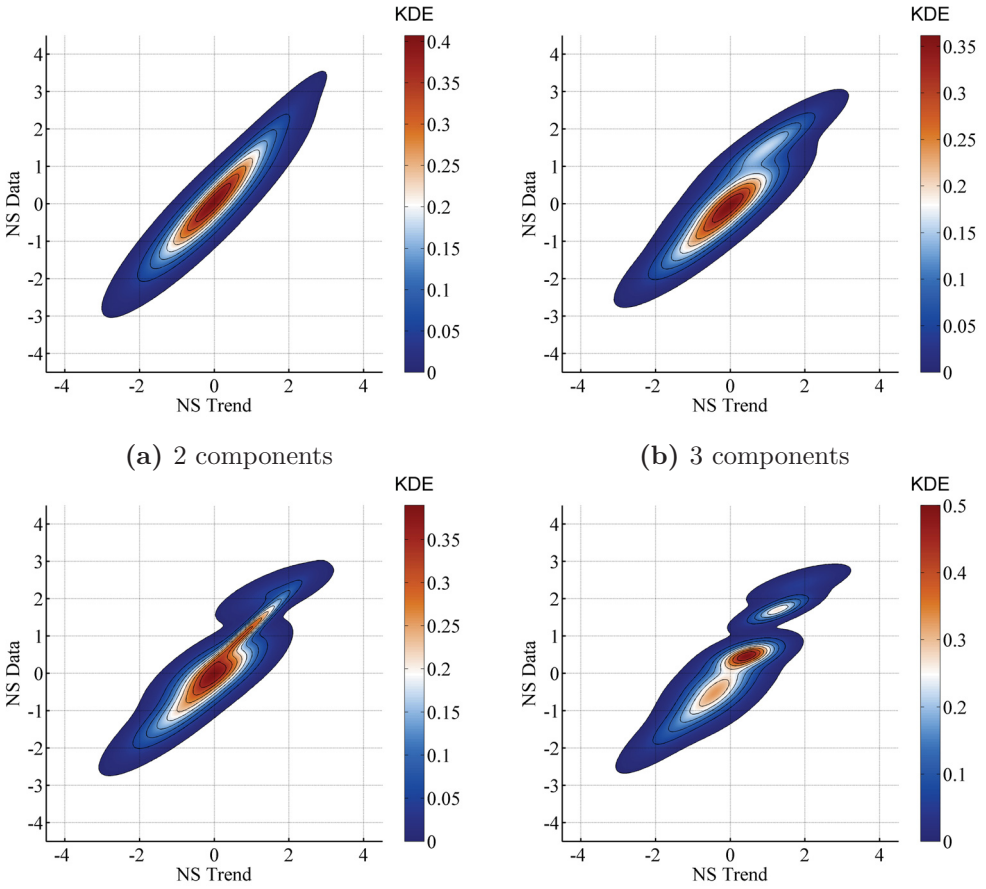
(b) 3 components

(c) 4 components

(d) 5 components

Figure 4.8: Univariate distributions of the GMM fitting with 51 data using different Gaussian components

different Gaussian components are shown in Figure 4.10. The variables after the stepwise conditional transformation yield a bivariate Gaussian distribution. Figure 4.11 shows the crossplots of the trends with 51 transformed data using different Gaussian mixtures in stepwise units. The correlation between the transformed data and the trend is almost zero. These two variables are assumed independent and the transformed data is considered in sequential Gaussian simulation.



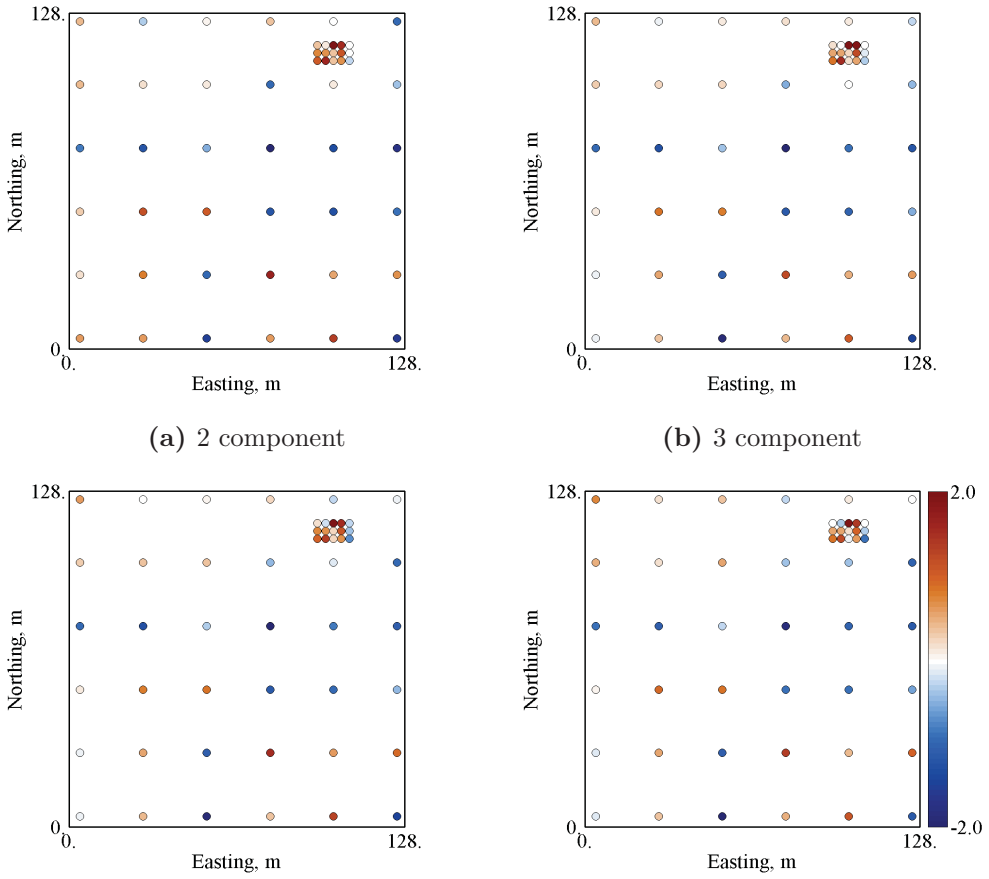
(a) 2 components **(b)** 3 components

(c) 4 components **(d)** 5 components

Figure 4.9: Bivariate distributions of the GMM fitting using different Gaussian components

The spatial statistics of the transformed data can be calculated and modeled by variogram models. The experimental variograms are calculated in the Northeast (green markers) and Northwest (blue markers) directions, as well as the omni-directional (red markers), and shown in Figure 4.12. A zero nugget effect is considered in this synthetic example. One spherical structure is considered due to the simplicity of the variogram. A fitted variogram model (black line) of the transformed variable is shown in Figure 4.12.

Sequential Gaussian simulation is used to generate many realizations at

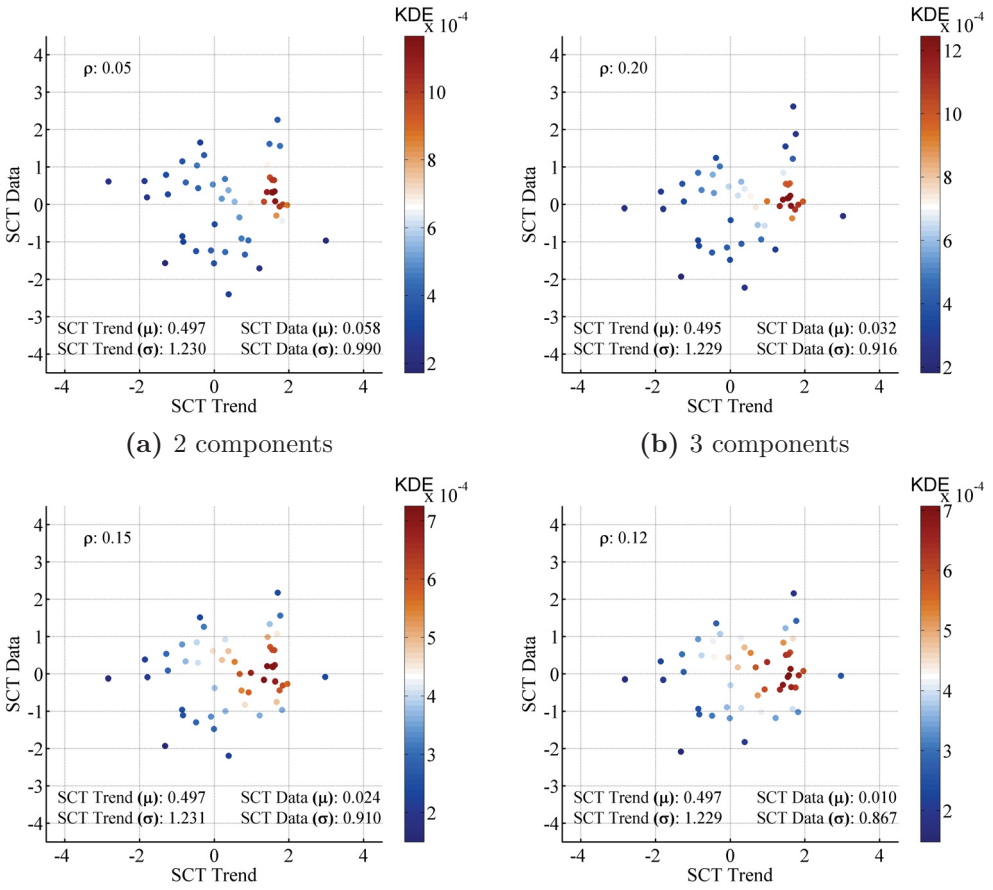


(a) 2 component **(b)** 3 component

(c) 4 component **(d)** 5 component

Figure 4.10: Location maps from 51 transformed variables with different Gaussian components in SC units

a cell size of 1 m × 1 m in North and East directions. The trend model is exhaustively known and it is not simulated. The simulation is performed with the transformed observed data values. Figure 4.13 shows the first three realizations with different Gaussian components in stepwise units. No obvious trend features are seen. The simulation results are back transformed to original units. The first back transformation is the stepwise conditional back transformation. Similar to the forward transformation, the back transformation for the simulated variable should be performed in a conditional fashion using Gaussian mixtures.

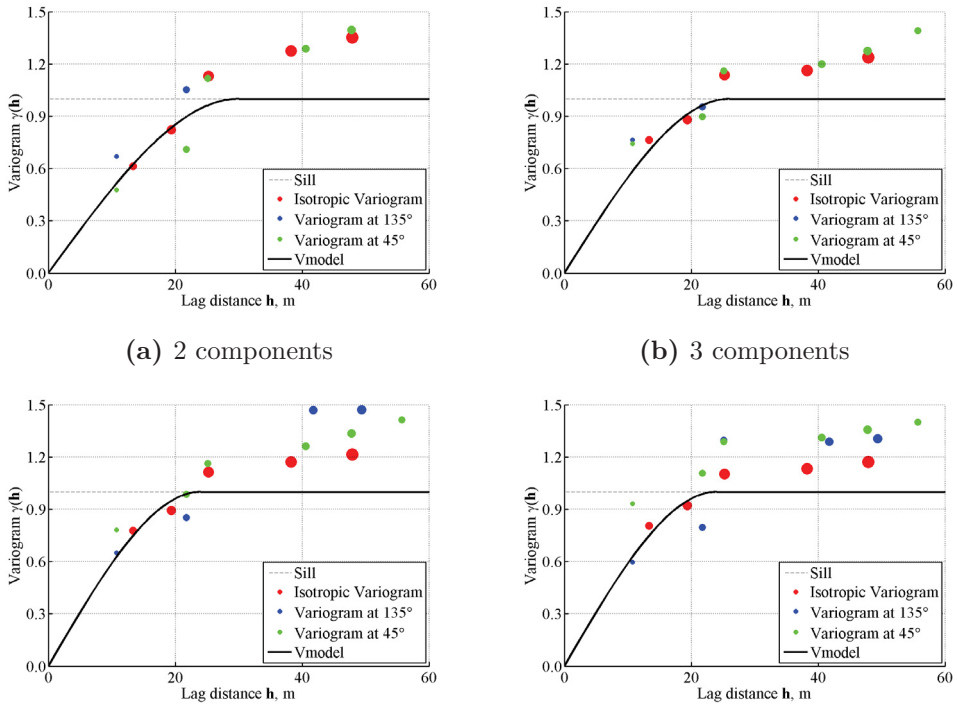


(a) 2 components **(b)** 3 components

(c) 4 components **(d)** 5 components

Figure 4.11: Crossplots of the trend model with 51 transformed variables with different Gaussian components in SC units

Figure 4.14 shows the first three realizations with different Gaussian components in NS units and the trend features are restored. The second back transformation is the normal score back transformation. After the back transformation, all realizations are in original units. Figure 4.15 shows the first three realizations with different Gaussian components in original units. The maps show that high values are located in the Northeast and the values are reduced to the Southwest. The simulation results also show good consistency with the true subsurface in Figure 3.4c.



(a) 2 components **(b)** 3 components

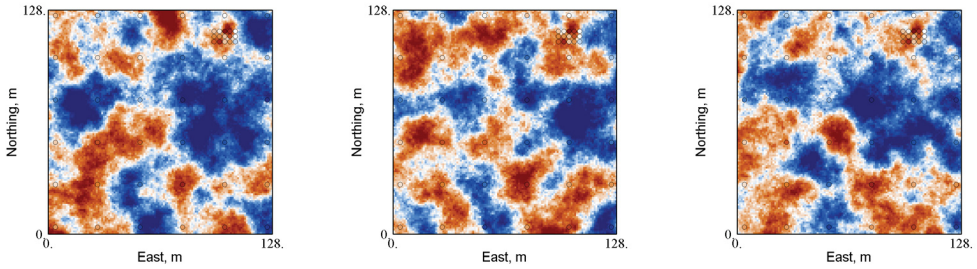
(c) 4 components **(d)** 5 components

Figure 4.12: Variogram models from 51 transformed variables with different Gaussian components in SC units. The sizes of the markers represent the relative pairs of the experimental variogram in each lag distance

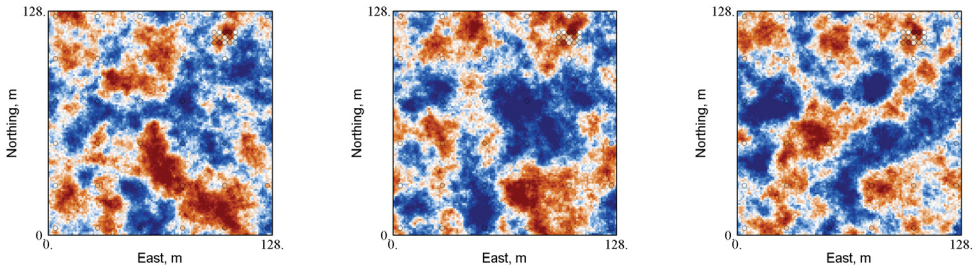
4.3.4 Validation of Simulation

Some checks should be considered to validate the simulated results. The input information including the correlation structures, the data distribution and the spatial relationship in the simulation should be honored.

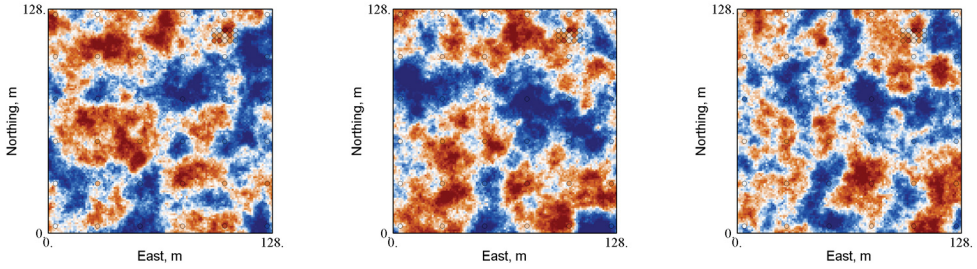
The relationship between the data values and the trend values is checked. The correlation from the input distribution is 0.97 in Figure 3.20d. Figure 4.16 shows the crossplots of 100 simulated values and the trend values with different Gaussian components. All plots show good reproductions of the bivariate distributions with strong correlations. The results show that the proposed stepwise conditional transformation with Gaussian mixture models performs well and no



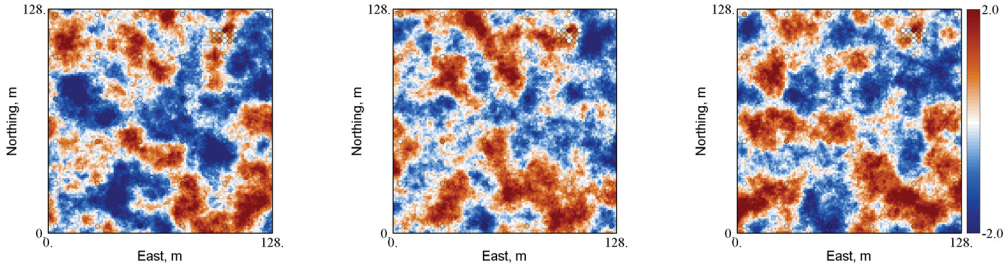
(a) First three realizations with 2 Gaussian components



(b) First three realizations with 3 Gaussian components

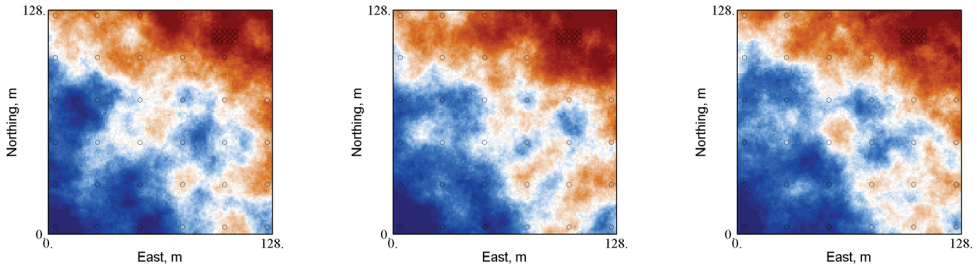


(c) First three realizations with 4 Gaussian components

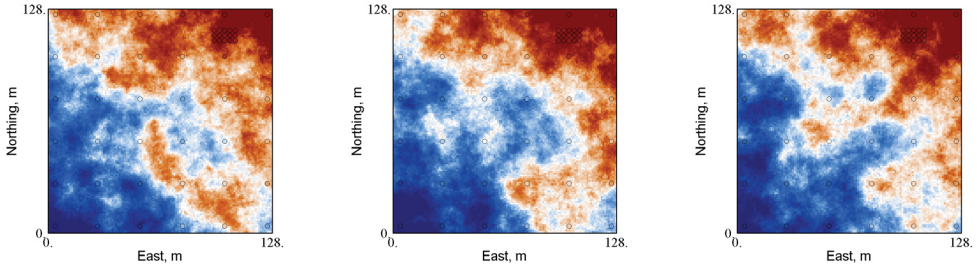


(d) First three realizations with 5 Gaussian components

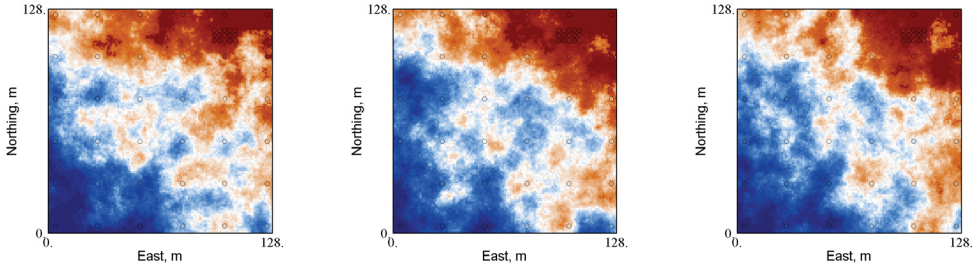
Figure 4.13: First three realizations with different Gaussian components in SC units. The GMM is fitted with exhaustive trend values



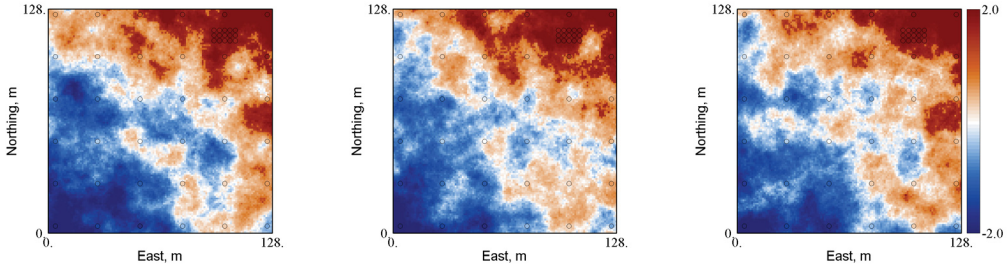
(a) First three realizations with 2 Gaussian components



(b) First three realizations with 3 Gaussian components

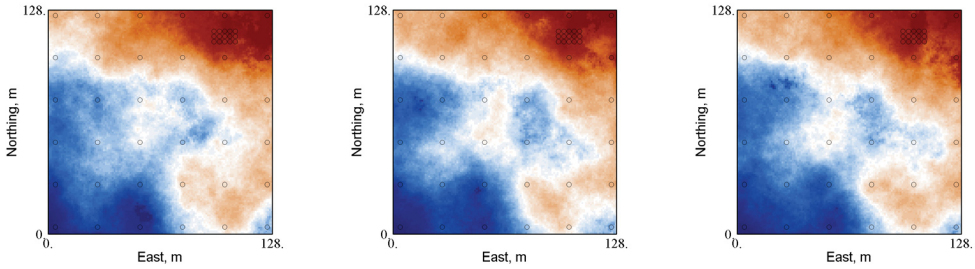


(c) First three realizations with 4 Gaussian components

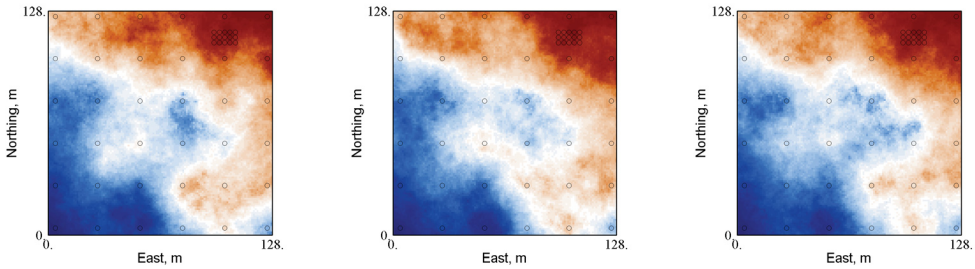


(d) First three realizations with 5 Gaussian components

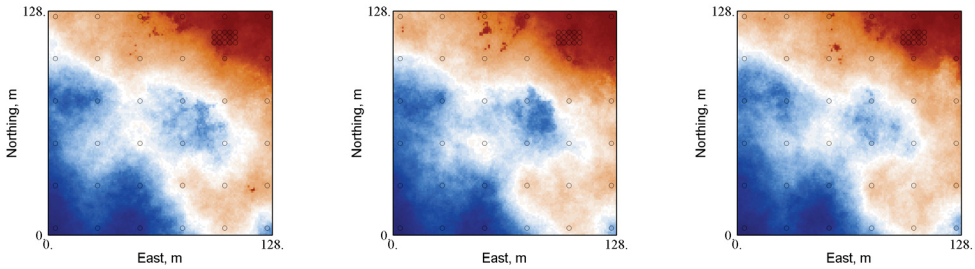
Figure 4.14: First three realizations with different Gaussian components in NS units. The GMM is fitted with exhaustive trend values



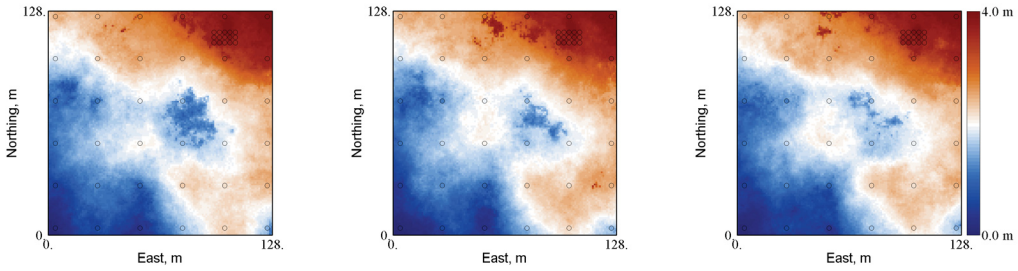
(a) First three realizations with 2 Gaussian components



(b) First three realizations with 3 Gaussian components



(c) First three realizations with 4 Gaussian components



(d) First three realizations with 5 Gaussian components

Figure 4.15: First three realizations with different Gaussian components in original units. The GMM is fitted with exhaustive trend values

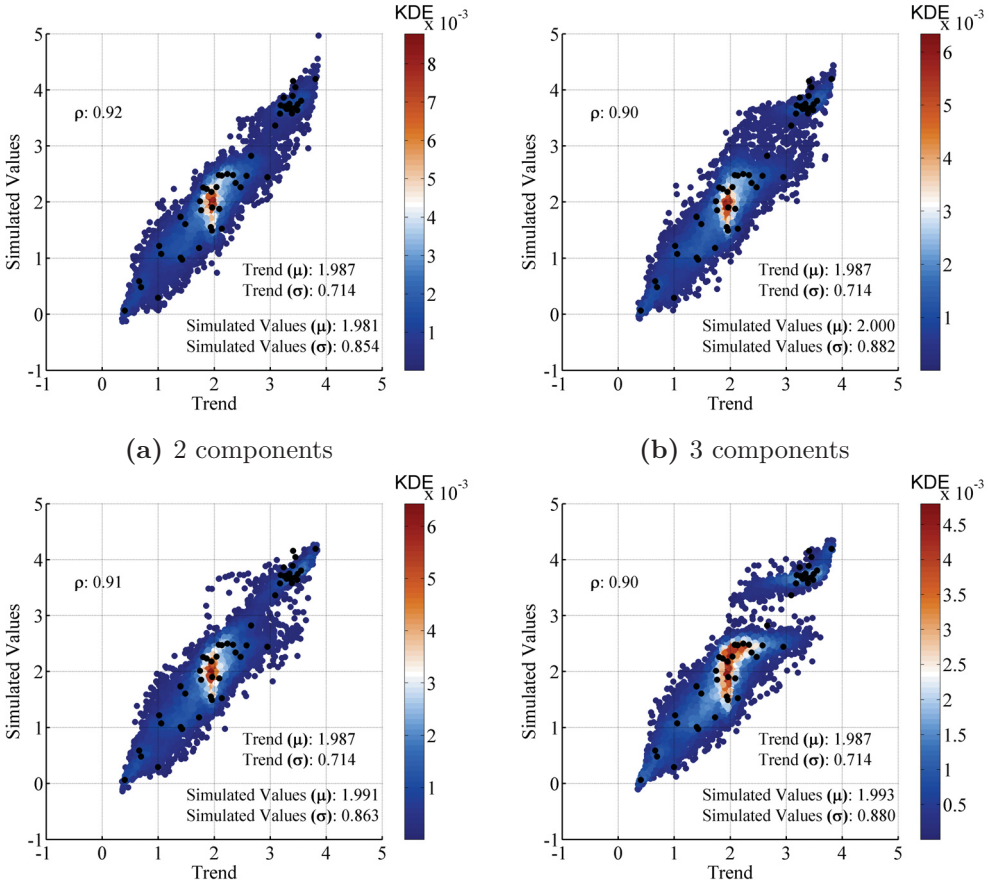
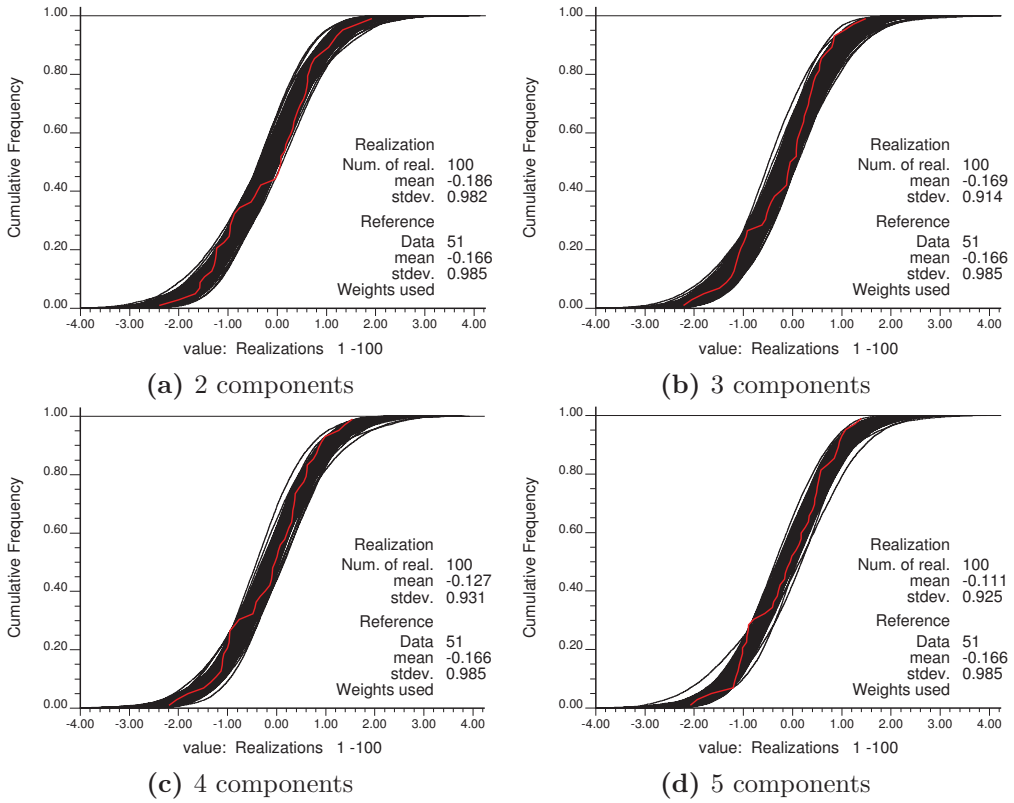


Figure 4.16: Crossplots of the trend model with 100 simulated realizations in original 2-D KDE units. 51 conditioning data are shown with black markers

artifacts are introduced. However, there is a gap in the crossplot with 5 components in Figure 4.16d due to the complexity in the high number of Gaussian mixtures in Figure 4.9d.

The histogram of the simulated values is an important parameter. The histograms over many realizations should be similar to the data histogram with comparable statistics. Figure 4.17 shows the histogram reproduction in stepwise units. A mean value and a standard deviation associated with the global distribution of each realization are calculated and plotted with the black lines,

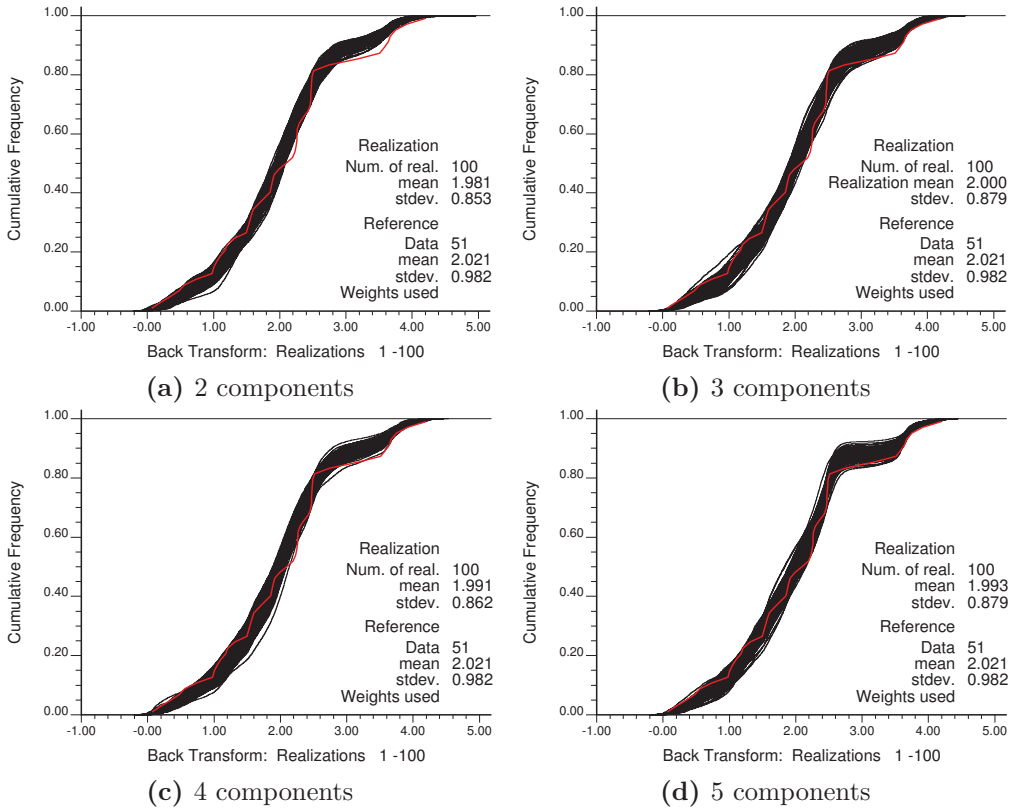


(a) 2 components **(b)** 3 components

(c) 4 components **(d)** 5 components

Figure 4.17: Histogram reproduction over 100 realizations in SC units. A mean value and a standard deviation associated with the global distribution of each realization are plotted with the black lines. 51 conditioning data are shown with a red line as a reference

while the input conditionally transformed variable is shown with a red line as a reference. The first and second order statistics of realizations with different Gaussian mixtures are close to the mean and the standard deviation of the input variable. Figure 4.18 shows the performance of the histogram reproduction with the summarized statistics over 100 realizations in original units. The reference of the 51 original data showing with a red line is approximately in the middle of the realizations. Overall, the distribution and the statistics of the input data in original units are satisfactorily reproduced.

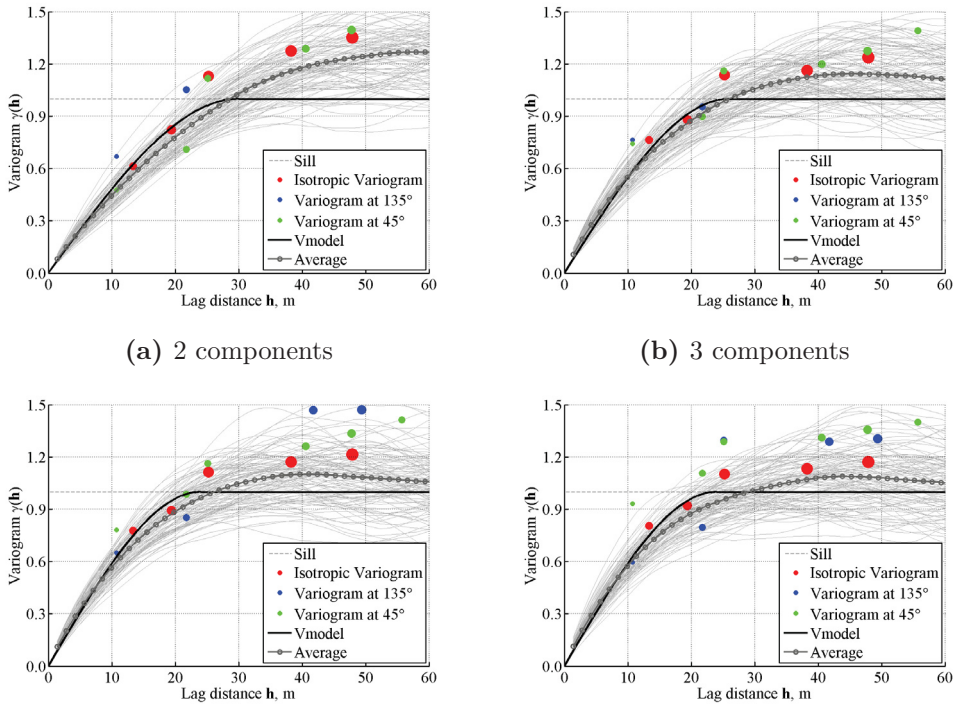


(a) 2 components **(b)** 3 components

(c) 4 components **(d)** 5 components

Figure 4.18: Histogram reproduction over 100 realizations in original units. A mean value and a standard deviation associated with the global distribution of each realization are plotted with the black lines. 51 conditioning data are shown with a red line as a reference.

The variogram should be checked in stepwise units since the input variogram is considered in simulation and should be reproduced. Figure 4.19 shows the result of the variogram reproduction in stepwise units. The experimental variograms in different directions (Northeast, Southwest and omni-directional) are shown with different colors of markers. The variogram model of each realization is plotted with the light gray line and the average of 100 realizations is shown with a dark gray line. The variograms over 100 realizations closely follow the input stepwise variograms (black line). The variogram in original units is also checked. Figure 4.20 shows the variogram reproduction over 100 realizations

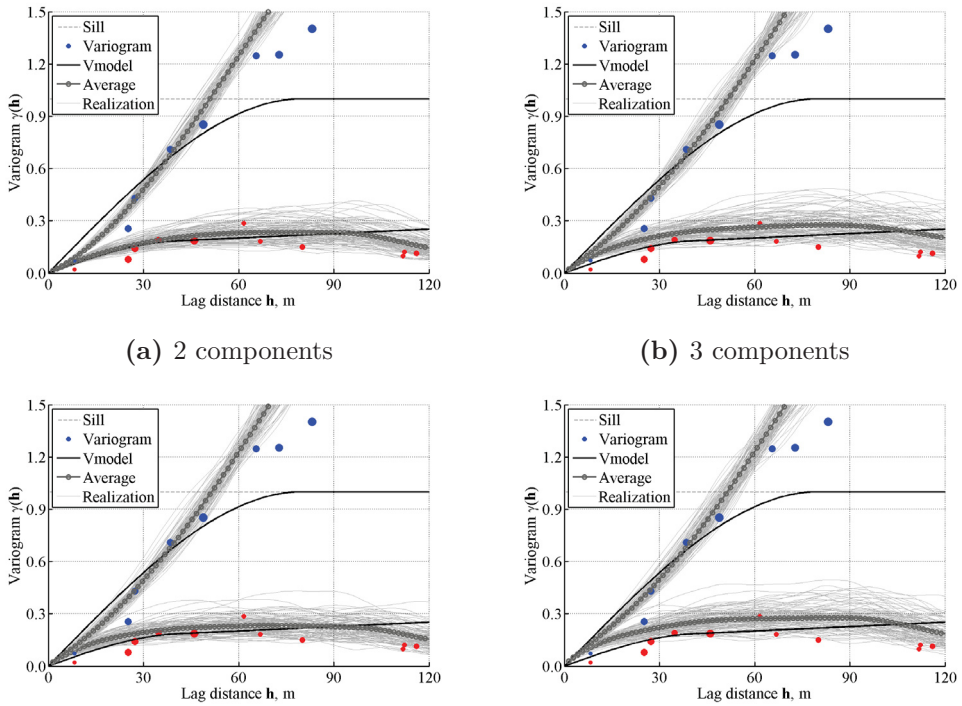


(a) 2 components **(b)** 3 components

(c) 4 components **(d)** 5 components

Figure 4.19: Variogram reproduction over 100 realizations in SC units. The experimental variograms in different directions are shown with different colors of markers. The sizes of the markers represent the relative pairs of the experimental variogram in each lag distance. The variogram model of each realization is plotted with the light gray line and an average realization is shown with a dark gray line.

with different Gaussian mixtures. The overall variogram models from realizations (gray lines) match with the experimental variogram models (markers) in original units, although the variogram reproduction is better with fewer Gaussian components.



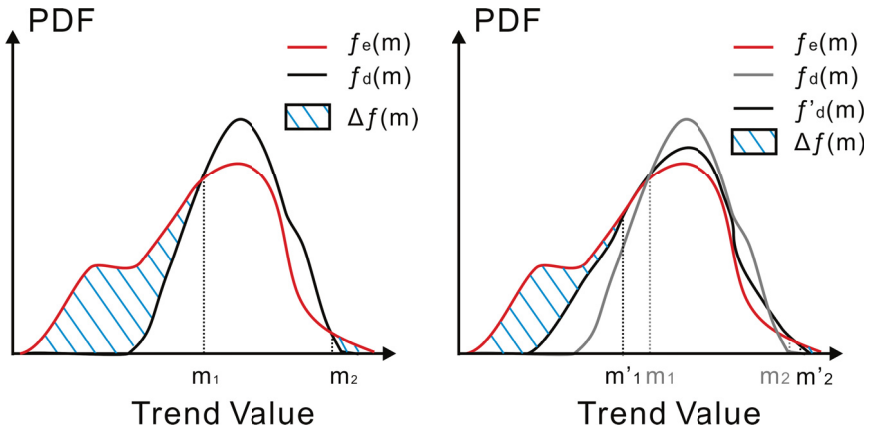
(a) 2 components **(b)** 3 components

(c) 4 components **(d)** 5 components

Figure 4.20: Variogram reproduction over 100 realizations in original units. The experimental variograms in different directions are shown with different colors of markers. The sizes of the markers represent the relative pairs of the experimental variogram in each lag distance. The variogram model of each realization is plotted with the light gray line and an average realization is shown with a dark gray line.

4.4 Special Topic of Modeling with a Trend

The missing values in the trend model must be considered, especially when the trend model is exhaustively known, see Figure 4.3. An alternative method, instead of using an arbitrary data imputation mechanism in EM algorithm, is proposed and implemented. The marginal distribution of the updated trend values can be imposed on fitting the bivariate distribution to improve the estimates.



(a) Initial distribution of the trend **(b)** Distribution of the trend after the first iteration

Figure 4.21: Schematic illustration of the updating trend distribution

4.4.1 Fitting with Pseudo Data

The idea is to correct the bivariate distribution of the trend and the data values by providing some pseudo data, that is, data that do not exist but could be generated to modify the bivariate distribution in a desirable fashion. An assumption is that the expected value of data equals the trend values at each location, that is, $E\{Z(\boldsymbol{u})\} = m(\boldsymbol{u})$. The procedure is to draw multiple representative points from the exhaustive trend distribution so that the updated trend distribution matches the exhaustive trend distribution. Such correction leads to provide an unbiased and accurate conditional transformation.

Figure 4.21 shows the iterative updated probability function of the trend values. Figure 4.21a shows the initial distribution of the trend values. The red line shows the distribution of the exhaustive trend $f_e(m(\boldsymbol{u}))$, while the black line represents that distribution only at data locations $f_d(m(\boldsymbol{u}))$. The distribution of the trend is more informed if the probability of the co-located trend is higher than that of the exhaustive trend, that is, $f_d(m(\boldsymbol{u})) > f_e(m(\boldsymbol{u}))$, and the distribution of the trend is not required to be updated; otherwise, the distribution of the

trend is required an update. The regions that require an update are shown with the blue stripes in Figure 4.21a. Once one value is drawn from the distribution of the exhaustive trend, the distribution of the updated trend $f_{\mathrm{d}}^{'}(m(\boldsymbol{u}))$ is shown with the gray line in Figure 4.21b. The regions that require the update are changed. The target difference between the exhaustive trend distribution and the updated trend distribution is denoted as $\triangle f(m(\boldsymbol{u}))$ which is given as:

$$\triangle f(m(\boldsymbol{u})) = \begin{cases} f_{\mathrm{e}}(m(\boldsymbol{u})) - f_{\mathrm{d}}(m(\boldsymbol{u})) & \text{if} f_{\mathrm{d}}(m(\boldsymbol{u})) \leq f_{\mathrm{e}}(m(\boldsymbol{u})) \\ 0 & \text{otherwise} \end{cases} \quad \forall \boldsymbol{u} \in A \quad (4.8)$$

here $f_{\mathrm{e}}(m(\boldsymbol{u}))$ represents the probability function of the exhaustive trend values and $f_{\mathrm{d}}(m(\boldsymbol{u}))$ represents the probability function of the updated trend values. $\triangle f(m(\boldsymbol{u}))$ represents the difference between these distributions that can be iteratively updated only when the distribution $f_{\mathrm{d}}(m(\boldsymbol{u}))$ is improved. The similarity sim between the exhaustive trend and the updated trend is discretized and can be quantified by:

$$sim = \frac{1}{ndis} \cdot \sum_{i=1}^{ndis} \triangle f^2(m(\boldsymbol{u})) \quad \forall \boldsymbol{u} \in A \quad (4.9)$$

here $ndis$ represents the number of the discretization of the probability function. The number of iterations should be balanced with the similarity measure and the computational cost. A perfect match $sim = 100\%$ is highly unlikely to be reached, whereas a lower similarity does not improve the bivariate distribution too much. Higher than 85% similarity is recommended in practice. Some deviations from the exhaustive trend distribution always exist.

The trend distribution continues to be updated. The pseudo data value is corrected by this updated trend value that is given as:

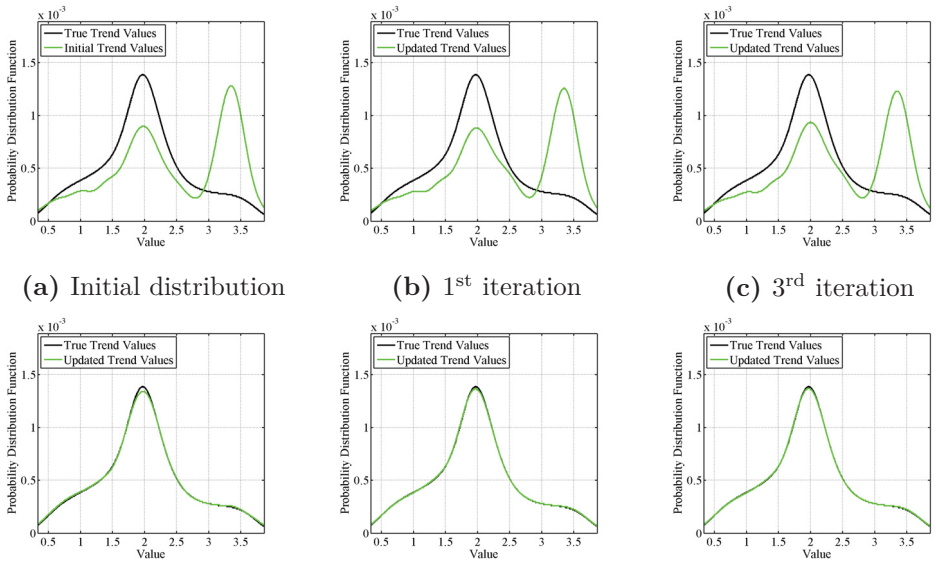
$$\begin{aligned} z(\boldsymbol{u}_j) = \hat{r}(\boldsymbol{u}_i) \cdot \varphi + m(\boldsymbol{u}_p) \quad \forall \boldsymbol{u}_i, \boldsymbol{u}_j, \boldsymbol{u}_p \in A \\ j = 1, 2, \cdots \quad i = 1, 2, \cdots, n \text{ and } p = 1, 2, \cdots, N \end{aligned} \tag{4.10}$$

where $\{j = 1, 2, \cdots\}$ represent the number of the updated data values, $\{i = 1, \cdots, n\}$ represent the set of original n data locations and $\{p = 1, \cdots, N\}$ are the number of the exhaustive trend values. $\{\hat{r}(\boldsymbol{u}_i), i = 1, \cdots, n\}$ are the global distribution of the residuals at n original data locations. $\{m(\boldsymbol{u}_p), p = 1, \cdots, N\}$ donate the trend values that are randomly drawn from the distribution of the exhaustive trend. φ is a known function of the residual data conditional to the trend model. The trend value is iteratively drawn from the exhaustive distribution and the pseudo data is computed. The iterative procedure stops when the similarity indicator *sim* reaches to the target.

4.4.2 Implementation of Fitting with Pseudo Data

The aim of the pseudo data addition helps to build a representative conditional distribution. The same 2-D example is considered. The known exhaustive trend is discretized into 2000 intervals. The distribution of the trend model is calculated in each interval. Due to the simplicity of the example, the similarity target is set to 99%.

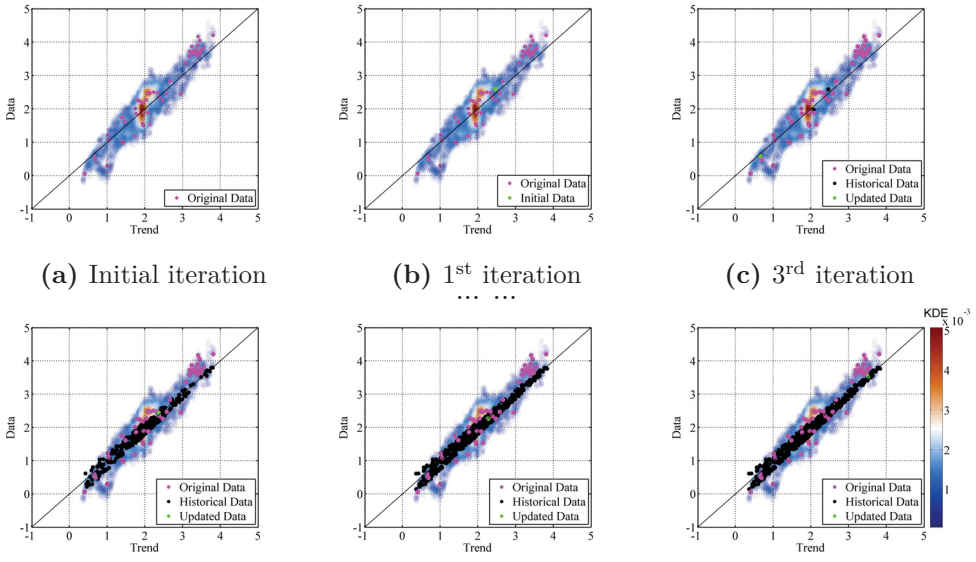
Figure 4.22 shows the probability distributions of the trend models with iterative draws from the exhaustive trend distribution. The black line represents the target distribution that is the probability function of the exhaustive trend model. The initial trend distribution comes from the co-located trend values that are represented by a blue line, see Figures 4.22a. The differences between these distributions are clear. One value is randomly drawn from the exhaustive trend



(a) Initial distribution **(b)** 1st iteration **(c)** 3rd iteration

(d) 300th iteration **(e)** 595th iteration **(f)** 596th iteration

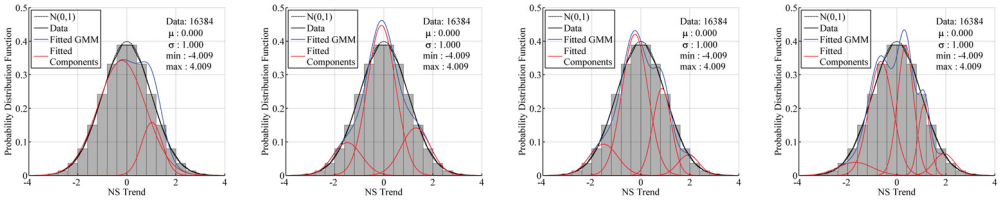
Figure 4.22: PDFs of the updated trend in original units. Values are randomly drawn from the exhaustive trend distribution. The iteration stops until the requirement on the similarity is satisfied



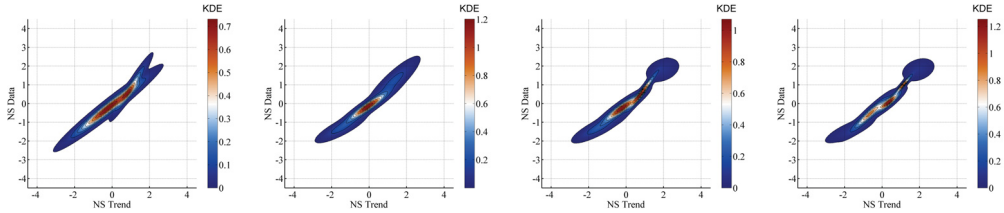
(a) Initial iteration **(b)** 1st iteration **(c)** 3rd iteration

(d) 300th iteration **(e)** 595th iteration **(f)** 596th iteration

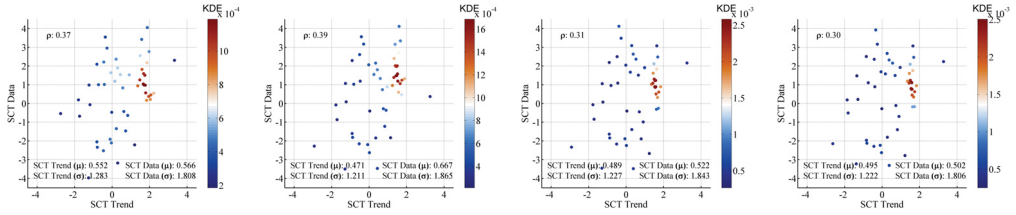
Figure 4.23: Crossplots of the updated trend model with 596 pseudo data in original units. Pseudo data are calculated by Equation 4.10



(a) Univariate distributions of the GMM fitting with updated trend values using different Gaussian components in NS units



(b) Bivariate distributions of the GMM fitting with pseudo data addition using different Gaussian components in NS units



(c) Crossplots of the trend model with 51 transformed variables with different Gaussian components in SC units

Figure 4.24: Univariate and bivariate distributions of the GMM fitting with pseudo data addition (top, middle) and crossplots of the trend model with 51 transformed variables using different Gaussian components (bottom)

model and the distribution of trend is updated and shown in Figure 4.22b. Note that the new distribution in Figure 4.22b does not display remarkably differences with the previous distribution in Figure 4.22a. The improvement is visible in more iterations, see Figure 4.22d with 300 iterations. The iteration continues until the similarity between the exhaustive trend distribution and the updated trend distribution is satisfied, see Figure 4.22f.

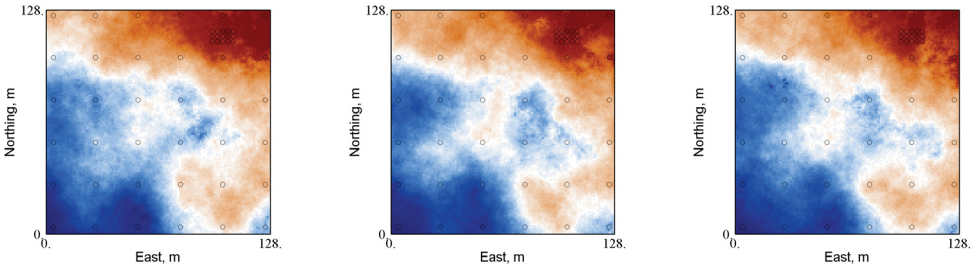
The pseudo values are calculated by Equation 4.10 as drawing new trend values. Figure 4.23 shows the crossplots of the drawn trend and with calculated

data values. 51 original data values are shown with pink markers, the updates in previous iterations are shown with black markers, and the current updated pair is shown with a green marker. The background values are shown in a 2-D density estimation plot (blue-red markers in transparent shades) that are inferred from the true distributions (true trend model and true subsurface model) in Figure 3.4. 596 pseudo data are generated until a target satisfaction of 99% is met. The plot in Figure 4.23f shows a close match to the 45° regression line and the background values.

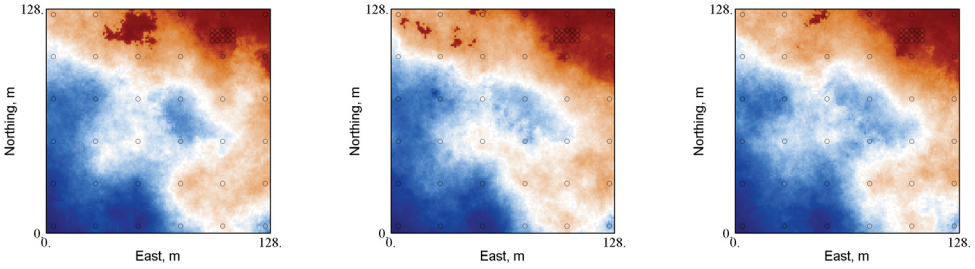
The conditional transformation is informed by a total of $51 + 596 = 647$ data and the corresponding trend values. The Gaussian mixture models with different components are constructed with 647 data. The univariate distributions of the trend model with different Gaussian components are shown in Figure 4.24a. The fitted Gaussian mixtures with pseudo data performs better than that with co-located trend values due to the fitted distribution closes to the exhaustive trend distribution, although the deviations are shown because of the mismatch. Figure 4.24b shows bivariate distribution of Gaussian mixtures with different components. Some unreasonable Gaussian components are generated when a large number of components is chosen.

51 data are conditionally transformed according to the bivariate distributions in Figure 4.24b into stepwise units. The correlations between the data and its trend are reduced to around 0.35 after the transformation in Figure 4.24c and some trend features are still stored in 51 data.

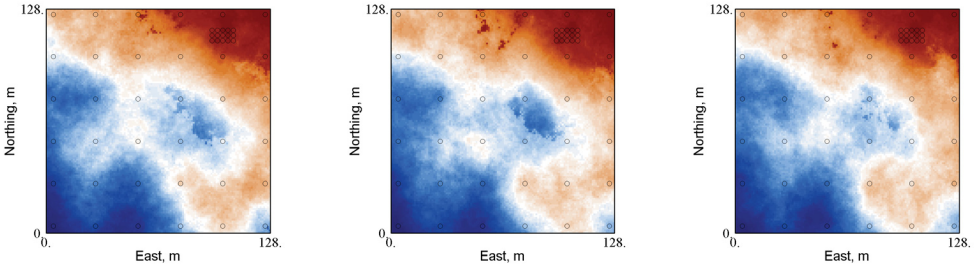
Geostatistical modeling with 51 transformed data is applied and the simulated values are back transformed into original units. Figure 4.25 shows the first three realizations with different numbers of Gaussian components. Realizations are compared with the true image in Figure 3.4c. Some extremely high values can been seen in middle-value locations. Figure 4.26 shows the crossplots of the trend models with 100 simulated values in original units. Some obvious artifacts



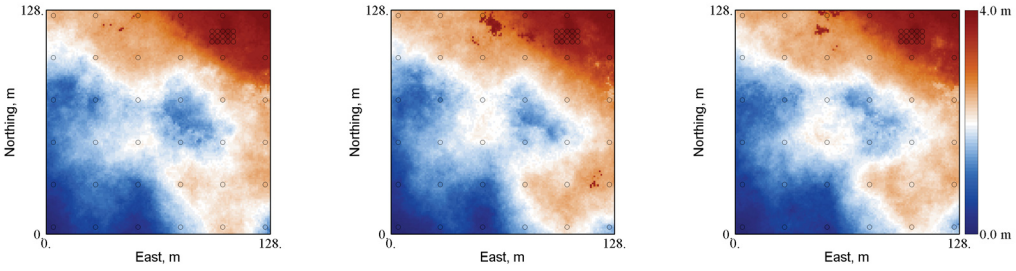
(a) First three realizations with 3 Gaussian components



(b) First three realizations with 3 Gaussian components



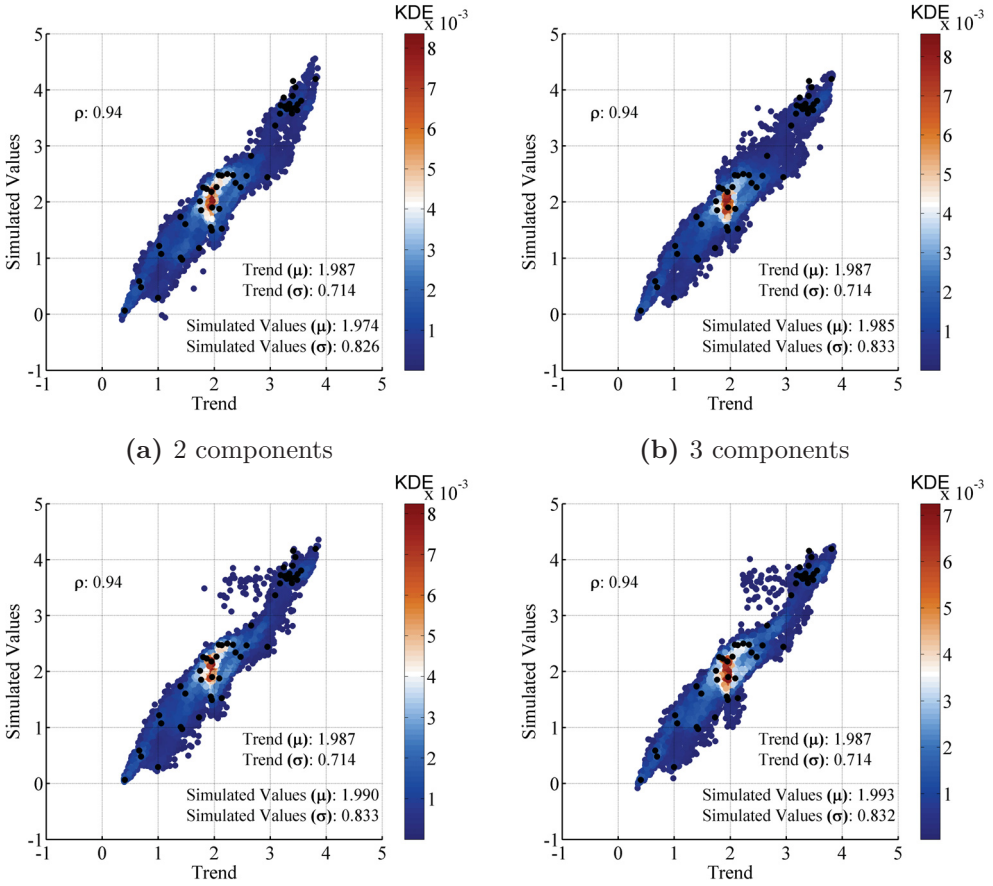
(c) First three realizations with 4 Gaussian components



(d) First three realizations with 5 Gaussian components

Figure 4.25: First three realizations with different Gaussian components in original units. GMMs are fitted with pseudo data

Crossplots of the trend model with 100 simulated realizations in original units



(a) 2 components **(b)** 3 components

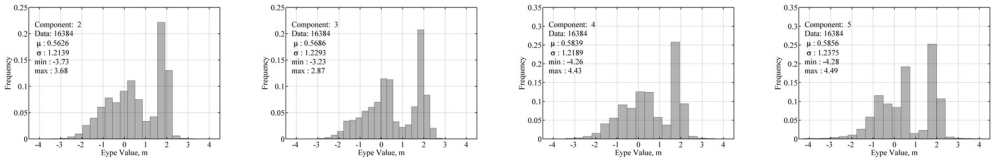
(c) 4 components **(d)** 5 components

Figure 4.26: Crossplots of the trend model with 100 simulated realizations in original units

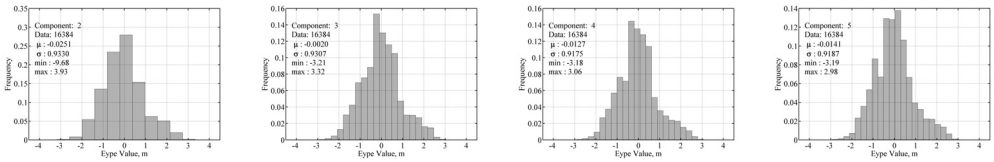
are seen far away from the regression line due to the unexpected features in mixture distributions in Figure 4.24b.

4.5 Validation of Parametric Conditional Transformation Methods

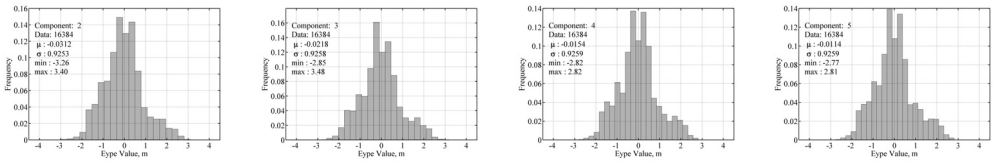
The main goal in this section is to validate the simulated results in Sections 4.3 and 4.4. The applications of the conditional transformations combining co-located trend values, exhaustive trend values and pseudo data are required to be validated.



(a) Uncertainty with GMM using co-located trend fitting



(b) Uncertainty with GMM using exhaustive trend fitting

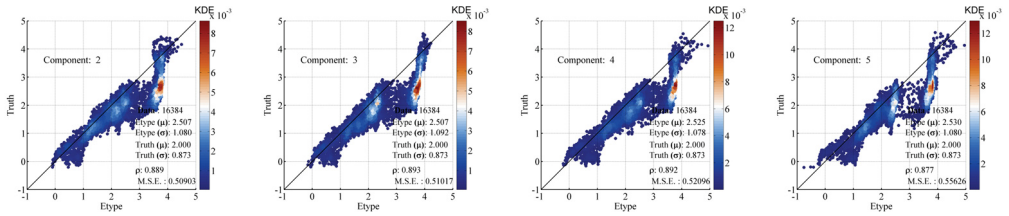


(c) Uncertainty with GMM using pseudo data fitting

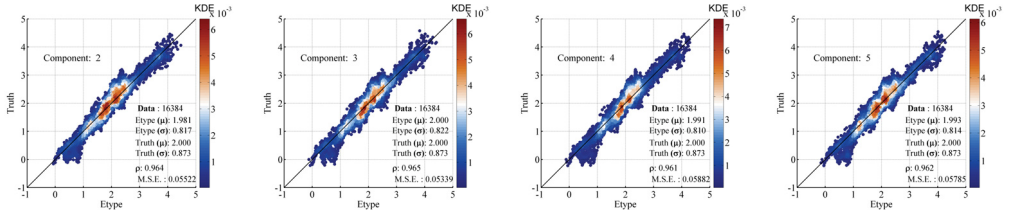
Figure 4.27: Uncertainties with different conditional transformations and Gaussian components in NS units

The uncertainty with different conditional transformations is checked in normal score units. The simulated realizations in stepwise units are conditionally back transformed into normal score units. All realizations are averaged. The standard deviation of the averaged simulation at $128 \times 128 = 16384$ checking locations could be understood as the uncertainty of the numerical model. The

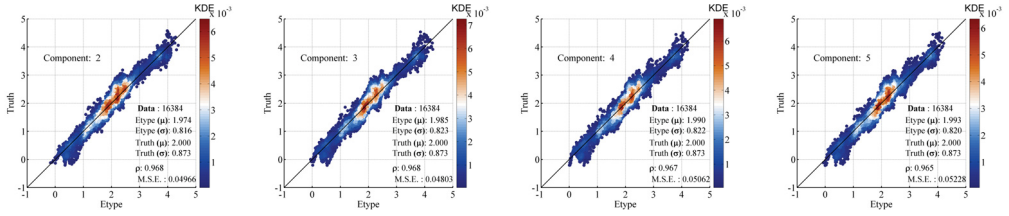
plots in Figure 4.27 show the uncertainty in the model in terms of Gaussian mixtures fitting with co-located trend, exhaustive trend and pseudo data. The uncertainty is quite high with a fitting of the co-located trend values and low with other two methods. The results highlight that the missing trend values should be considered in the conditional transformation and the uncertainty is reduced in the model.



(a) Cross validation using the conditional transformation with co-located trend fitting



(b) Cross validation using the conditional transformation with exhaustive trend fitting



(c) Cross validation using the conditional transformation with pseudo data fitting

Figure 4.28: MSE comparisons with different conditional transformations and Gaussian components in original units

The cross validation is designed to assess the goodness of a algorithm with data. The true values that are included with the input 51 values are compared

with the simulated averages in original units. Figure 4.28 shows the comparisons between the true data and the etype results (average values from 100 realizations). The strong positive correlations between true and etype are seen. Discarding unequally sampled trend values impacts the final estimates and some unexpected relationships are found, see Figure 4.28a.

The mean squared error value measures the difference between true data and what is being estimated, and further, summarizes the prediction performance. The results indicate that the conditional transformation using pseudo data fitting with 3 components shows a minimum mean squared error value. Although this method leads to a minimum mean squared error value, this application is not recommended due to some unstable Gaussian mixture models, see Figure 4.24b, and some artifacts in realizations, see Figure 4.25.

The statistics are other factors to quantify the performance of the implementation. The mean and the standard deviation in the simulated results in Figure 4.28b close to the truth. Gaussian mixture model using exhaustive trend model produces a stable and reasonable Gaussian components and provides artifact-free estimates.

4.6 Remarks

The developed geostatistical modeling algorithm considers the deterministic features of the continuous variable in an artifact-free fashion. Data with an apparent trend are transformed conditionally by a parametric transformation. The stepwise conditional transformation with Gaussian mixture models removes the complexity multivariate features. The variables at zero lag distances are decorrelated and form uncorrelated multivariate Gaussian distributions. The improved performance of the geostatistical algorithm is attributed to the stationarity of the transformed result after the parametric conditional transformation

in artifacts-free fashion.

The trend is estimated and exhaustively known after trend modeling. The trend values at unequally locations could be considered in the conditional transformation. As the missing values are not random, some pseudo data can be introduced to correct the bivariate distribution of the Gaussian mixtures and generate appropriate results. However, some unreasonable mixtures may be produced and they obstruct the estimates. EM algorithm also provides an stable imputation mechanism that the multivariate probabilities are marginalized by the observed data and the exhaustive trend.

In this chapter, a 2-D example illustrates the fit with different Gaussian components, although the optimal number of components for the mixture models is not known in general. Data in the presence of a trend are transformed by a parametric conditional transformation. Sequential Gaussian simulation is applied with the transformed variable. Original complex features are re-introduced in the back transformation. The performance of numerical models, the reproduction of geological characterizations and the local uncertainty are analyzed. The results show that the proposed Gaussian mixture approach produces a consistent estimate and too many mixtures would increase the complexity of the model.

Chapter 5

Case Study 1: 2-D Sub-Surface Mapping

The theories on modeling a trend and incorporating it into geostatistical prediction are given in Chapters 3 and 4. The objectives of this chapter are to model a 2-D sub-surface in the presence of a trend and demonstrate the importance of considering the trend in geostatistical modeling. Figure 5.1 shows the overall workflow. Visual inspection and some basic statistical analyses of the data including the location map and the variogram indicate the existence of trend-like features. Conventional geostatistical modeling without considering a trend is considered for comparison. The procedure involves a normal score transformation, variogram fitting, sequential Gaussian simulation and a normal score back transformation. Then, the assumption of stationarity is relaxed, and the developed geostatistical modeling workflow is implemented. The developed method includes the averaged anisotropy calculation, trend modeling, the selection of the optimal trend model, a normal score transformation, Gaussian mixture model determination, the stepwise conditional transformation, variogram fitting, sequential Gaussian simulation and two back transformations. Finally,

the performances of the geostatistical predictions are assessed by cross validations. The cross validation shows that the model with the proposed method is better than that with the conventional method.

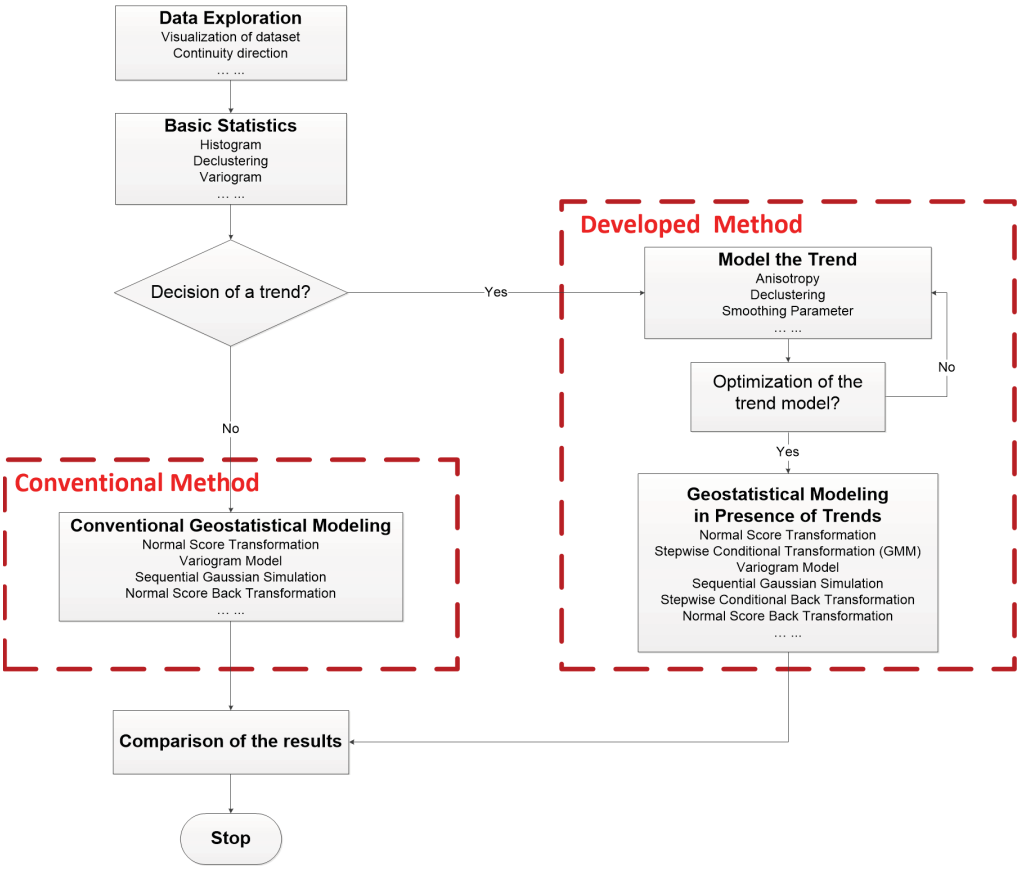
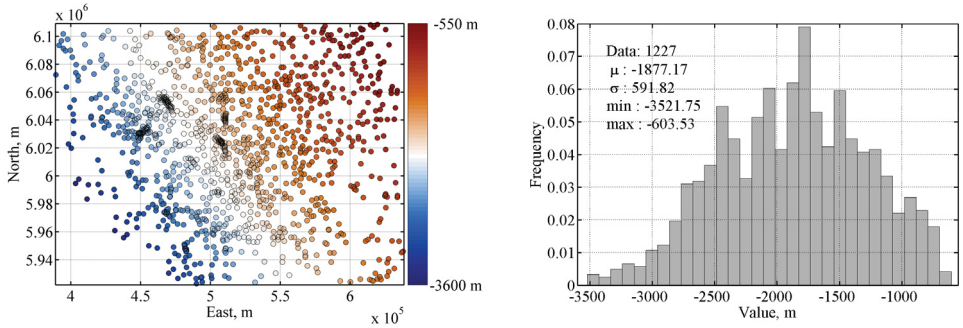


Figure 5.1: Overall workflow description for sub-surface mapping with and without a trend

5.1 Data Analysis

Data provided by Alberta Energy Regulator come from the sub-surface of the Blue Ridge Member inside the Graminia Formation located in West Central Alberta (Choquette, 1955; Switzer et al., 1994). The variable to be modeled is the

sub-surface top elevation. Data contains $1,227$ valid values. The location range of this variable is from $389,200$ to $638,000$ meters in the East and from $5,922,800$ to $6,108,600$ meters in the North. The location map and the histogram are shown in Figure 5.2. The value of the elevation gradually increases from the Southwest to the Northeast. The mean of the variable is $-1,877.17$ meters and the standard deviation is 591.82 meters.



(a) Location map **(b)** Histogram

Figure 5.2: Location map and histogram from $1,227$ data

A global simple kriging with a high nugget effect of 20%, and a Gaussian structure model with an isotropic range of $2,000,000$ meters is considered. The somewhat high nugget effect is used to filter the short-scale variability. Figure 5.3 shows a map of the global kriging result that considers all the data in each estimate. The map indicates a gradual increase from the Southwest to the Northeast. The most continuous direction is at an azimuth of 146°.

5.2 Data Preparation

Total $1,227$ data are divided randomly into a modeling set and a test set. Around 75% of the data, that is 827 data, are selected for conducting the geostatistical prediction and the remainder are considered for checking the simulated results. The prediction quality is determined by comparing the actual elevation values

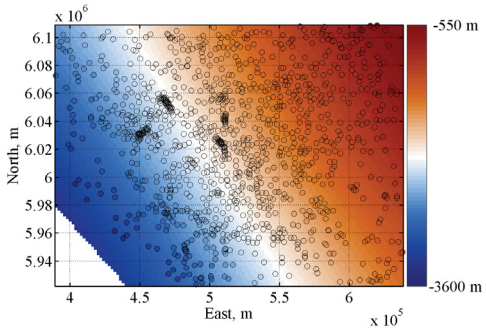


Figure 5.3: Global simple kriging from $1,227$ data

with the simulated values using 400 test data. The prediction quality criterion are considered by the uncertainty and the mean square error value.

Figure 5.4 shows the location map and the histogram from 827 modeling data, while Figure 5.5 shows the location map and the histogram from 400 test data. The statistics for these dataset are summarized in Table 5.1.

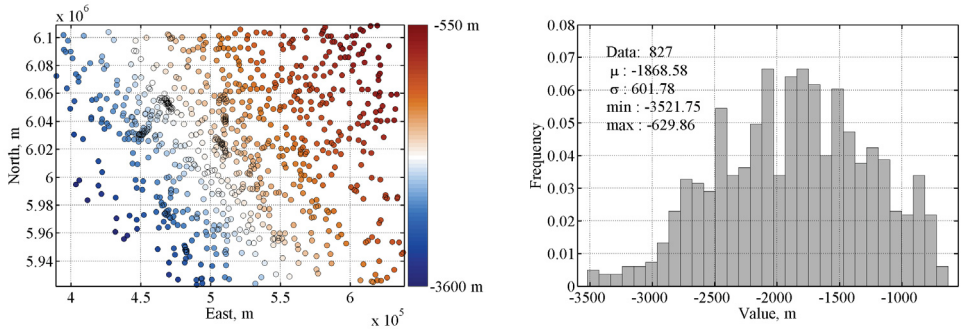
Name of Data	**Number of Data**	**Mean, m**	**Std. Dev., m**
Total Data	1,227	-1,877.17	591.82
Modeling Data	827	-1,868.58	601.78
Test Data	400	-1,894.93	571.02

Table 5.1: Summary statistics for each dataset

5.3 Conventional Geostatistical Modeling

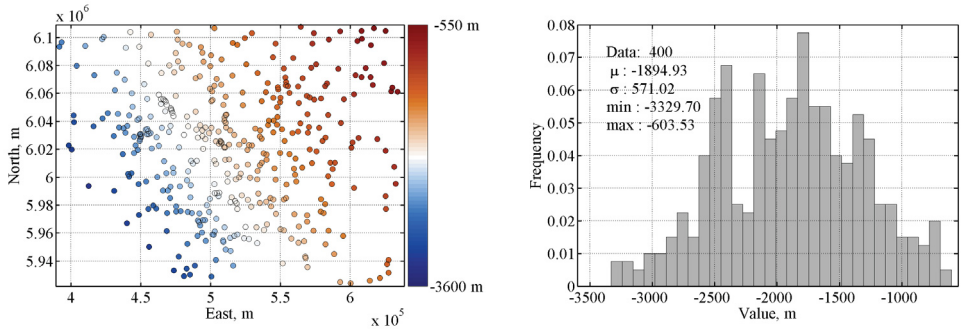
Conventional geostatistical modeling assumes that the attribute of interest over the domain is stationary. The conventional conditional simulation workflow is presented in Figure 5.1. The key points can be summarized as follows:

1. **Preprocess data**



(a) Location map **(b)** Histogram

Figure 5.4: Location map and histogram from 827 modeling data



(a) Location map **(b)** Histogram

Figure 5.5: Location map and histogram from 400 test data

- calculate cell declustering weights with 827 modeling data
- transform 827 modeling data into normal score units with declustering weights
- transform 400 test data into normal score units using 827 modeling data as a reference distribution

2. **Conduct sequential Gaussian simulation**

- calculate the experimental variograms in important directions using 827 modeling data
- define the principle direction of continuity

- fit the variogram model in the major and minor continuity directions
- generate many realizations using the normal transformed data
- transform all realizations back to original units
- check the histogram and variogram reproduction over realizations
- summarize the averaged simulated results

3. **Cross validate the numerical model**:

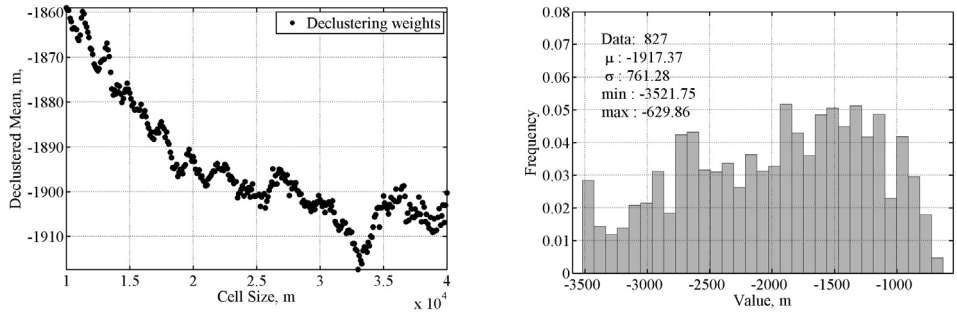
   - calculate the accuracy and the precision of the numerical model with 400 test data
   - check the mean squared error between the numerical model with 400 test data

Details of this conventional simulation workflow are described below.

5.3.1 Normal Score Transformation

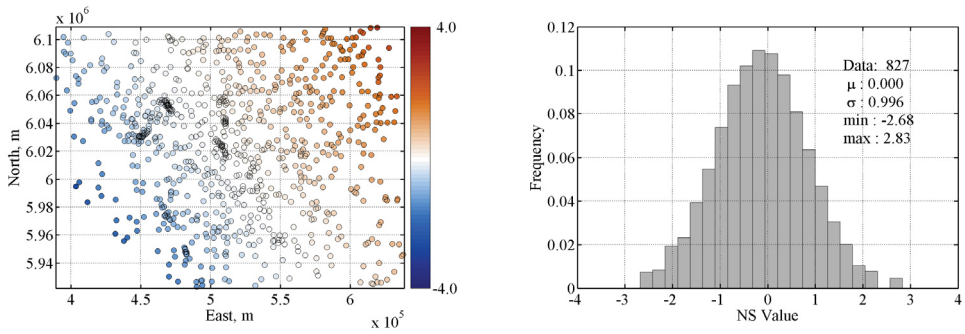
827 modeling data in Figure 5.4a shows some clustering and cell declustering is applied. Figure 5.6a shows a diagnostic plot of the declustered mean versus the cell size. The declustered mean reduces with increasing cell size. $33,000$ meters is used for defining the cell size when the declustered mean closes to the minimum value. Figure 5.6b shows the declustered histogram with a mean of $-1,917.37$ meters and a standard deviation of 761.28 meters.

Declustering weights are incorporated into the normal score transformation. The 827 modeling data are transformed into a standard Gaussian distribution with the defined declustering weights. The location map and the histogram with the transformed 827 modeling data are shown in Figure 5.7. The 400 test data are also normal transformed with the reference distribution from 827 modeling data. The location map and the histogram from 400 test data are shown in



(a) Cell declustering **(b)** Histogram with declustering

Figure 5.6: Histogram with cell declustering from 827 modeling data



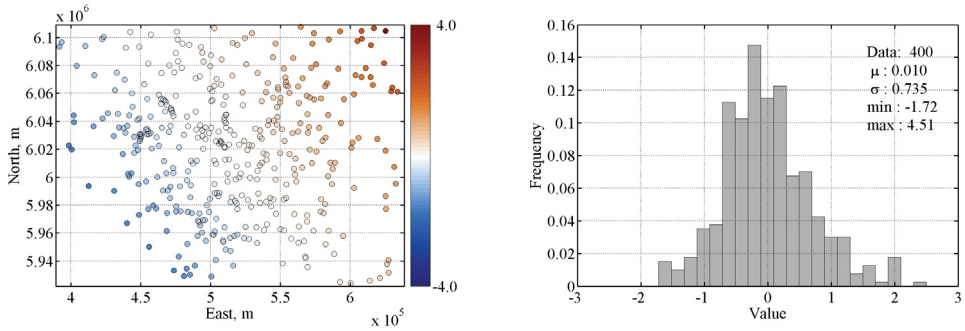
(a) Location map **(b)** Histogram

Figure 5.7: Location map and histogram from 827 modeling data in NS units

Figure 5.8. The mean and the standard deviation are 0.010 and 0.735, respectively.

5.3.2 Sequential Gaussian Simulation

A variogram model is required to capture the spatial variability of the normal score regionalized variable. The experimental variograms are calculated in the primary directions of 146° as the principle direction of continuity and 56° as the minor direction. The directional variograms are plotted with markers in Figure 5.9. The variogram shows a strong continuous structure, and a Gaussian type is considered to fit the experimental variograms. The variogram model is



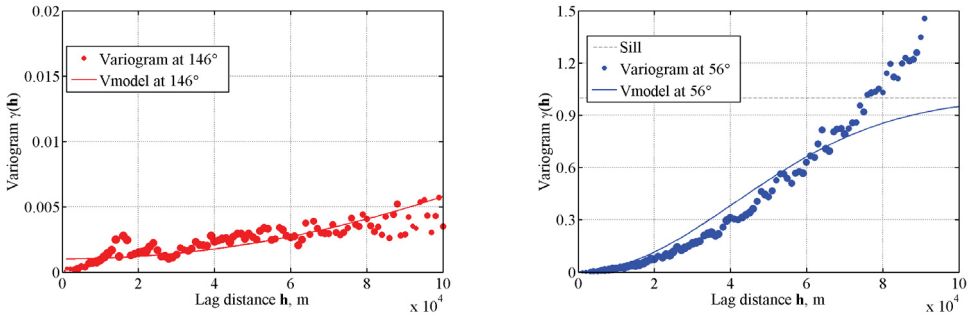
(a) Location map **(b)** Histogram

Figure 5.8: Location map and histogram from 400 test data in NS units

given by:

$$\gamma(\mathbf{h}) = 0.001_{\text{Nugget}} + 0.999 \cdot \text{Gauss} \underset{\substack{a_{\text{hmax}}=2{,}500{,}000 \text{ m} \\ a_{\text{hmin}}=100{,}000 \text{ m}}}{\alpha=146^\circ} \tag{5.1}$$

The stationary variogram is fitted with a unit sill. A small nugget effect is used to minimize the numerical instability common with the Gaussian-type variogram. This variogram model that is shown with lines in Figure5.9 will be considered in simulation.



(a) Variogram model in major direction **(b)** Variogram model in minor direction

Figure 5.9: variogram models from 827 modeling data in NS units.

A regular grid is established. The grid system is specified by the number of grid nodes (nx, ny), the center of the first grid node (xmin, ymin) and the size

of the grid nodes (xsize, ysize), see in Table 5.2. Note that no data are sampled in the Southwest areas. These areas will be omitted.

nx, xmin, xsize	312	389,400.0	800.0
ny, ymin, ysize	234	5,922,400.0	800.0

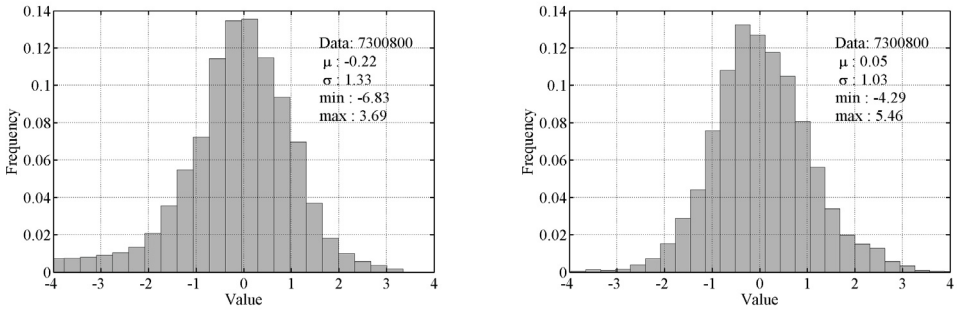
Table 5.2: Grid nodes specification for simulation

Sequential Gaussian simulation is a widely used geostatistical technique for modeling continuous variables. Simple kriging is employed in the simulation. The minimum search data is 8 and the maximum search data is 24 within a search ellipse. The search ellipse is defined based on the variogram ranges. A random path is defined. Many realizations are generated. Figure 5.10a shows the histogram of simulated values with the first 100 realizations. Note that the results do not exactly follow the standard normal score distribution.

The simulation parameters should be checked using the same parameters with unconditional simulation. Figure 5.10b shows the histogram of the unconditional simulated values with the first 100 realizations that closely follow the standard Gaussian distribution. The results indicate that the parameters in the conditional simulation are acceptable.

The first three realizations and the average simulated results over 100 realizations with the conditioning 827 modeling data are shown in Figure 5.11. The results show that the simulated values gradually increase from the Southwest to the Northeast. High variability at the edges is due to a lack of conditioning data.

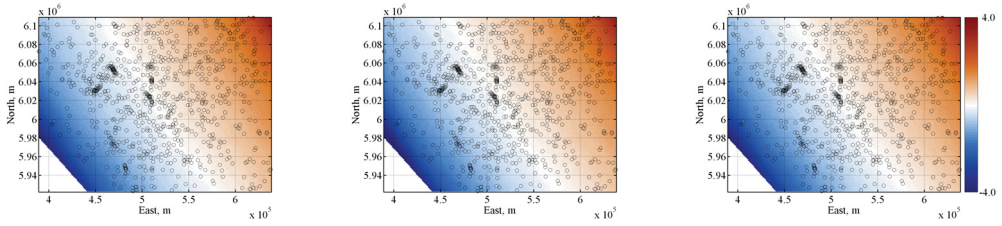
Normal score back transformation brings all realizations to original units. Figure 5.12 shows the results with the first three realizations and the average results over 100 realizations in original units. The simulated elevation values gradually change from Southwest to Northeast. Figure 5.12c shows the variance over 100 realizations. This plot shows strong artifacts. These artifacts exist



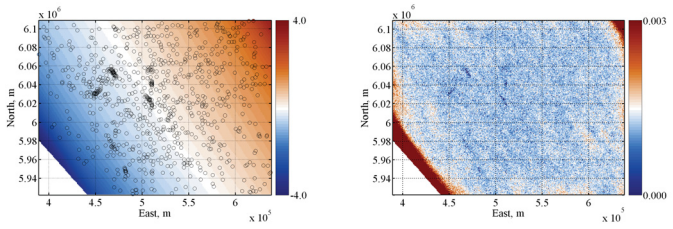
(a) Histogram with conditional simulation

(b) Histogram with unconditional simulation

Figure 5.10: Histogram from 100 realizations in NS units



(a) First three realizations

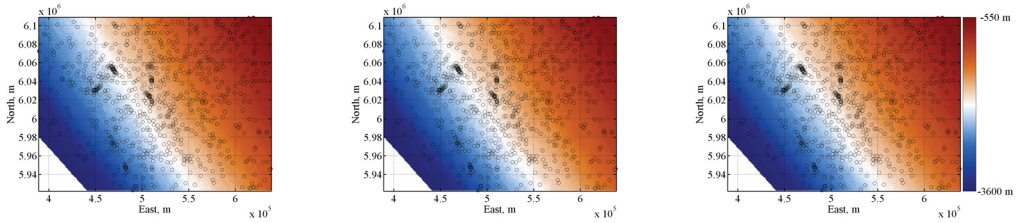


(b) Average values over 100 realizations

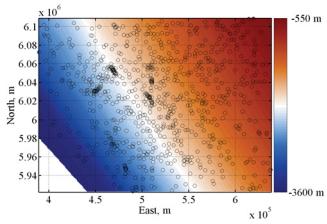
(c) Local variance over 100 realizations

Figure 5.11: First three realizations and average results over 100 realizations in NS units. 827 modeling data are shown with black circles

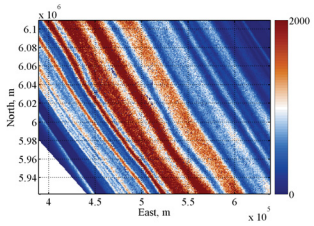
due to the transformation table from normal score units to original units. The transformation table does not provide a smooth and reasonable forward/back transformation. The transformed values are smooth in normal score units; however, the elevation values change substantially in original units due to small spikes and flat spots in the empirical distribution. A smooth transformation table must be considered.



(a) First three realizations



(b) Average values over 100 realizations



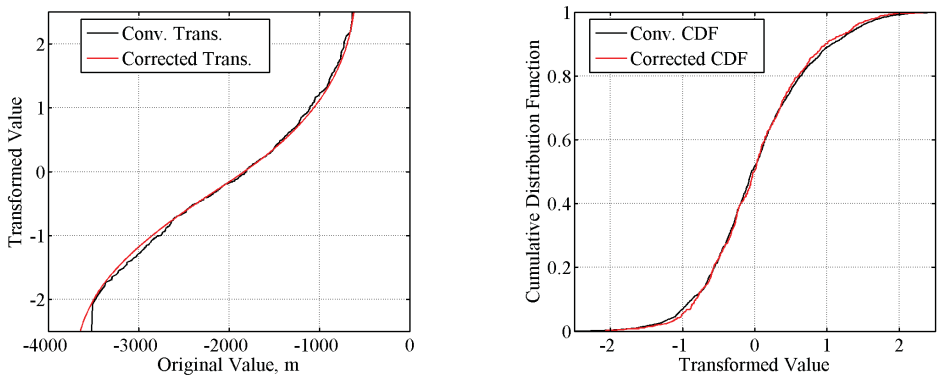
(c) Local variance over 100 realizations

Figure 5.12: First three realizations and average results over 100 realizations in original units. 827 elevation values are shown with black circles

5.3.3 Transformation Correction

Normal score transformation is a quantile-to-quantile transformation that transforms the original elevation values to a standard Gaussian distribution. The transformation table is used for viewing the original values and the corresponding normal score values. The black line in Figure 5.13a shows the transformation table from the conventional normal score transformation. It shows obvious

abrupt jumps that leads to a high variance in the transformation. It is the reason for the zebra-like artifacts shown in Figure 5.12c.

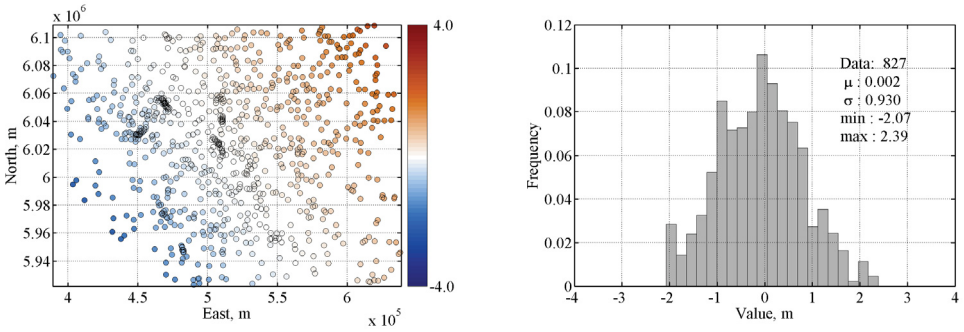


(a) Transformation table **(b)** Cumulative distribution

Figure 5.13: NS transformation table comparison from 827 modeling data

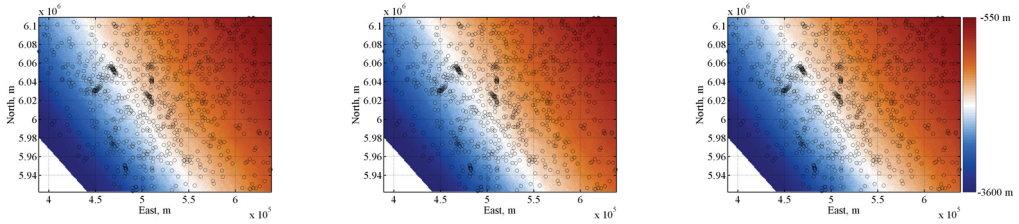
Silva and Deutsch (2015) have introduced a tool for estimating the univariate cumulative distribution function that allows to generate a smooth empirical distribution. A normal score transformation with a kernel density estimation is considered to correct the transformation table. The red line in Figure 5.13a shows a smooth transformation table so that the variable could be transformed continuously. Figure 5.13b shows the cumulative distribution function of the elevation value after the correction in normal score units. The black line is the cumulative distribution from the conventional normal score transformation, while the red line performs the cumulative distribution function using the kernel density estimation. A smooth cumulative distribution is generated if the kernel density estimation is applied.

Figure 5.14 shows the location map and the histogram from 827 modeling data in normal score units with a transformation correction. All procedures of the conventional simulation are conducted with the corrected normal score variable. The first three realizations, the average simulated elevation values and the

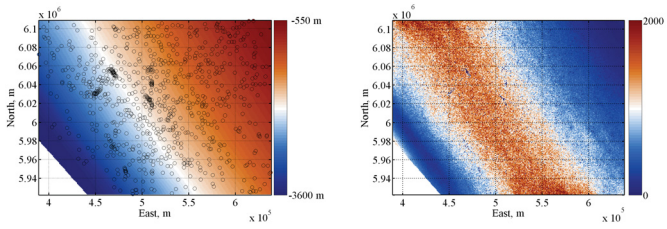


(a) Location map **(b)** Histogram

Figure 5.14: Location map and histogram from 827 modeling data with a transformation correction in NS units



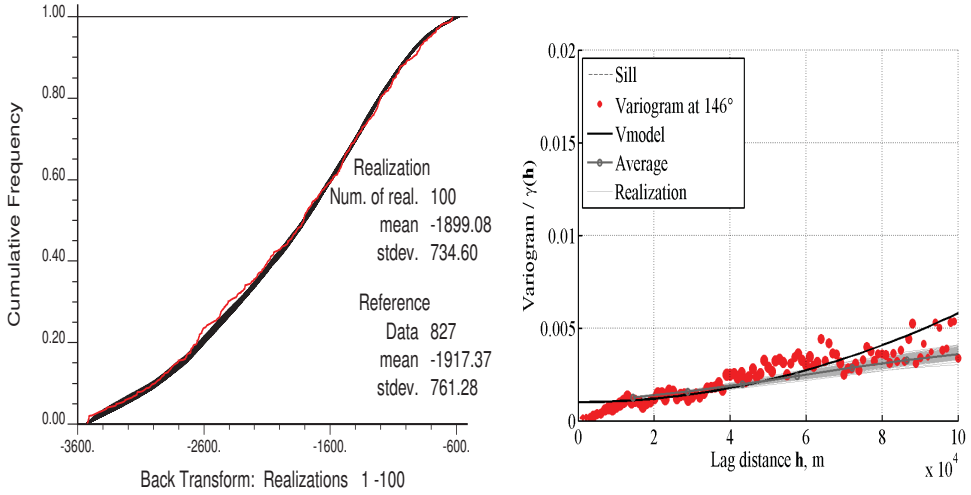
(a) First three realizations



(b) Average values over 100 realizations **(c)** Local variance over 100 realizations

Figure 5.15: Corrected first three realizations and average results over 100 realizations in original units

variance with the first 100 realizations in original units are shown in Figure 5.15. No obvious artifacts from the average variance are shown in Figure 5.15c. The elevation values in the middle area contain more uncertainty than the values in the Southwest and Northeast. The existence of the high variation in the middle is due to the incomplete smoothing of the distribution.



(a) Histogram reproduction **(b)** Variogram reproduction

Figure 5.16: Histogram and variogram reproduction with the conventional geostatistical modeling in original units

The histogram of the first 100 realizations is shown in Figure 5.16a. The black lines represent 100 realizations, and the red line indicates 827 weighted conditioning data. The histogram reproduction in the middle value zone is slightly different from the conditioning data. The mean of the realizations, $-1,899.08$ meters, is close to the reference mean, $-1,917.37$ meters; the standard deviation, 734.60 meters, is slight lower than the conditioning variance, 761.28 meters. Overall, the reproduction of the distribution seems reasonable.

Figure 5.16b checks the spatial correlation from the variogram model. The experimental variogram model in the direction of 146° from 827 modeling values is plotted with red points. Light gray lines are the variograms of each realization

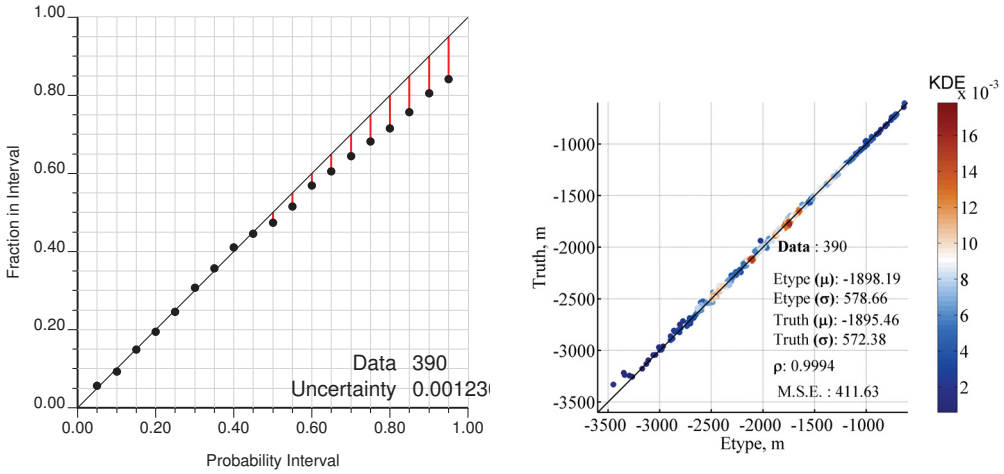
and the dark gray dash-dot line represents the average of the first 100 realizations. The variogram model from realizations is well reproduced the expectation variogram from 827 modeling data (the black line).

5.3.4 Cross Validation

Two tests are run to check the numerical model versus the true values at the 400 test data locations. The simulated values at the grid nodes closest to the test data are extracted from the gridded realizations. 2 data out of the 400 test data fall into the same grid cell and another 8 data fall into the same grid cell as one of the 827 modeling data. Finally, 390 test data are considered for the cross validation.

The first validation is to compare the true data and the simulated average values in normal score units. The accuracy plot in Figure 5.17a checks the reliability of the probabilistic estimates. The accuracy is checked by calculating the number of values that fall within an interval versus the probability of being in that interval (Deutsch, 2010b). The dots in intervals fall on the 45° line represent the predictions are accuracy and precision. The predictions are accurate but not precise if the dots fall above the 45° line, while the predictions are neither accurate nor precise if the dots fall below the 45° line. The result shows that the uncertainty represented in simulated values is close to the truth for the first half of the probability intervals and underestimated for the second half of the intervals. The uncertainty is summarized by an average variance of 0.00123 over 100 realizations. The accuracy and the precision are 0.211 and 0.997.

The second validation is to compare the true values and the average of the simulated values in original units. The mean squared error comparison between the true values and the simulated values is shown in Figure 5.17b. The mean of the model, $-1,898.19$ meters, is close to the true mean of $-1,895.46$ meters.



(a) Accuracy plot **(b)** Mean squared error

Figure 5.17: Accuracy plot in NS units and the mean squared error comparison with the test data in original units

The mean square error between the true values and the simulated average values is 411.63 square meters at the 390 test data locations.

5.4 Developed Geostatistical Modeling - Gaussian Mixture Model

The assumption of stationarity is relaxed. The trend is incorporated into the geostatistical prediction and the non-stationary geostatistical method is considered. The developed geostatistical method is presented in Figure 5.1. The major points of the developed method can be summarized as follows:

1. **Model the trend**
    - define the principle direction of the continuity from 827 modeling data
    - calculate the averaged anisotropy ratio
    - model the trend with increasing smoothing parameter
    - find the optimal trend model

2. **Reconstruct data**

    - assemble the input data including 827 modeling data and the optimal trend model in regular grids
    - transform the exhaustive optimal trend model into normal score units
    - calculate cell declustering weights from 827 modeling data
    - transform 827 modeling data into normal score units with the declustering weights
    - transform 400 test data into normal score units using the 827 modeling data as a reference distribution

3. **Define the conditional transformation**

    - fit the Gaussian mixture model between the variable and the trend model in normal score units
    - transform the variable according to the Gaussian mixture models

4. **Conduct sequential Gaussian simulation**

    - fit the variogram model with the directions of anisotropy
    - generate many realizations with the transformed variable
    - transform many realizations back to original units
    - check the histogram reproduction and variogram reproduction over realizations
    - summarize the averaged simulated results from realizations

5. **Cross validate the numerical model**:

    - check the accuracy and precision of the distributions of uncertainty with true test data

- check the mean squared error between the simulated averages and true test value

The steps with the stepwise conditional transformation using Gaussian mixture models are described below.

5.4.1 Trend Modeling

The location map in Figure 5.2 and the global kriging in Figure 5.3 show a very strong trend with high values concentrated in the Northeast and values decreasing to the Southwest. Further evidence of an existence of a trend comes from the crossplot of the East coordinate and the elevation values shown in Figure 5.18. The crossplot is fitted by a linear regression model with a black line. The map shows gradual changes from low values to high values and a clear trend is shown.

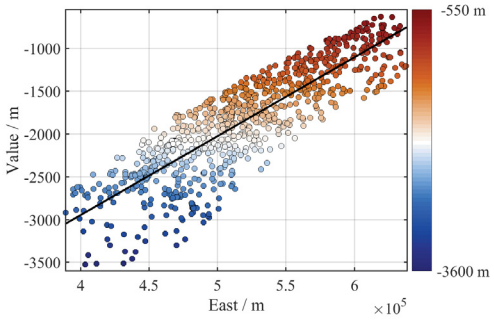
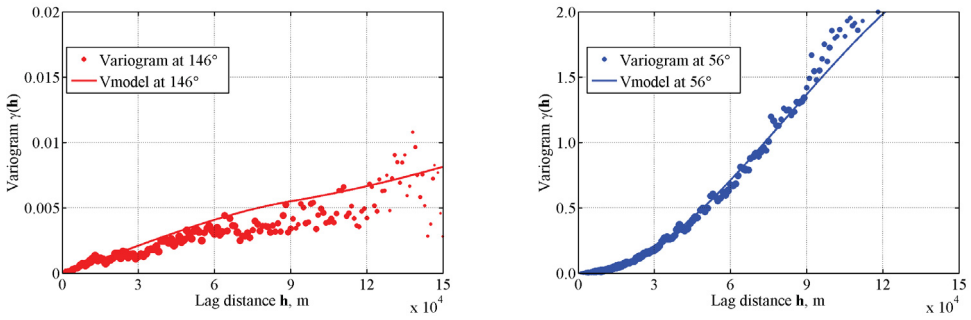


Figure 5.18: Crossplot of the location with the elevation values fitted with a linear regression model

5.4.1.1 Averaged Anisotropy Calculation

827 data are used to model the variogram. The anisotropy direction should follow the greatest continuity direction of the variable that is 146° in the horizontal direction. The proposed range of the variable is determined based on



(a) Variogram model in major direction **(b)** Variogram model in minor direction

Figure 5.19: Variogram models with non-stationary features from 827 modeling data

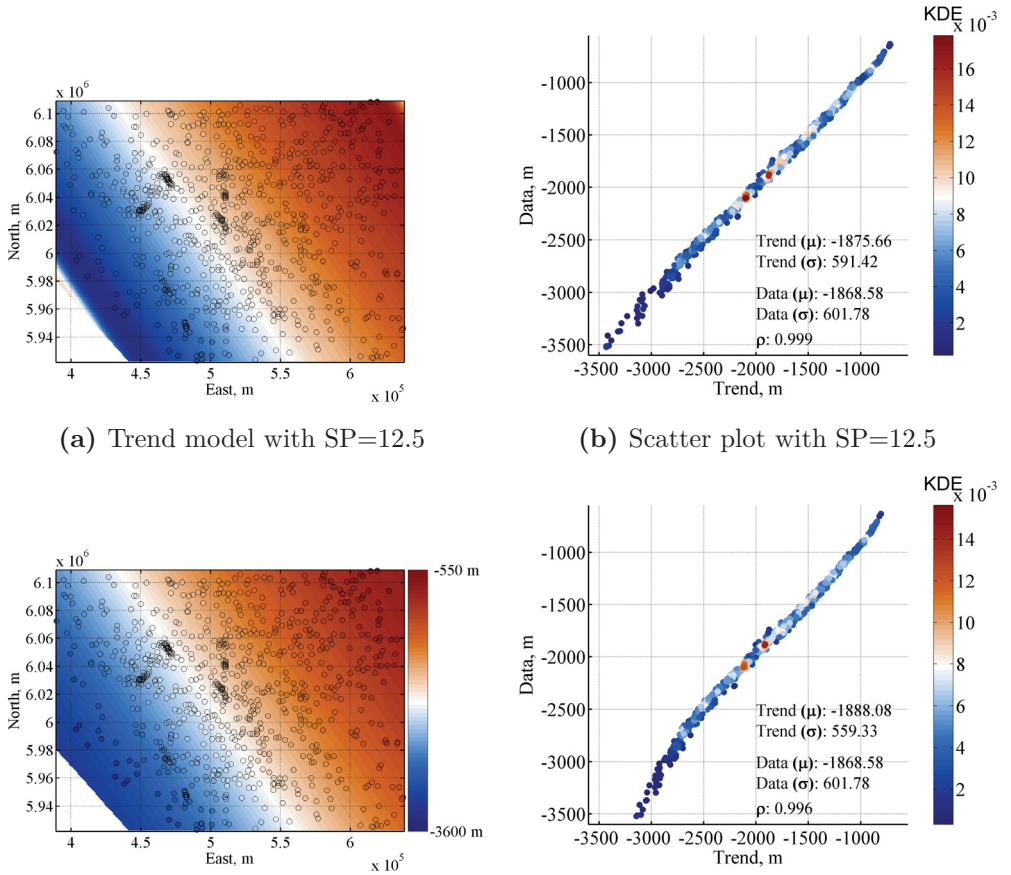
the experimental variogram within a half of the domain size. Variogram models with non-stationary features in an azimuth angle of 146° is given as:

$$\gamma(\mathbf{h}) = 0.004 \cdot \text{Sph}_{\substack{\alpha=146^\circ \\ a_{\text{hmax}}=90{,}000 \text{ m} \\ a_{\text{hmin}}=200{,}000 \text{ m}}} + 1.999 \cdot \text{Gauss}_{\substack{\alpha=146^\circ \\ a_{\text{hmax}}=7{,}000{,}000 \text{ m} \\ a_{\text{hmin}}=200{,}000 \text{ m}}} \tag{5.2}$$

Figure 5.19 shows the variogram models with non-stationary features from 827 modeling data fitting within a half of the domain size. The anisotropy ratio based on areas under the variogram curves is around 247 : 1. This anisotropy ratio is extremely high, but it is more robust than the range ratio.

5.4.1.2 Trend Construction and Optimization

Several parameters are required for trend modeling. The first parameter is the background value. A small amount of background value 0.001 is considered to ensure the computational stability. The second parameter is the anisotropy direction and ratio. The anisotropy model should be followed by the variogram models with non-stationary features in Figure 5.19 showing a direction of 146° and a ratio of 247. The third parameter is the smoothing parameter that controls the smoothness of the trend model. Several trends are modeled with increasing



(a) Trend model with SP=12.5

(b) Scatter plot with SP=12.5

(c) Trend model with SP=58.5

(d) Scatter plot with SP=58.5

Figure 5.20: Modeled trends and crossplots with different smoothing parameters. 821 elevation values are shown with trend models with black circles

smoothing parameter. Figure 5.20 shows trend models and the scatter plots of the trend models and 827 data with smoothing parameters of 12.5 and 58.5, respectively. The trend models require some extrapolated values, see Northeast and Southwest corners in Figure 5.20a, that close to the global mean. The color bar in scatter plots represents the accurate state-of-the-art bivariate density estimation. The trend model becomes smoother with increased smoothing parameter.

The selection of an optimal trend for further geostatistical modeling is

required. The objective function considers the correlation between the trend and its residual data. According to the developed objective function, the optimal trend model is the one when the smoothing parameter is SP $= 2 \times 19.5 = 39$, shown in Figure 5.21. Figure 5.22 shows the optimal trend model with the smoothing parameter of 39.

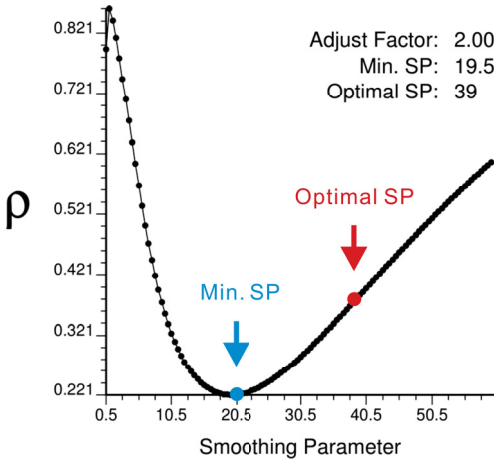


Figure 5.21: Determination of the optimal smoothing parameter

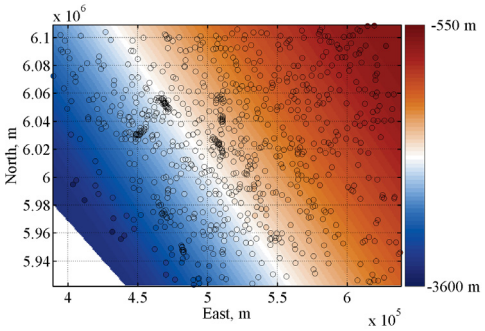
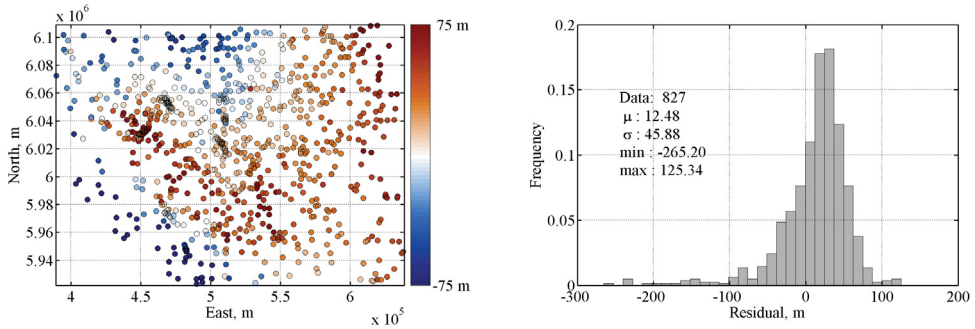


Figure 5.22: Optimal trend model with a smoothing parameter of 39. 827 elevation values are shown with black circles

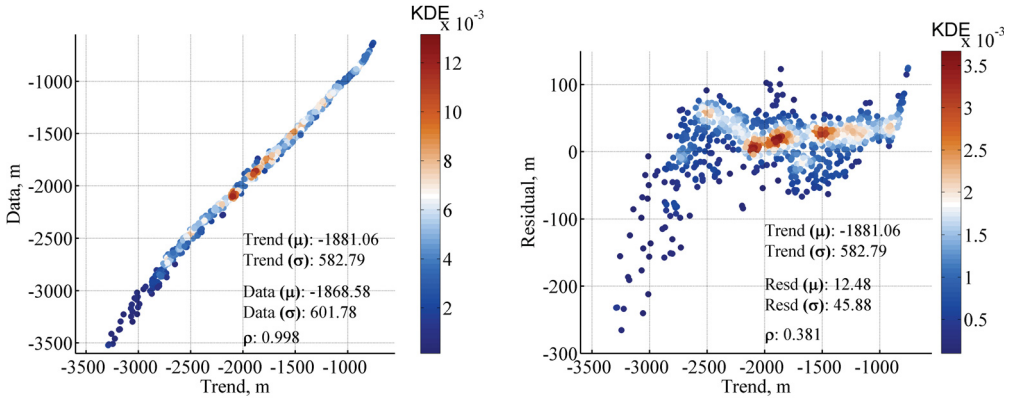
Figure 5.23a shows the residuals that remove the trend from 827 modeling data. The mean of the residuals is 12.48 meters and the standard deviation is

45.88 meters. The residual model reaches to a stationary model and there is much less trend in the residual model. Figure 5.24a shows the scatter plot between the data and the optimal trend with a high correlation of 0.998. Figure 5.24b shows the scatter plot between the optimal trend and the residuals with a correlation of 0.381.



(a) Residuals **(b)** Histogram

Figure 5.23: Visualization of location map and histogram from residuals

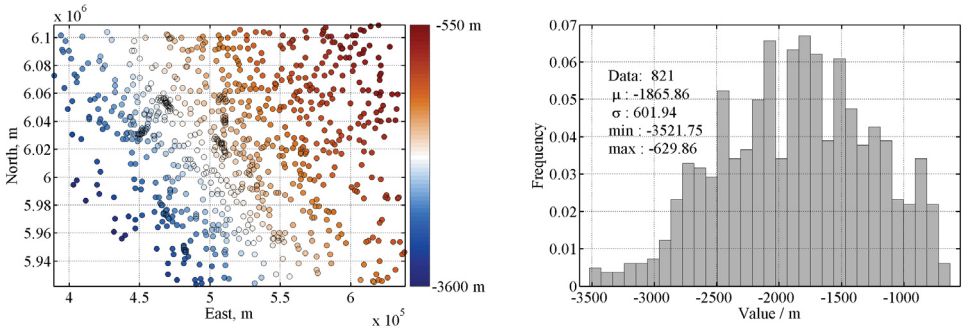


(a) Crossplot of the elevation values with trend model **(b)** Crossplot of the residual data with trend model

Figure 5.24: Crossplots of the observed data and residual data with trend values

5.4.2 Data Transformation

827 modeling data should be pre-processed so that they are assigned to grid nodes. The aim is to combine the 827 modeling data with the optimal trend for subsequent geostatistical modeling. The number of data is reduced from 827 to 821 due to multiple data falling into the same grid cell. The statistical analysis should be also re-calculated. The mean of the 821 modeling changes from $-1,868.58$ meters to $-1,865.86$ meters and the standard deviation changes from 601.78 meters to 601.94 meters.

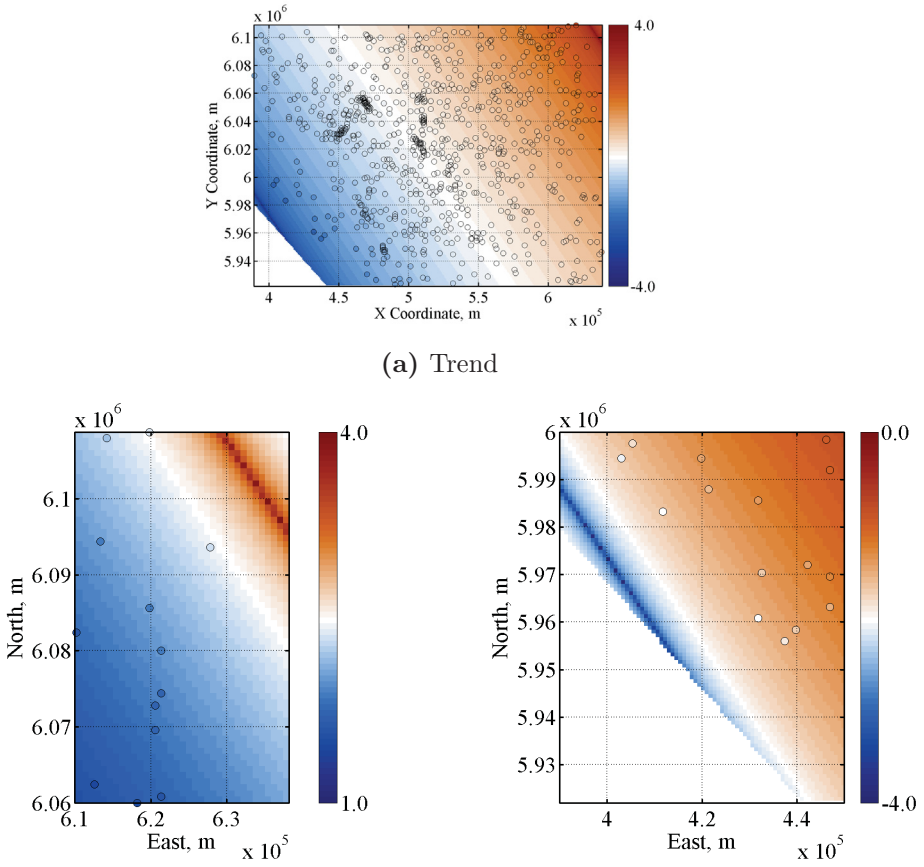


(a) Location map **(b)** Histogram

Figure 5.25: Location map and histogram from the re-constructed 821 modeling data in original units

The declustering weights are also re-calculated with the 821 modeling data. $33,000$ meters is kept for defining the cell size. The 821 modeling data are transformed into normal score units with declustering weights. The corrected transformation table based on a kernel density estimation is required to transform the 821 modeling data into normal score units.

The optimal trend is exhaustively normal transformed and is shown in Figure 5.26a. There is no need to correct the trend transformation table; however, the transformed trend shows some artifacts at the edge of the Northeast (displayed in Figure 5.26b) and the Southwest (displayed in Figure 5.26c). These artifacts exist due to the lack of conditioning data in these areas and trend values



(a) Trend

(b) Artifacts in Northeast

(c) Artifacts in Southwest

Figure 5.26: Trend model and artifacts in Northeast and Southwest in NS units. 821 elevation values are shown with black circles

come close to the global mean of the data. Figure 5.27 shows the crossplot of the trend with 821 modeling data in normal score units with a high correlation of 0.996. The 400 test data are also transformed into normal score units with a reference of 821 modeling data.

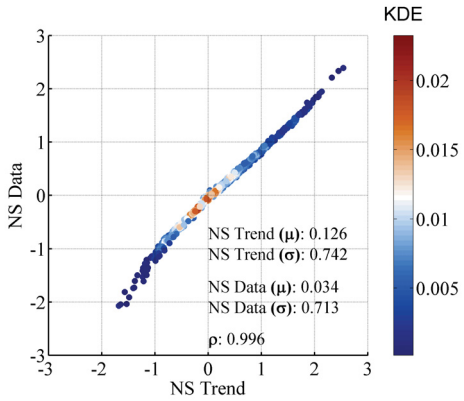
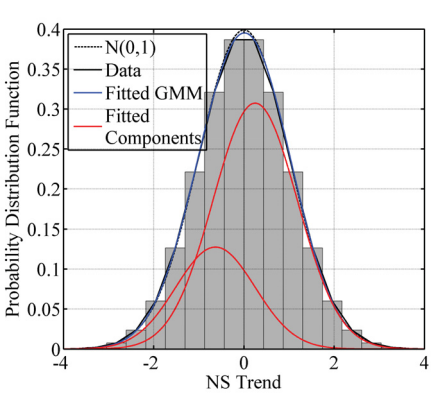


Figure 5.27: Crossplot of the trend values with 821 modeling data in NS units

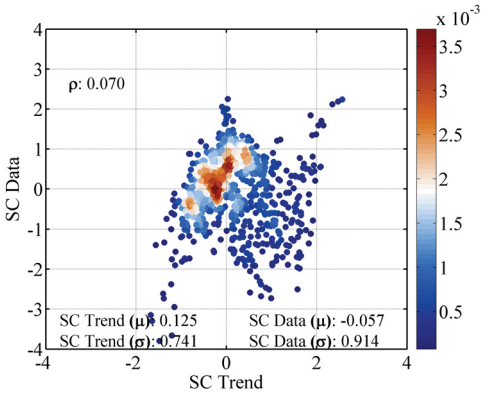
5.4.3 Stepwise Conditional Transformation with Gaussian Mixture Model

The developed method, stepwise conditional transformation with Gaussian mixture models, is considered. The Gaussian mixture model fits to crossplot of the trend with 821 modeling data in Figure 5.27. Two components are considered to fit the scatter plot. The univariate distribution of the normal score transformed trend is shown in Figure 5.28a that is closely fitting the mixture models. The univariate distribution of the normal score transformed data is shown in Figure 5.28d that is nearly a standard Gaussian distribution. The bivariate distribution is shown on a 2-D density estimation plot in Figure 5.28c. Transformed variables of the trend and the data in stepwise units are uncorrected, 0.070, in Figure 5.28b.

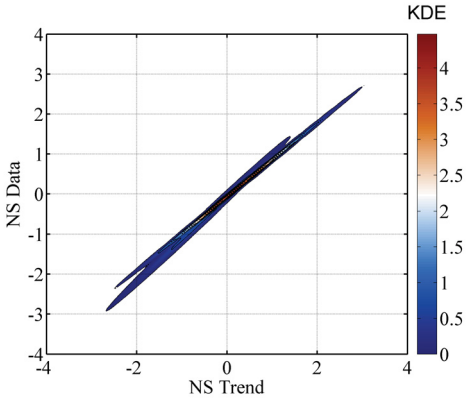
After a stepwise conditional transformation, the transformed 821 modeling data are shown in Figure 5.29a. The large-scale trend-like features are removed, and some spatial structures, such as high values in the center, still exist. The variogram model of this transformed data is shown in Figure 5.29b with an



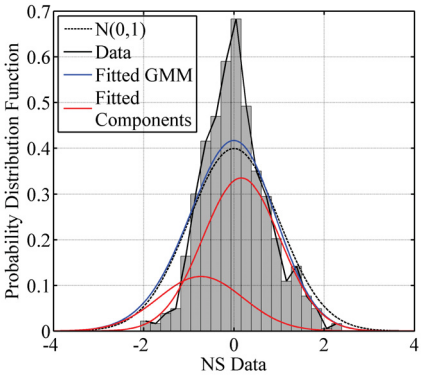
(a) Univariate distribution of the exhaustive trend



(b) Scatter plot of transformed variable



(c) Bivariate distribution

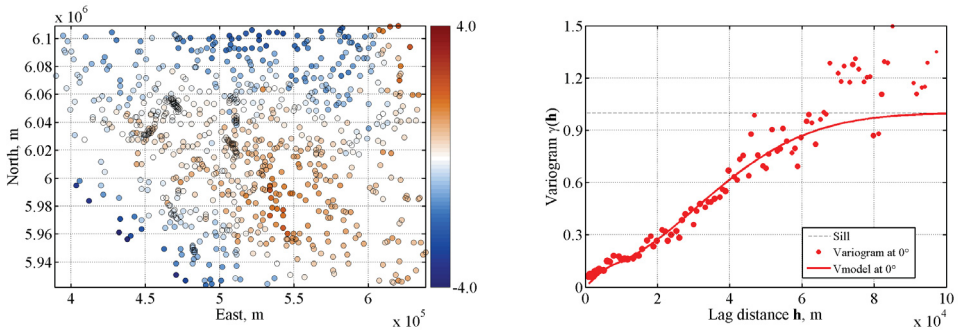


(d) Univariate distribution of 821 data

Figure 5.28: Bivariate and univariate marginals of the Gaussian mixture model fitted to 821 modeling data

isotropic variogram model given by:

$$\gamma_{\text{direct}}(\mathbf{h}) = 0.001_{\text{Nugget}} + 0.099 \cdot \text{Sph}\underset{\substack{a_{\text{hmax}}=8{,}000\text{ m}\\ a_{\text{hmin}}=8{,}000\text{ m}}}{{}^{\alpha=0^\circ}} + 0.900 \cdot \text{Gauss}\underset{\substack{a_{\text{hmax}}=75{,}000\text{ m}\\ a_{\text{hmin}}=75{,}000\text{ m}}}{{}^{\alpha=0^\circ}} \tag{5.3}$$



(a) Location map **(b)** Variogram

Figure 5.29: Location map and variogram model from transformed 821 modeling data in SC units

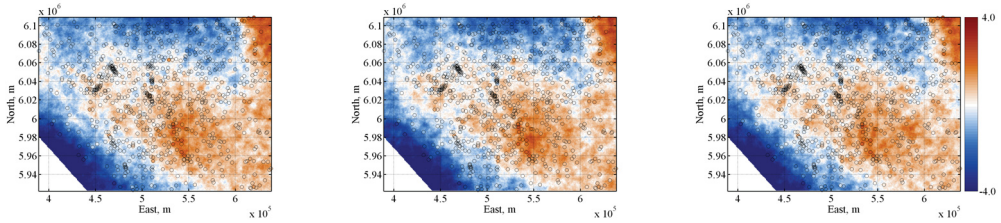
5.4.4 Sequential Gaussian Simulation

Sequential Gaussian simulation is applied on the transformed variable. The grid system is defined in Table 5.2. The minimum and maximum search data are 8 and 24, respectively. The search ellipse is defined based on the variogram ranges in Equation 5.3.

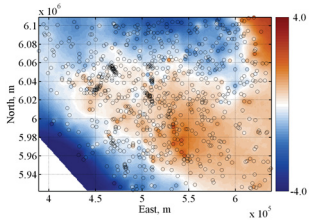
Many realizations are generated. Figure 5.30 shows the first three realizations, an average simulated map and the variance overs the first 100 realizations. High values that are concentrated in the center are shown in the average realization. The high variance is located at the edge of the domain due to lack of conditioning data.

A stepwise conditional back transformation with Gaussian mixture models is then conducted. Data would be back transformed conditional to the trend. Figure 5.31 shows the first three realizations, an average simulated map and an average variance over the first 100 realizations. The average realization performs a smooth result, and the average variance over 100 realizations shows some artifacts in the Northeast/Southwest and West areas.

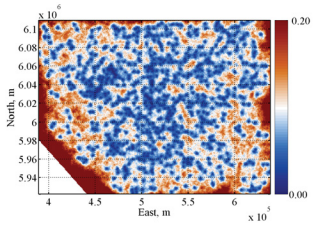
Let us consider the Northeast corner first. The Northeast corner is shown



(a) First three realizations

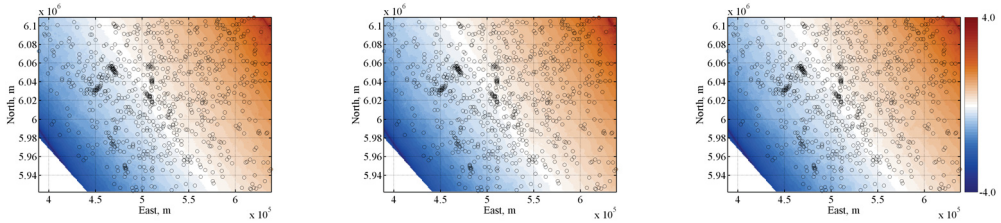


(b) Average values over 100 realizations

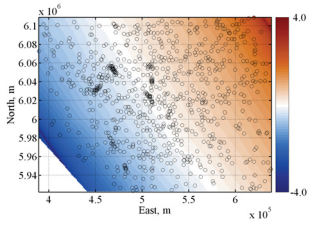


(c) Local variance over 100 realizations

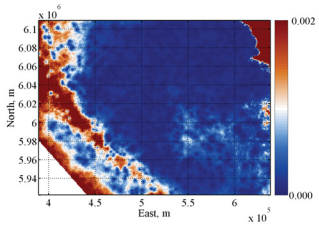
Figure 5.30: First three realizations and average results over 100 realizations in SC units



(a) First three realizations

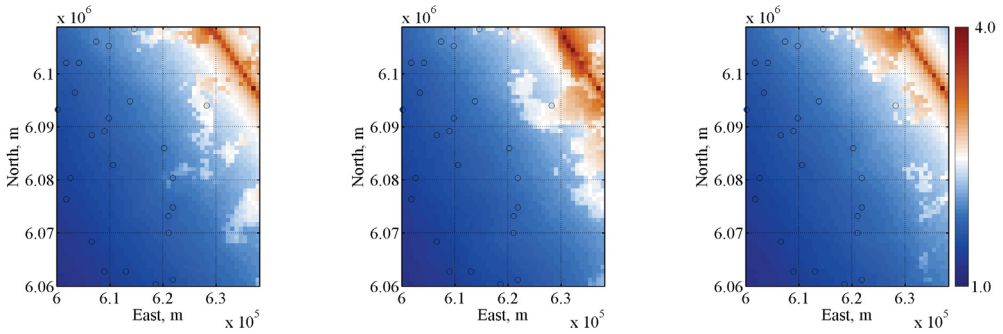


(b) Average values over 100 realizations

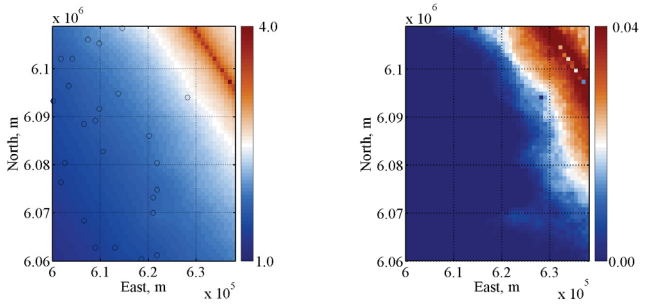


(c) Local variance over 100 realizations

Figure 5.31: First three realizations and average results over 100 realizations in NS units



(a) First three realizations



(b) Average values over 100 realizations

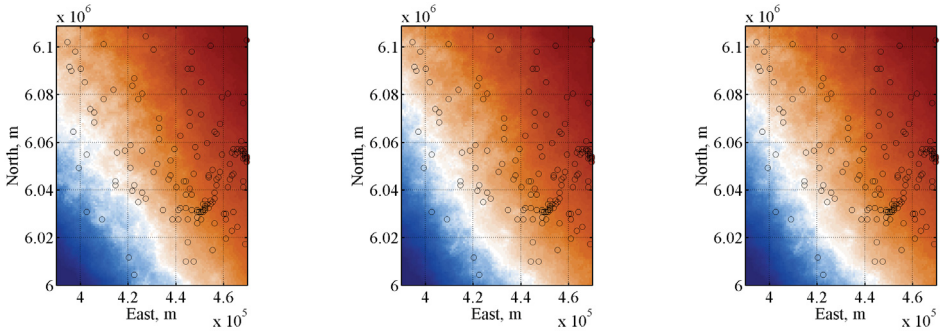
(c) Local variance over 100 realizations

Figure 5.32: Realizations and the average simulated results in the Northeast corner in NS units

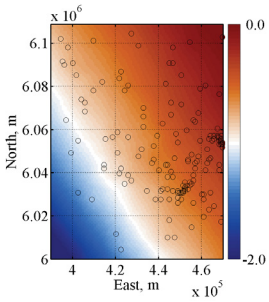
in Figure 5.32 in normal score units. The upper figures are the first three realizations that show a visible artifact at the edge due to the artifacts in the trend in Figure 5.26b. The extrapolation of the trend in the Northeast affects the conditional transformation. All realizations are averaged, and the variance over 100 realizations shows a high variability at the Northeast corner. Such artifact also holds for the Southwest.

Let us now consider the West corner. The West corner is shown in Figure 5.33 in normal score units. The upper figures are the first three realizations and the average results are shown in the bottom. The high variance comes from the bivariate distribution of the Gaussian mixture model in Figure 5.28c. Two

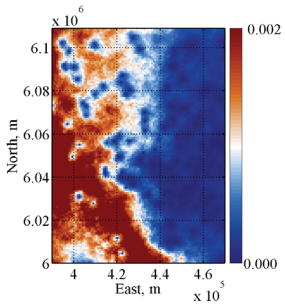
components of mixture models could not allow a good convergence and is over-fitting the distributions from the data in this example. Less components should be used in Gaussian mixture models.



(a) First three realizations



(b) Average values over 100 realizations



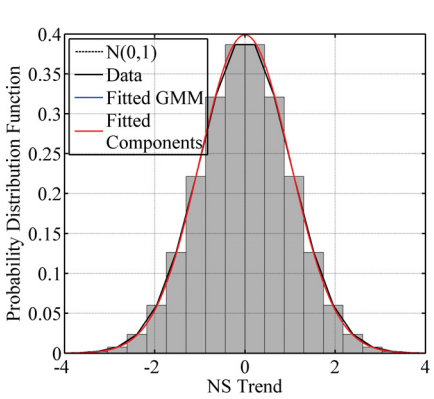
(c) Local variance over 100 realizations

Figure 5.33: Realizations and the average simulated results in the West corner in NS units

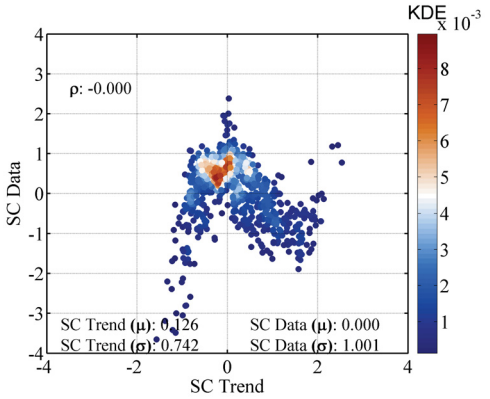
5.4.5 Gaussian Mixture Model Correction

One Gaussian component is considered to avoid the over-fitting in this example, although this does not take full advantage of the concept of a Gaussian mixture model. Figure 5.34a and 5.34d show the univariate distribution of the trend and

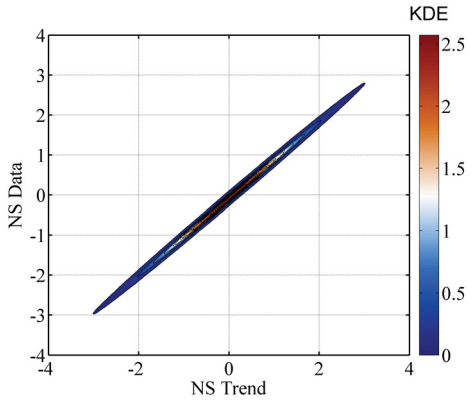
the data, respectively. Figure 5.34c shows the bivariate distribution of the Gaussian mixture model and Figure 5.34b shows the scatter plot of the transformed variable with a zero correlation.



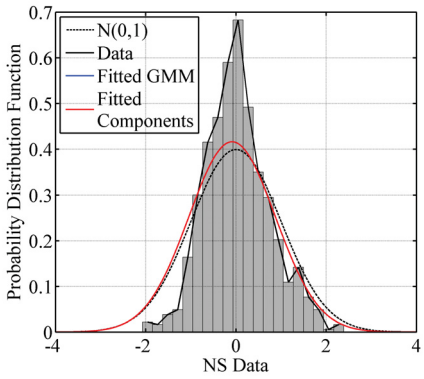
(a) Univariate distribution of the exhaustive trend



(b) Scatter plot of transformed variable



(c) Univariate distribution of the data

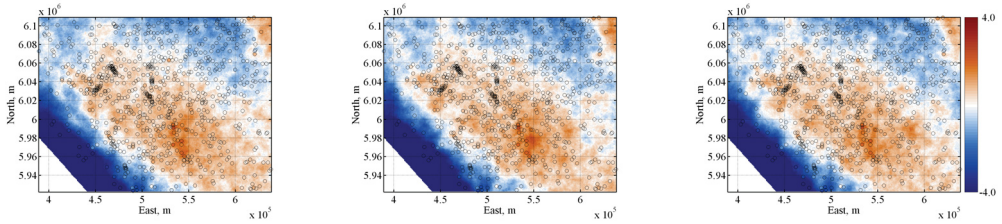


(d) Scatter plot of transformed variable

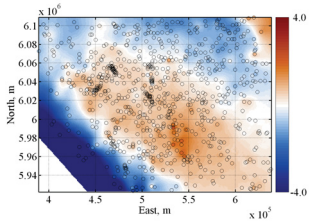
Figure 5.34: Corrected bivariate distribution of the Gaussian mixture model fitted to 821 modeling data

Sequential simulation is conducted on the stepwise transformed data and many realizations are generated. Figure 5.35 shows the simulated results. Two back transformations would apply on 100 realizations. The first back transformation is the stepwise back transformation. Simulated results would be back

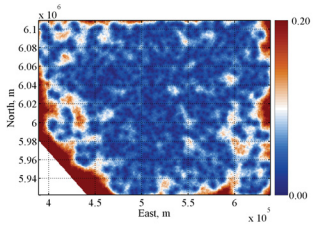
transformed with the Gaussian mixture model. Figure 5.36 shows the first three realizations, the average realization and the average variance over 100 realizations in stepwise units. The simulated map shows a smooth result. No obvious artifacts exist except in the Northeast and the Southwest corners due to the trend extrapolation. The second back transformation is the normal score back transformation. All realizations are transformed back into original units. Figure 5.37 shows the first three realizations and the averages over 100 realizations. The values gradually increase from Southwest to Northeast. No artifacts can be found in the average variance over 100 realizations.



(a) First three realizations



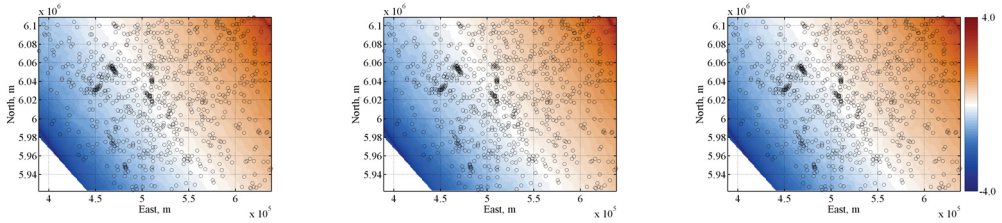
(b) Average values over 100 realizations



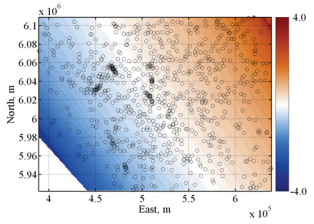
(c) Local variance over 100 realizations

Figure 5.35: First three realizations and average results over 100 realizations in SC units

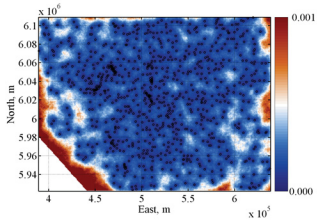
The reproductions of the histogram and the variogram are checked. Figure 5.38a shows histogram reproduction. The black lines represent the 100 realizations, while the red line represents the weighted conditioning data. The simulated values are well reproduced, although they are slightly lower than the conditioning data in middle quantiles and slightly higher in high quantiles. The mean



(a) First three realizations

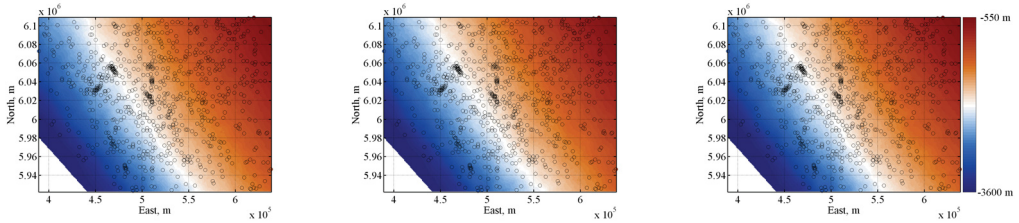


(b) Average values over 100 realizations

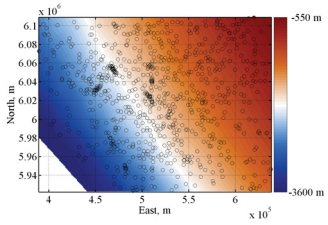


(c) Local variance over 100 realizations

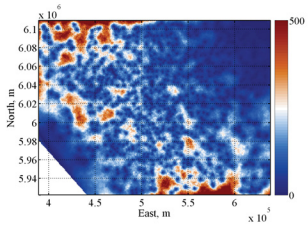
Figure 5.36: First three realizations and average results over 100 realizations in NS units



(a) First three realizations



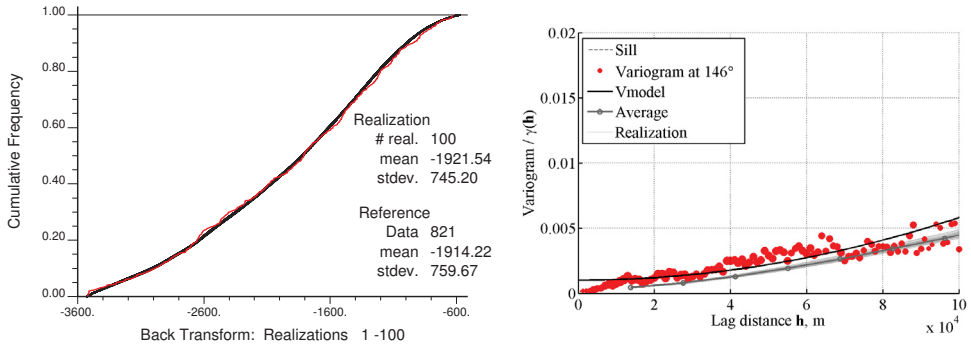
(b) Average values over 100 realizations



(c) Local variance over 100 realizations

Figure 5.37: First three realizations and average results over 100 realizations in original units

of the realizations, $-1,921.54$ meters, is close to the reference mean, $-1,914.22$ meters and the standard deviation, 745.20 meters, is close to the conditioning standard deviation, 759.67 meters. The realizations successfully reproduce the global mean/standard deviation and the global distribution of the conditioning data.



(a) Histogram reproduction **(b)** Variogram reproduction

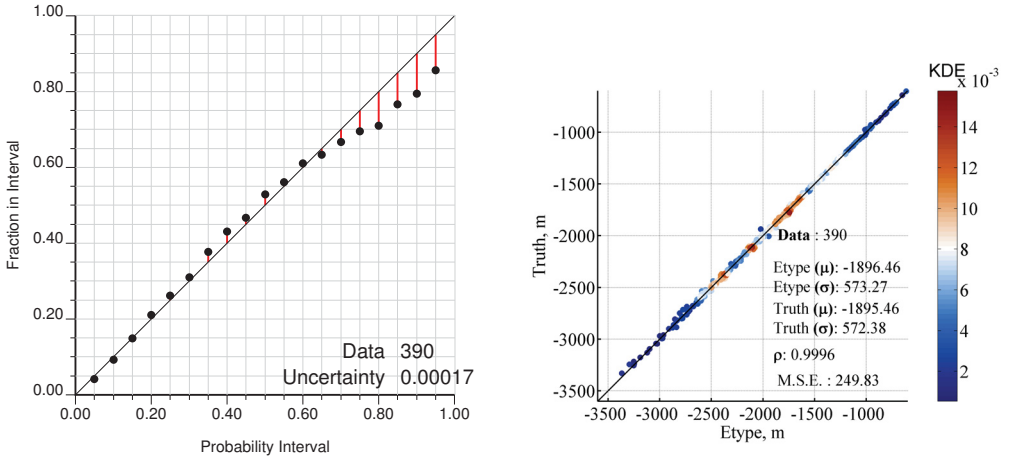
Figure 5.38: Histogram and variogram reproduction with the proposed geostatistical modeling in original units

The spatial correlation from the variogram model is shown in Figure 5.38b. The red markers represent the calculated experimental variogram in the most continuous direction. The black line is the expectation variogram model from the experimental variogram. The light gray lines indicate the variograms inferred from 100 realizations and the gray dash-dot line represents the average model over 100 variogram models. The spatial relationship is reproduced and the model has less uncertainty than the model from the conventional method.

5.4.6 Cross Validation

The first validation is to check the uncertainty in normal score units. 390 test data are used. Figure 5.39a shows that the probability intervals are slightly overestimated in the first-three-quartile of the intervals and become wider underestimated in the upper-quartile probability intervals. The uncertainty is a

relatively small value, 0.00017, indicating an improvement of 86.18% compared with the conventional method. The prediction with the proposed method also shows more accurate than the one with the conventional method.



(a) Accuracy plot **(b)** Mean square error value

Figure 5.39: Accuracy plot in NS units and mean squared error comparison with test data in original units

The comparison of the mean square error between the truth and the simulated values is shown in Figure 5.39b. The mean square error value is 249.83 square meters that is smaller than that in conventional simulation procedure, 411.63 square meters. The developed method shows a significant improvement.

5.5 Remarks

The goal of this case study is to demonstrate a geostatistical modeling framework to account for non-stationary features of regionalized variables. A conventional geostatistical modeling workflow and the developed geostatistical modeling workflow with trend modeling are implemented. The assumption of stationarity is made in the conventional geostatistical prediction. The kernel density estimation is involved to provide a smooth empirical distribution. The simulated average

shows a fair smooth map and the simulated average variance shows a high variance in the middle. The benefits of incorporating the trend into geostatistical modeling is clearly visible in the simulated results. The non-stationary features can be addressed with the technique by modeling the trend deterministically and inferring an optimal trend objectively. The stepwise conditional transformation with Gaussian mixture models is robust and is able to reproduce the complex features of the regionalized variable.

Prediction with a trend can be an important consideration for improving the predictions. This chapter shows an advantage of trend modeling as part of geostatistical modeling with the cross validation. The cross validation is used to guide the selection of parameters in the absence of independent validation data. The numerical model contains less uncertainty with the proposed method. The mean square error is much less than conventional geostatistical modeling; a 39.31% improvement is shown in this case study.

Chapter 6

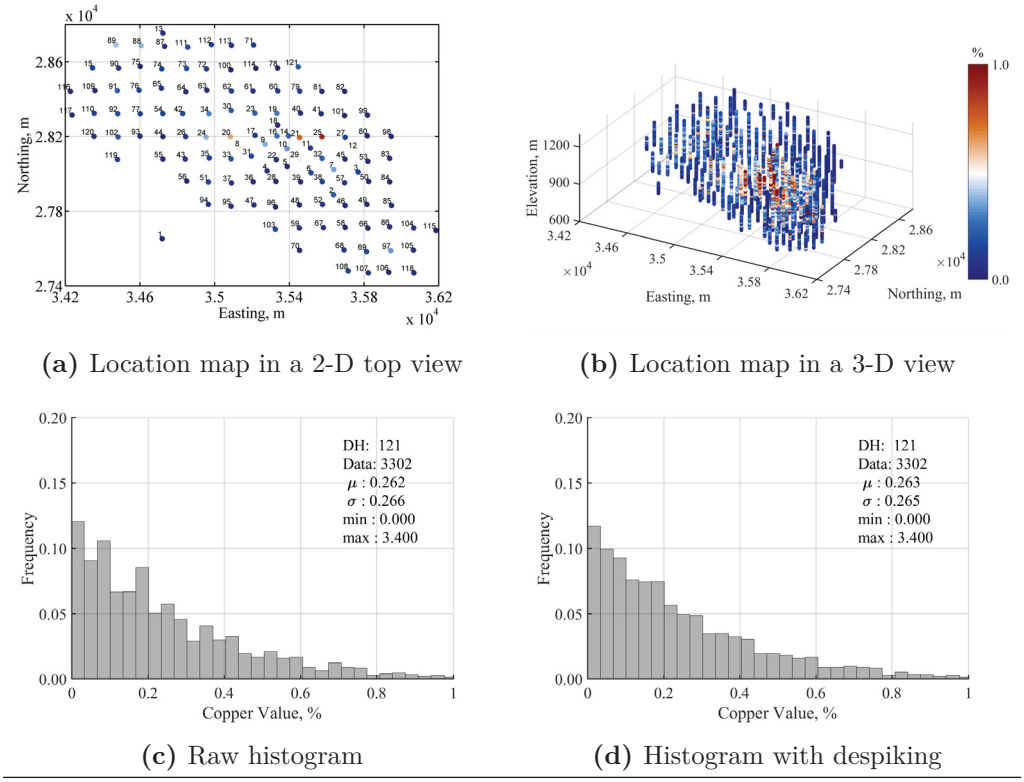
Case Study 2: 3-D Mapping

This case study shows the predictive performance of non-stationary geostatistical modeling in a 3-D realistic scenario. A porphyry copper deposit with $3,302$ data is considered where the grades show a trend. Trend detection methods are discussed. The trend modeling technique is applied to establish an optimal trend. Comparisons to conventional geostatistical calculations are made. The results show that the proposed geostatistical modeling with trend modeling outperforms the conventional geostatistical modeling techniques with less error and better reproduction of essential features of the regionalized variable.

6.1 Data Analysis

The data shown in Figures 6.1a (2-D top view with drillhole ID) and 6.1b (3-D view) contain 121 drillholes with $3,302$ grade measurements from a porphyry copper deposit. The location coordinates range from $34,200$ to $36,200$ meters in the East, from $27,400$ to $28,800$ meters in the North, and from 600 to $1,300$ meters with an interval of 9 meters in the Elevation. The value of the copper ranges from 0.0% to 3.4% with a mean of 0.262% and a standard deviation of 0.266%. The histogram is shown in Figure 6.1c. Some copper grades are

constant values at the detection limit and due to the number of decimal places in the database. Such ties of constant values should be broken so that the variogram calculation and the normal score transformation work properly (Rossi and Deutsch, 2014). Despiking of the data is considered and the histogram after the despiking is shown in Figure 6.1d with a mean of 0.263% and a standard deviation of 0.265%.



(a) Location map in a 2-D top view

(b) Location map in a 3-D view

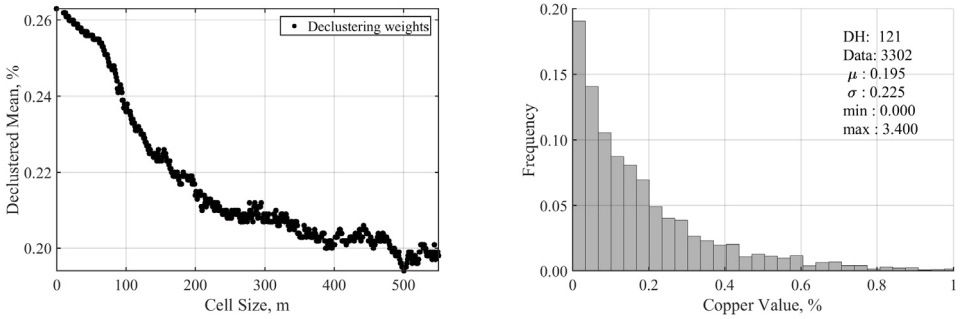
(c) Raw histogram

(d) Histogram with despiking

Figure 6.1: Location maps and histograms from 3,302 data

More samples are taken in the high grade center area in Figure 6.1 and the equal-weighted statistics are not representative of the entire population. Declustering should be considered so that the distribution of the regionalized variable is representative. Figure 6.2a shows the diagnostic plot of the cell size against the declutered mean. The decluster reduces with increasing the cell size. A

500-meter cell-size is determined for the cell declustering where it reaches to the minimized declustered mean. The corrected histogram from 3, 302 data is shown in Figure 6.2b with a mean of 0.195% and a standard deviation of 0.225%.



(a) Cell-size declustering **(b)** Histogram with declustering

Figure 6.2: Histogram with the cell declustering from 3, 302 data

6.2 Trend Modeling

Spatial trends are not consistent with the assumption of stationarity that the conventional geostatistical modeling considers. The spatial trend should be detected before the geostatistical modeling. After deciding to model a trend, an artifact-free trend should be modeled to investigate the performance of the non-stationary geostatistical modeling.

6.2.1 Trend Detection

Visualization of the data is the simplest way to determine if the attribute of interest is stationary. Figures 6.1a (2-D top view) and 6.1b (3-D view) show the location maps from 3, 302 data. Drillholes with high values are concentrated in the center of the domain and values are decreased to the edge. Such gradually change of values is an evidence of a trend.

The regression analysis from the data can be also considered for the trend detection. Figure 6.3 shows the crossplot of the data against the elevation on a 2-D density estimation plot. A quadratic regression model that is accomplished by the least square method is shown with a black line where a change in the local mean is noticed.

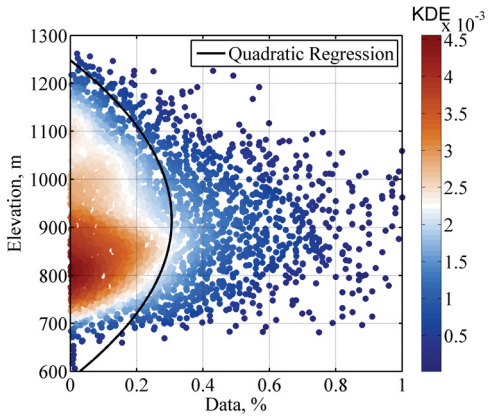


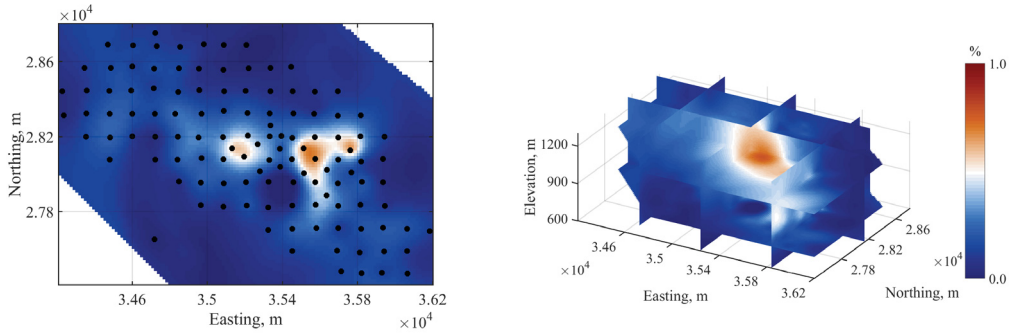
Figure 6.3: Visualization of the trend by the quadratic regression model from 3,302 data

Global kriging is another way to map the large-scale trend-like features. A global kriging is performed with an isotropic variogram with a 20% nugget effect and a range of 1,000 meters. The variogram model is summarized as:

$$\gamma(\mathbf{h}) = 0.20_{\text{Nugget}} + 0.80 \cdot \text{Sph}(\mathbf{h}) \begin{array}{l} \alpha{=}0^\circ \\ a_{\text{hmax}}{=}1{,}000 \text{ m} \\ a_{\text{hmin}}{=}1{,}000 \text{ m} \\ a_{\text{hvert}}{=}1{,}000 \text{ m} \end{array} \tag{6.1}$$

The global kriging result in Figure 6.4 reveals an obvious trend where high values are in the center and lower grades toward the edges. The most continuous direction is at an azimuth of 110° in the horizontal plane and at a dip of 90° in the vertical direction.

The additional evidence of the trend comes from the variogram. The standardized experimental variograms calculated from 3,302 data in the major (110°

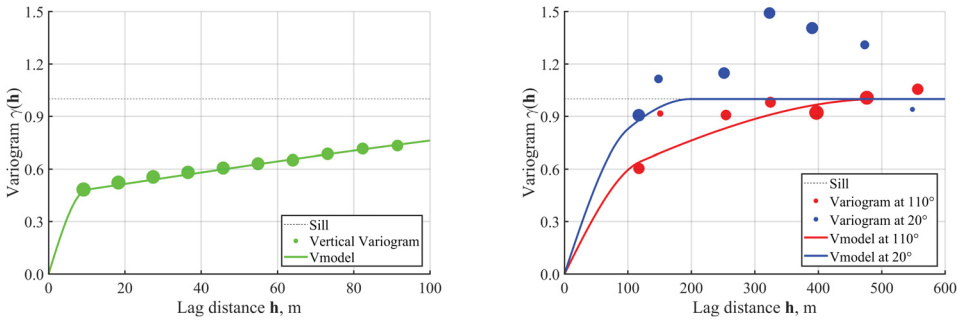


(a) Estimation at an elevation 775.5 meters

(b) Estimation in a 3-D view

Figure 6.4: Visualization of the trend by global kriging estimates from 3,302 data

in horizontal), minor (20° in horizontal) and vertical directions are calculated and plotted with markers. These variograms are fitted with lines and shown in Figure 6.5. No nugget effect is considered in variograms. The trend-like features are most noticeable in the horizontal plane. The experimental variogram in the minor direction (blue markers) climbs steadily above the sill which indicates the existence of a trend.



(a) Vertical variogram

(b) Horizontal variogram

Figure 6.5: Visualization of the directional variograms from 3,302 data. The sizes of the markers represent the relative number of pairs at each lag distance.

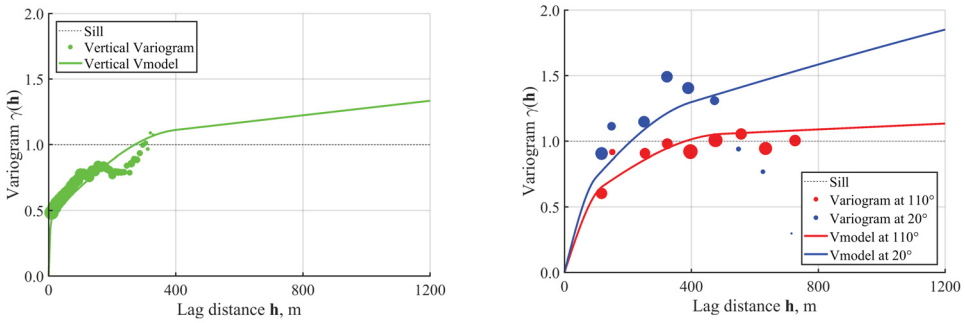
6.2.2 Trend Construction

Trend parameters should be optimized to minimize the subjectivity and improve the final geostatistical model to the greatest extent possible. In trend modeling, several parameters are required including: (1) the background value; (2) the declustering weights; (3) the anisotropy involving directions and ratios; and (4) the smoothing parameter.

The background value controls the tails of the weighting function. It can be used for averaging the values when no data are found in the defined length scales and avoiding some computational problems. The background value should not be too high so that it ignores the spatial correlation between the geological observations and overly averages the trend model. Meanwhile, the background value can not be extremely small which could cause an unstable weighting function. A background value of 0.001 is considered here for modeling the trend.

Declustering is used to remove the bias from the calculated summary statistics and the histogram. Such unbiased statistical distribution from the data should be considered for trend modeling. Figure 6.2b shows the diagnostic plot between the declustered mean and the cell size. The declustered mean reduces with increasing cell size. A 500-meter cell size is applied to the $3,302$ copper data when the declustered mean closes to the minimum value. The declustered mean from $3,302$ data is 0.195% and the standard deviation is 0.225%. These declustering weights are assigned to each sample to construct the trends.

The anisotropy parameters are required for trend modeling. The directions of the trend model should follow the global variogram directions that are 110° in the horizontal direction and along the drillholes in the vertical direction. The standardized experimental variograms are shown with markers in Figure 6.6. The experimental variogram in the horizontal direction shows a geometry anisotropy. The variograms with non-stationary features are modeled within a half of the



(a) Vertical variogram **(b)** Horizontal variogram

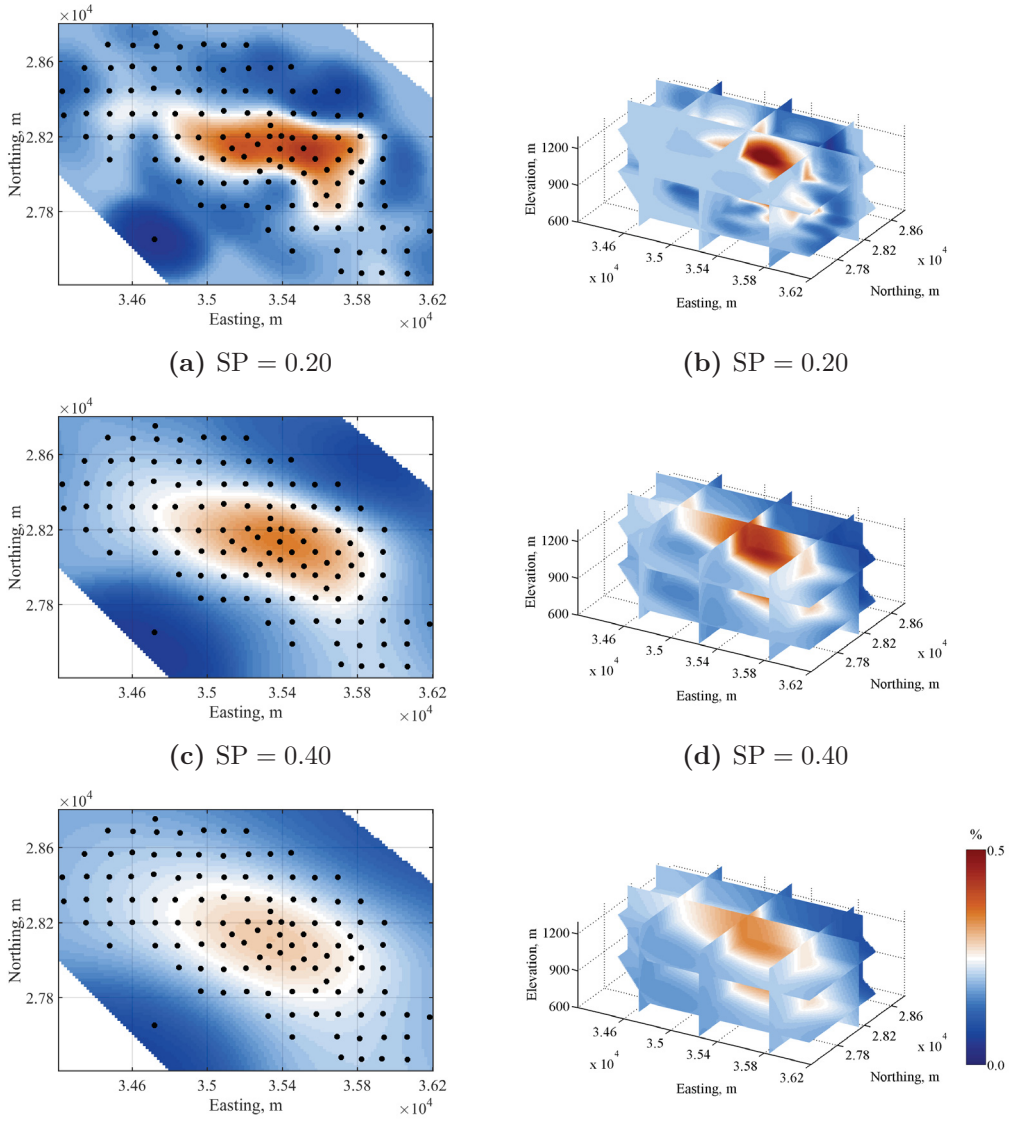
Figure 6.6: Variogram models with non-stationary features from 3,302 data. The sizes of the markers represent the relative number of pairs in each lag distance

domain size, that is, around 1,200 meters. The variogram model with non-stationary features is given as:

$$\gamma(\mathbf{h}) = 0.45\cdot\mathrm{Sph}(\mathbf{h})_{\substack{\alpha=110^\circ \\ a_{\mathrm{hmax}}=120\text{ m} \\ a_{\mathrm{hmin}}=100\text{ m} \\ a_{\mathrm{hvert}}=10\text{ m}}} + 0.55\cdot\mathrm{Sph}(\mathbf{h})_{\substack{\alpha=110^\circ \\ a_{\mathrm{hmax}}=500\text{ m} \\ a_{\mathrm{hmin}}=400\text{ m} \\ a_{\mathrm{hvert}}=400\text{ m}}} + 1.50\cdot\mathrm{Sph}(\mathbf{h})_{\substack{\alpha=110^\circ \\ a_{\mathrm{hmax}}=20{,}000\text{ m} \\ a_{\mathrm{hmin}}=3{,}000\text{ m} \\ a_{\mathrm{hvert}}=8{,}000\text{ m}}} \tag{6.2}$$

A zero nugget effect is considered due to the variability in the short-scale does not impact the results of trend modeling. The spatial correlation is relatively isotropic at short-scale distances and is more anisotropic at large-scale. The area ratio under the variogram curves is considered to calculate the anisotropy ratios in Chapter 3. Anisotropy ratios of 1.42 : 1 and 1.14 : 1 are calculated for the minor and vertical directions.

The smoothing parameter is the last parameter that is tuned for the trend modeling. A variety of trends are modeled with increasing smoothing parameter. Figure 6.7 shows several trend models with the smoothing parameters of 0.20, 0.40 and 0.60, respectively. The black markers represent the 3,302 data. The result confirms the apparent areal trend where high values are concentrated in the center and reduced to the edges.



(a) SP = 0.20 **(b)** SP = 0.20

(c) SP = 0.40 **(d)** SP = 0.40

(e) SP = 0.60 **(f)** SP = 0.60

Figure 6.7: Modeled trends with different smoothing parameters at an elevation 775.5 meters in 2-D/3-D views

Some artifacts exist at the edge due to a limited number of the data in those areas. The trend value approaches the declustered mean in extrapolation regions due to the algorithm. 3-D trend extrapolation is an important topic for future research and the predictions beyond the limits of the data will not be considered here for trend modeling.

6.2.3 Trend Optimization

The objective function of the absolute correlation between the trend and the residuals is shown in Figure 6.8. According to the developed objective function, the optimal trend model is the one when the smoothing parameter doubles the value that minimizes the absolute correlation between the trend and the residuals, that is, $2 \times 0.20 = 0.40$. The optimal trend is shown in Figure 6.7c (2-D top view) and 6.7d (3-D view). Figure 6.9a shows the cross plot between the data and the optimal trend with a correlation of 0.49. Figure 6.9b shows the scatter plots between the optimal trend and the residuals with a correlation of 0.26.

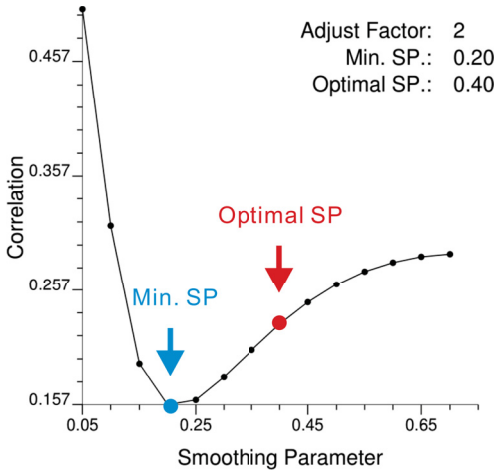
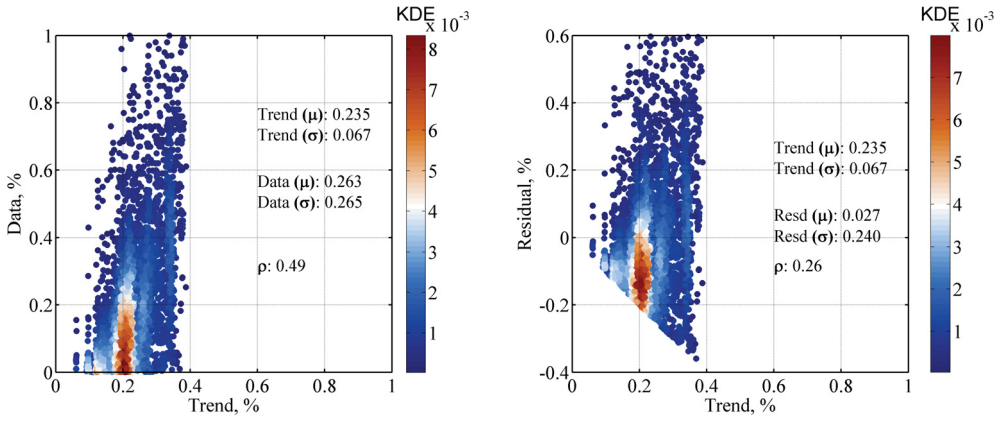


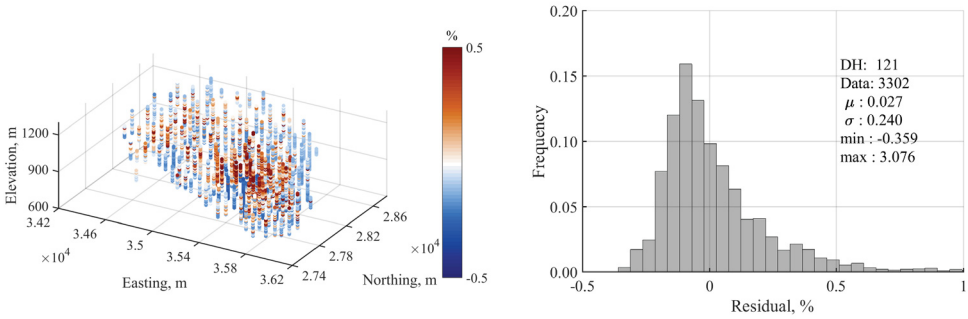
Figure 6.8: Determination of the optimal smoothing parameter

The residual model is checked to ensure the trend-like features are reasonably removed. Figure 6.10 shows the location map and the histogram of



(a) Scatter plot of $m(\boldsymbol{u})$ and $Z(\boldsymbol{u})$ **(b)** Scatter plot of $m(\boldsymbol{u})$ and $R(\boldsymbol{u})$

Figure 6.9: Crossplots of the observed data and residual data with trend values

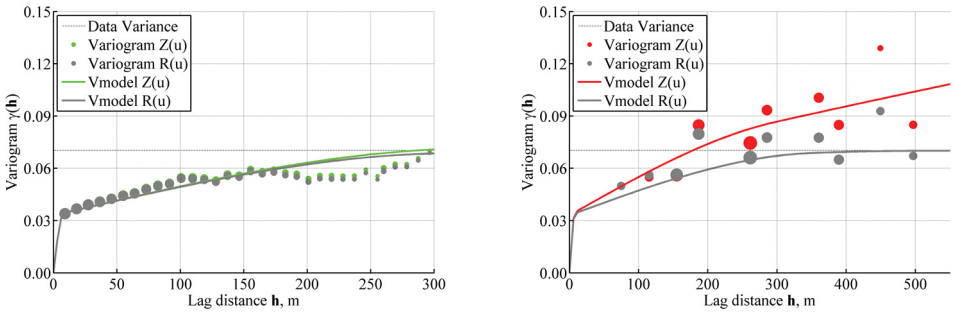


(a) Location map in a 3-D view **(b)** Histogram

Figure 6.10: Visualization of the location map and histogram from 3,302 residuals

the residuals, that is, the data values minus the trend values. The mean of the residuals is 0.027% and the standard deviation is 0.240%. The variogram of the residuals is also modeled. Data are closely sampled along the drillholes and sparsely sampled in the horizontal direction. In this case, the small-scale features from the data are captured from the variogram in the vertical direction and the trend-like features are captured in the horizontal direction. The experimental variograms of the grade values and the residuals are compared and

plotted in Figure 6.11 where the sill is the data variance, 0.070%. The vertical variogram models of the grade values with green markers/line and the residuals with gray markers/line show that no obvious spatial continuity is removed due to the large-scale features are not captured in the vertical direction. The horizontal variogram model of the observation shows that the experimental variogram of the grade values (red markers/line) goes above the data variance (sill), while the horizontal variogram model of the residuals shows that the experimental variogram (gray markers/line) reaches to the data variance and stays constant afterwards which is consistent with a stationary regionalized variable. The trend features are removed from the data and the result could be considered for the subsequent stationary geostatistical modeling.



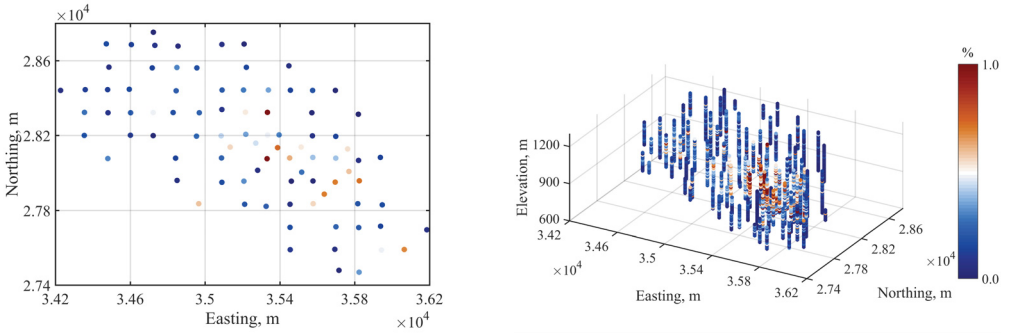
(a) Vertical variogram **(b)** Horizontal variogram

Figure 6.11: Variogram models of the data and residuals from 3,302 data

6.3 Data Preparation

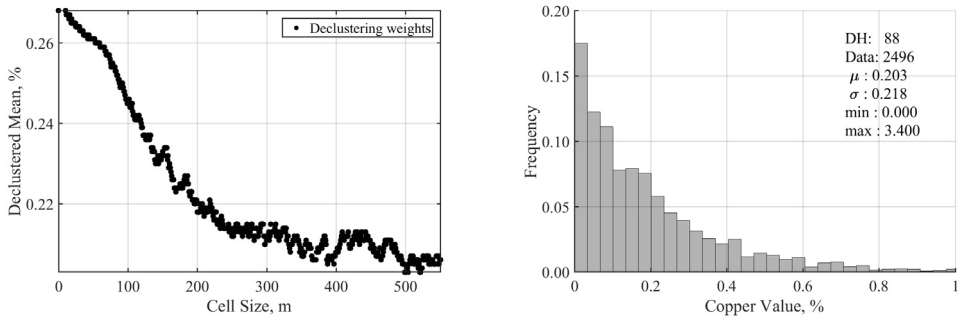
Figure 6.1 shows the despiked 3,302 copper grade values. In order to check the importance of the trend modeling in the geostatistical prediction, the total data are randomly divided into a modeling set and a test set. In this case, 80% of the total data, that is 2,496 data from 88 drillhole, will be used for the geostatistical modeling. The remaining will be considered for checking the simulated results.

Figures 6.12 show location maps of the 2,496 modeling grades from 88 drillholes in the modeling data. The cell declustering is performed with a 500-meter cell-size in Figure 6.13a. The corrected histogram with a mean of 0.203% and a standard deviation of 0.218% is shown in Figure 6.13b.



(a) Location map in a 2-D top view **(b)** Location map in a 3-D view

Figure 6.12: Location maps from 2,496 modeling data



(a) Cell-size declustering **(b)** Histogram with despiking

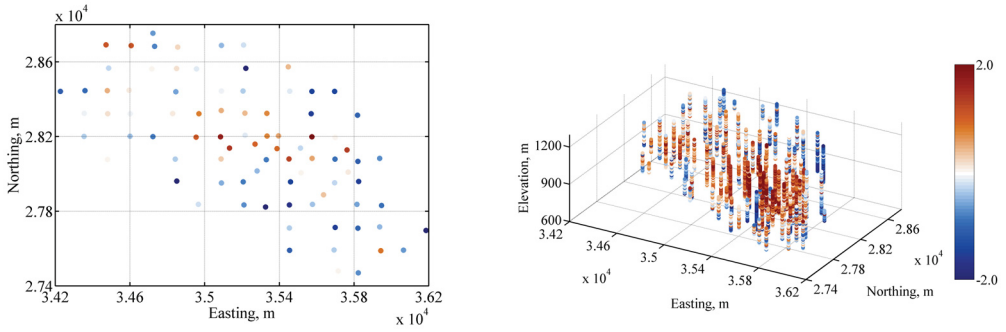
Figure 6.13: Cell declustering and corrected histogram from 2,496 modeling data

2,496 modeling data are pre-processed with the normal score transformation and variogram model fitting. The normal score transform considers the declustering weights. The transformed 2,496 modeling data with declustering weights show a standard normal distribution, that is, the mean is 0.0% and the

standardized deviation is 1.0%. The location maps after the normal score transformation are shown in Figure 6.14. The directional experimental variograms in normal score units are modeled with an isotropic variogram given as:

$$\gamma(\mathbf{h}) = 0.4 \cdot \mathrm{Sph}(\mathbf{h})\underset{\substack{a_{\mathrm{hmax}}=20\text{ m}\\ a_{\mathrm{hmin}}=20\text{ m}\\ a_{\mathrm{hvert}}=20\text{ m}}}{\overset{\alpha=0^\circ}{}} + 0.6 \cdot \mathrm{Sph}(\mathbf{h})\underset{\substack{a_{\mathrm{hmax}}=550\text{ m}\\ a_{\mathrm{hmin}}=550\text{ m}\\ a_{\mathrm{hvert}}=550\text{ m}}}{\overset{\alpha=0^\circ}{}} \tag{6.3}$$

No nugget effect is considered. The variogram models are shown in Figure 6.15.



(a) Location map in a 2-D top view **(b)** Location map in a 3-D view

Figure 6.14: Location maps from 2,496 modeling data in NS units

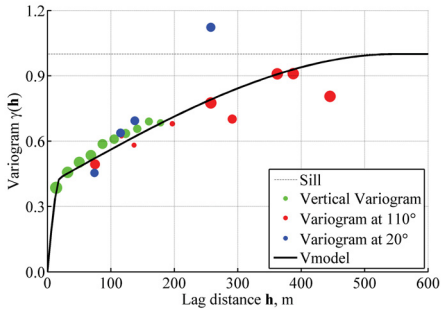
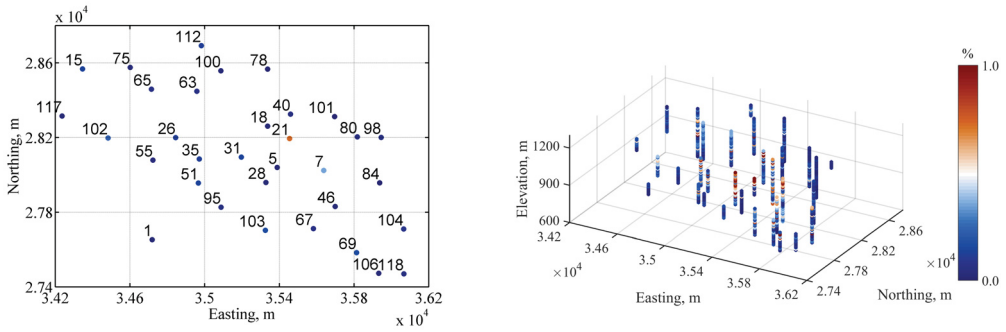


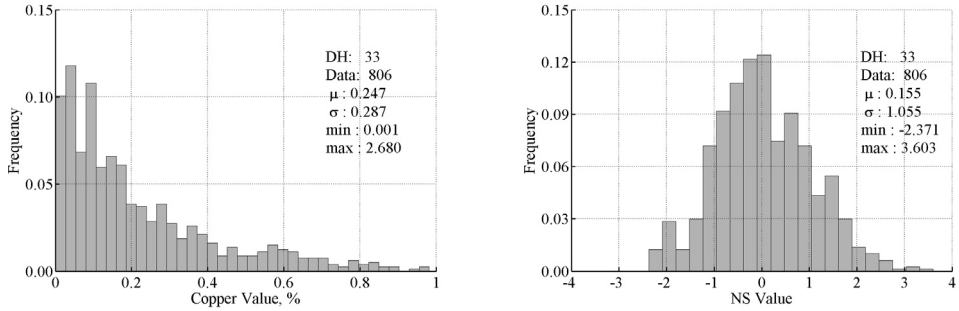
Figure 6.15: Isotropic variogram model from 2,496 modeling data in NS units. The sizes of the markers represent the relative number of pairs in each lag distance

The location of the remaining 806 test data are labeled with the drillhole IDs and shown in Figure 6.16. 806 values with a mean of 0.247% and a standard deviation of 0.287% in Figure 6.17a are transformed into normal score units with a reference distribution from 2,496 modeling data. Figure 6.17b shows the transformed test data with a mean of 0.155 and a standard deviation of 1.055.



(a) Location map in a 2-D top view **(b)** Location map in a 3-D view

Figure 6.16: Location maps from 806 checking data



(a) Histogram in original units **(b)** Histogram in NS units

Figure 6.17: Histograms from 806 test data

6.4 Conventional Stationary Geostatistical Modeling

Conventional geostatistical modeling assumes stationarity. Although the given data show a trend, the assumption of stationarity is made. Major points of the

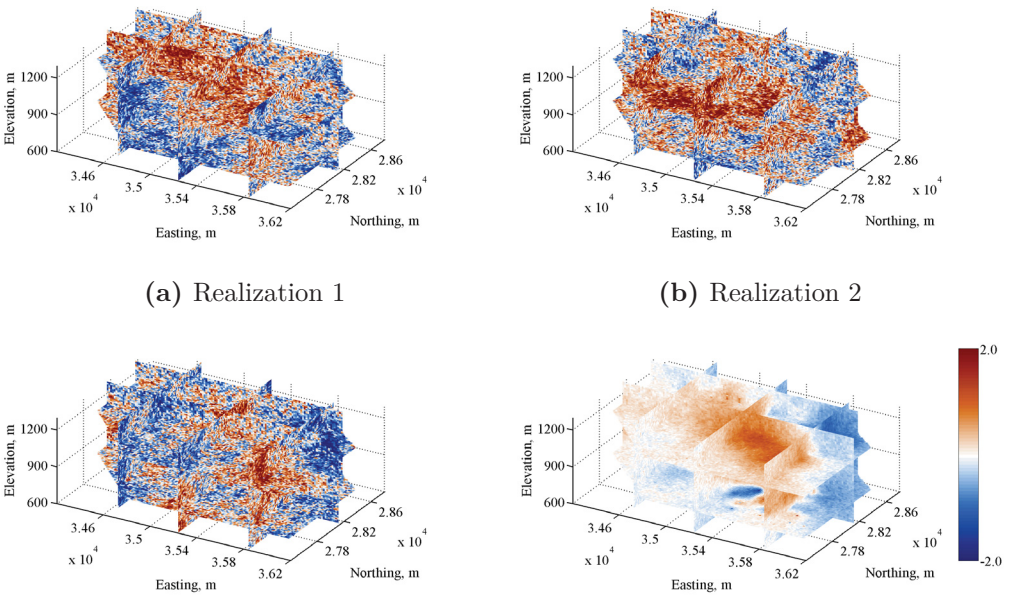
conventional simulation can be summarized as: (1) conduct sequential Gaussian simulation of the normal score transformed data; (2) generate many realizations; (3) transfer all realizations back to original units; and also (4) check the histogram and variogram reproduction of models over all realizations. The procedure is demonstrated step by step.

6.4.1 Sequential Gaussian Simulation

Sequential Gaussian simulation with simple kriging is well established, easy to implement and is considered with the normal score transformed data. Several parameters must be chosen for the simulation. The search data for the simulation must be large enough so that enough neighbors are considered. A search ellipse is defined by the variogram ranges in Equation 6.3. Many realizations are generated. The first three realizations and the average values over the first 100 realizations in normal score units are shown in Figure 6.18.

A normal score back transformation is considered to bring all realizations back to original units. Figure 6.19 shows the back transformed results with the first three realizations and the averages of all realizations in original units. The variance of 100 values at each location is shown in Figure 6.20. There is a low variance in the low-valued zones and a high variance in high-valued zones as expected with a positively skewed distribution. The variance is high in the margins because of few conditioning data.

Histogram reproduction can be checked to ensure that the geostatistical realizations reproduce the input histogram. The histograms of the first 100 realizations are shown with black lines, while the $2,496$ conditioning data are shown with a red line in Figure 6.21a. The mean of the realizations, 0.198%, is slightly less than the reference mean, 0.203%; the standard deviation, 0.199%, is lower than the conditioning variance, 0.218%. Although the statistics from the realizations are slightly lower than the $2,496$ conditioning data, the realizations



(a) Realization 1

(b) Realization 2

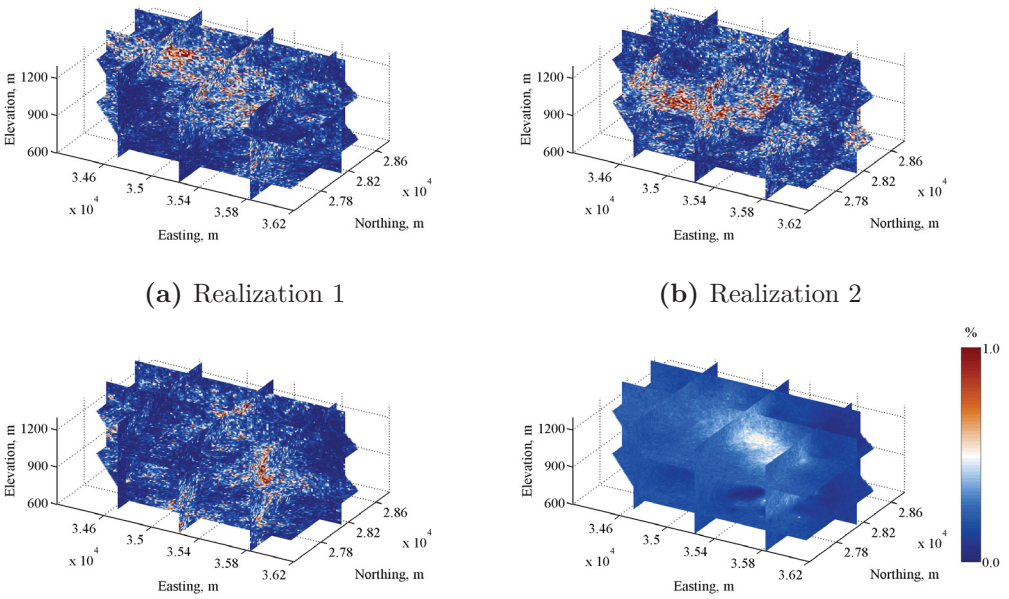
(c) Realization 3

(d) Average values over 100 realizations

Figure 6.18: First three realizations and averages over 100 realizations in NS units

successfully reproduce the global distribution of the 2,496 conditioning data. The histogram reproduction appears reasonable.

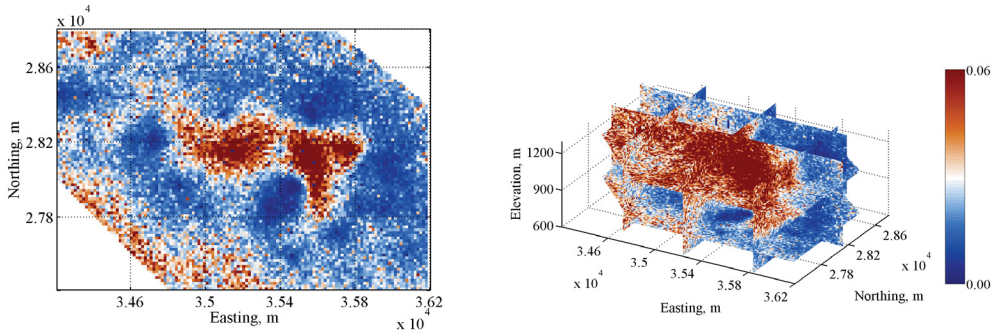
The variogram measures the spatial variability of the regionalized variable. Figure 6.21b shows the variogram reproduction in original units. Directional experimental variograms are plotted with markers. The black line is the expected isotropic variogram from the original 2,496 values. Light gray lines represent the variograms of the first 100 realization, while the dark gray dash-dot line represents the average of all realizations. The predicted model captures more variability from the grade values in the short-scale structure and contains more continuous in the large-scale structure. Overall, the variograms of all realizations compare closely with the original isotropic variogram.



(a) Realization 1 **(b)** Realization 2

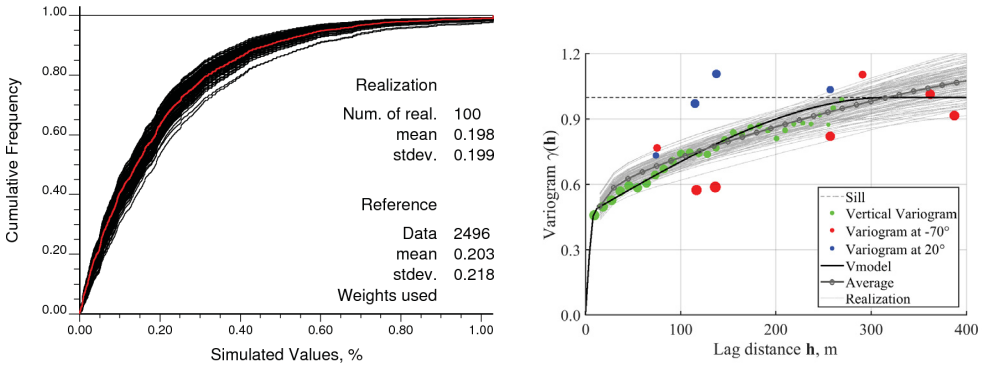
(c) Realization 3 **(d)** Average values over 100 realizations

Figure 6.19: First three realizations and the average over 100 realizations in original units



(a) A slice at an elevation 775.5 meters **(b)** A volume in a 3-D view

Figure 6.20: Local variance over 100 realizations in original units



(a) Histogram reproduction **(b)** Variogram reproduction

Figure 6.21: Histogram and variogram reproduction with the conventional geostatistical modeling in original units

6.4.2 Cross Validation

The first validation is to compare the true values with the distributions of the uncertainty at the 806 test locations. The distributions of the local uncertainty are specified by the normal equations in normal score units. The plot in Figure 6.22 shows the accuracy of the simulated distributions of the uncertainty in normal score units. The grid of light lines shows the probability intervals, while the red lines and black dots show the deviations of the actual proportions from the predicted probability intervals (Deutsch, 2010b). The prediction is neither accurate nor precise due to the observed fraction in each probability interval is less than the predicted values (45° black line). The model appears to slightly underestimate the uncertainty. The local uncertainty is 0.589 with the conventional method.

The second validation is to compare the 806 real values with the simulated average values in original units. Figure 6.23 shows the comparison. The mean of the average checks the bias of the predicted model and the standard deviation of the average measures the smoothing effect of the predicted model. The mean of the average values, 0.231%, is slightly lower than the true mean, 0.247%.

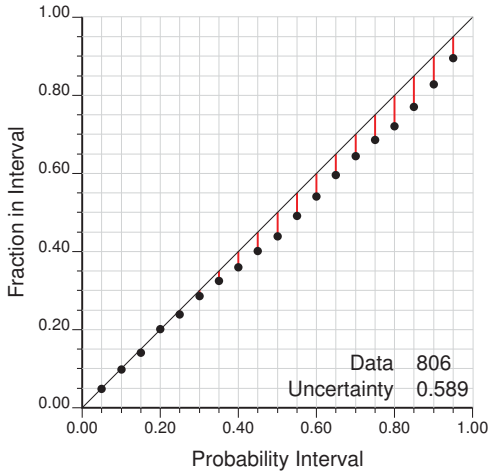
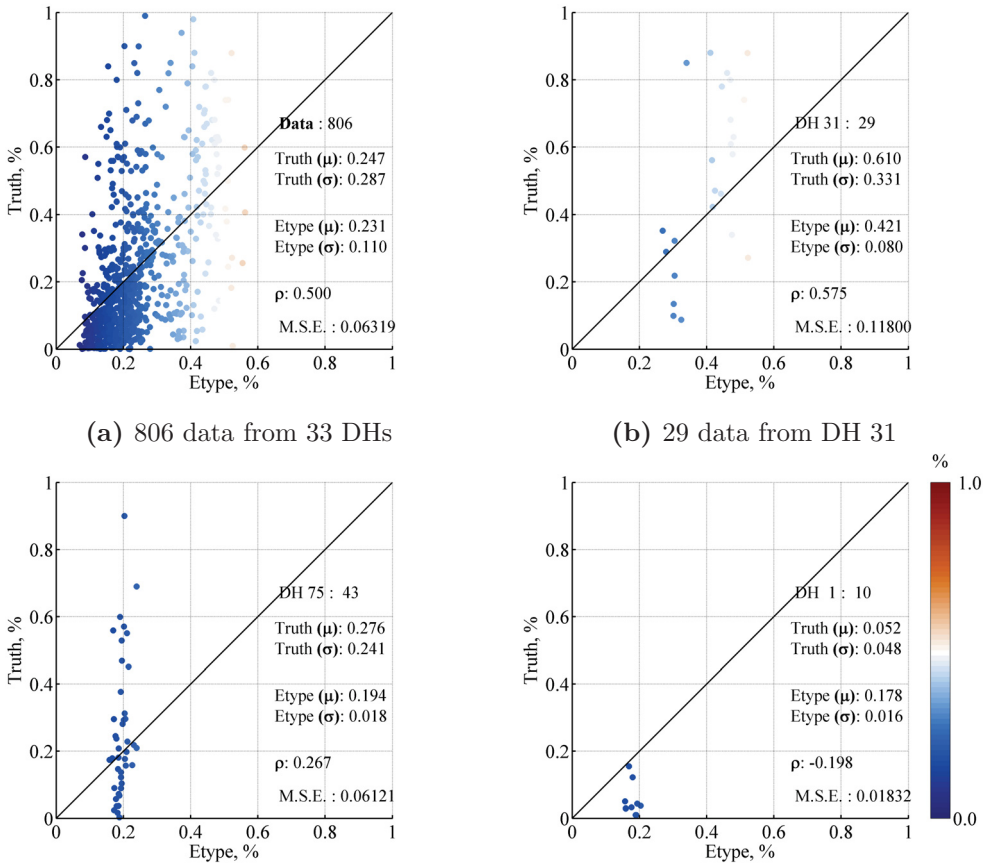


Figure 6.22: Cross validation from 806 test data with the conventional geostatistical modeling in NS units

The standard deviation of the average values is 0.108% that is much lower than the truth, 0.287%. This low standard deviation is due to the average of many realizations being overly smooth. The mean squared error value summarizes the prediction performance and it should be small. The mean squared error value between the true values and the average values is 0.06304. The statistics from 3 drillholes that are extracted from high-, medium- and low-valued zones are calculated showing the mean squared error values of 0.11989, 0.06173 and 0.01358, respectively.

6.5 Conventional Non-Stationary Geostatistical Modeling

The assumption of stationarity on the data is relaxed and the stationary assumption is applied on the residuals that removes from the optimal trend. The approach of modeling with residuals and adding the trend model back in the final model is considered. The major steps of this conventional non-stationary



(a) 806 data from 33 DHs

(b) 29 data from DH 31

(c) 43 data from DH 75

(d) 10 data from DH 1

Figure 6.23: Mean squared error values from 806 test data with the conventional geostatistical modeling in original units

geostatistical method can be summarized as: (1) obtain residuals that removes the optimal trend from data; (2) transform the residuals into normal score units; (3) conduct sequential Gaussian simulation on the normal score transformed residuals; (4) generate many realizations; (5) back transform all realizations into original units; (6) add the trend model back to the simulated residuals; and (7) check the histogram and variogram reproductions of the simulated residuals and the final simulated values over all realizations. Details will be demonstrated with the same data.

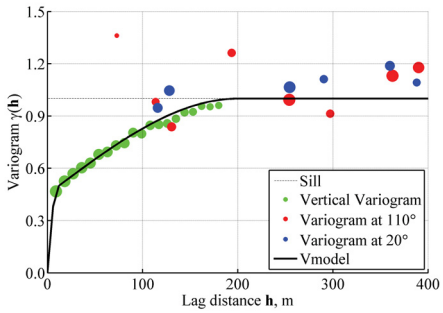


Figure 6.24: variogram models from 2,496 residuals in NS units

6.5.1 Sequential Gaussian Simulation with Residuals

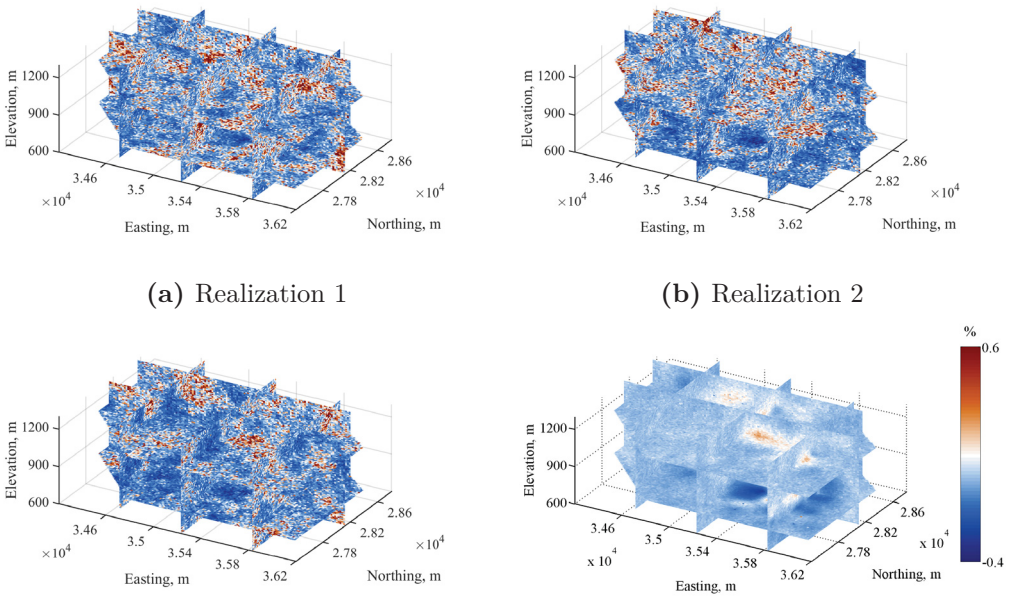
A scatter plot of the residuals against the trend model is shown in Figure 6.9b with a correlation of 0.26. The location map and the histogram from the residuals are shown in Figure 6.10. Still, the 500−meter cell size is considered for the cell declustering. Residuals are transformed into normal score units with the declustering weights. The isotropic variogram model in normal score units is given as:

$$\gamma(\mathbf{h}) = 0.45 \cdot \mathrm{Sph}(\mathbf{h})\underset{\substack{a_{\mathrm{hmax}}=10\text{ m}\\ a_{\mathrm{hmin}}=10\text{ m}\\ a_{\mathrm{hvert}}=10\text{ m}}}{\overset{\alpha=0^\circ}{}} + 0.55 \cdot \mathrm{Sph}(\mathbf{h})\underset{\substack{a_{\mathrm{hmax}}=200\text{ m}\\ a_{\mathrm{hmin}}=200\text{ m}\\ a_{\mathrm{hvert}}=200\text{ m}}}{\overset{\alpha=0^\circ}{}} \tag{6.4}$$

The variogram model is shown in Figure 6.24. The markers with the red, blue and green colors represent the experimental variograms in the major, minor and vertical directions, respectively, and the black line indicates the isotropic variogram model.

Sequential Gaussian simulation is conducted on the transformed residuals and many realizations are generated. A normal score back transformation is considered to bring all realizations of residuals back to the original units. Figure 6.25 shows the back transformed residuals with the first three realizations

and the averages of the first 100 in original units. The simulated residuals show more randomness.



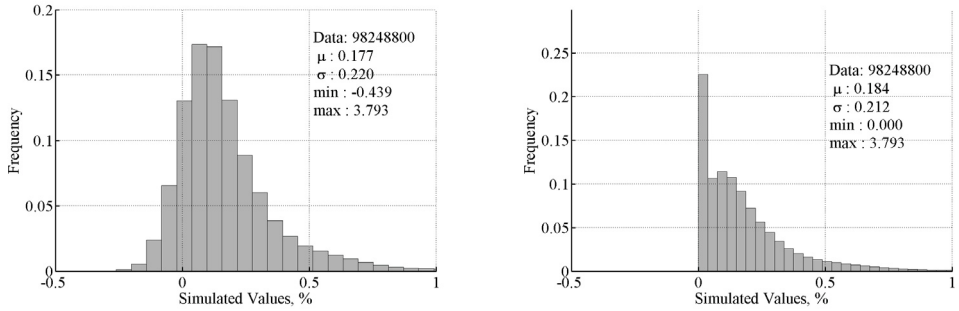
(a) Realization 1

(b) Realization 2

(c) Realization 3

(d) Average values over 100 realizations

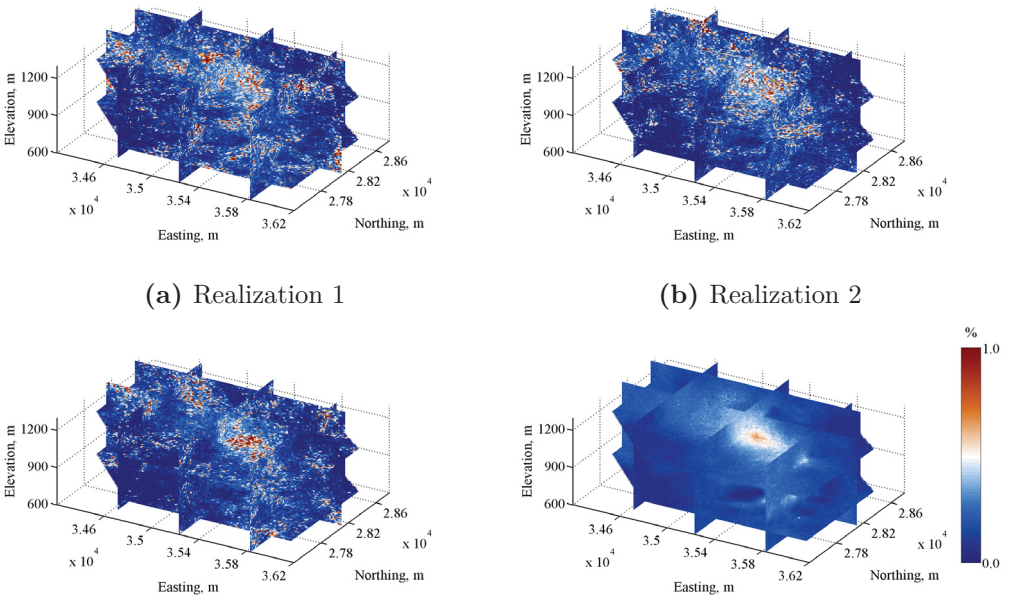
Figure 6.25: First three realizations and the average over 100 realizations of residuals in original units



(a) Raw simulate values

(b) Corrected simulated values

Figure 6.26: Histograms correction of the final simulated values



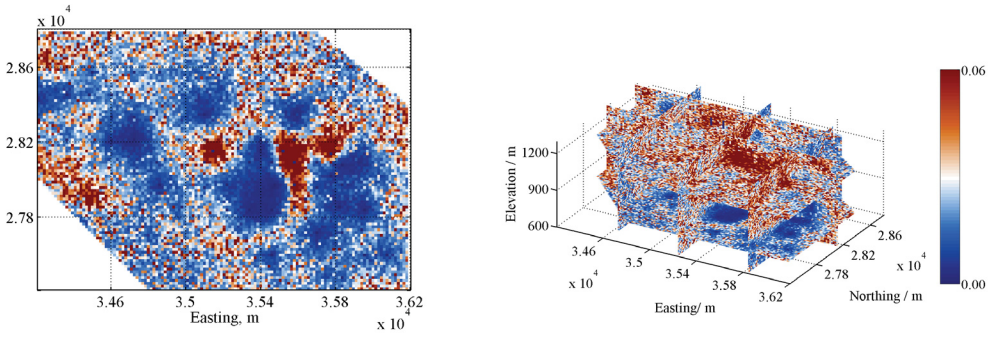
(a) Realization 1 **(b)** Realization 2

(c) Realization 3 **(d)** Average values over 100 realizations

Figure 6.27: First three realizations and the average of the final model over 100 realizations in original units

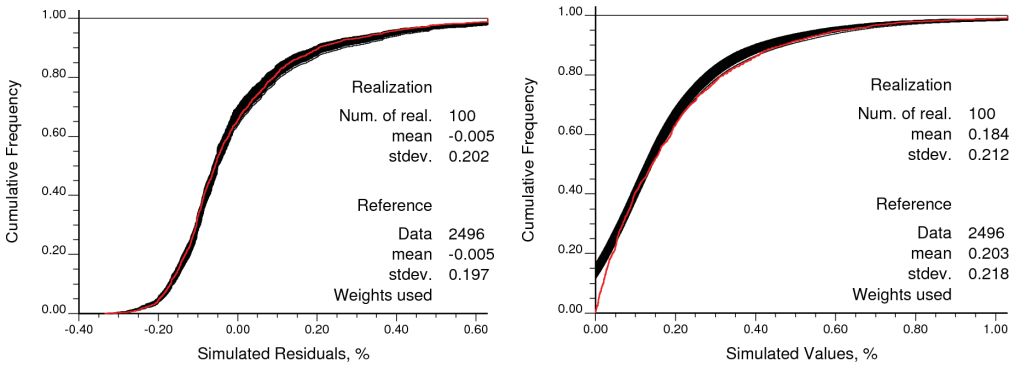
The trend in Figure 6.7d is added back to the simulated residuals. A constraint that the final model should not contain negative grade values in original units $z(\boldsymbol{u}) \geq 0.0$ is considered so that the existed negative values in the final simulated model should be reset to zeros. Figure 6.26a shows the naive histogram of the first 100 realizations after adding the trend back and Figure 6.26b shows the constraint corrected histogram from realizations of the final model so that no negative values exist. Figure 6.27 shows the first three realizations and the averages of the final model over all realizations in original units. Figure 6.28 shows the variance over all realizations. This model has more variability than the preceding model of sequential Gaussian simulation directly.

Histogram reproductions of the residuals and the final grades over 100 realizations are checked. The red line represents the cumulative distribution function



(a) A slice at an elevation 775.5 meters **(b)** A volume in a 3-D view

Figure 6.28: Local variance of the final model over 100 realizations in original units



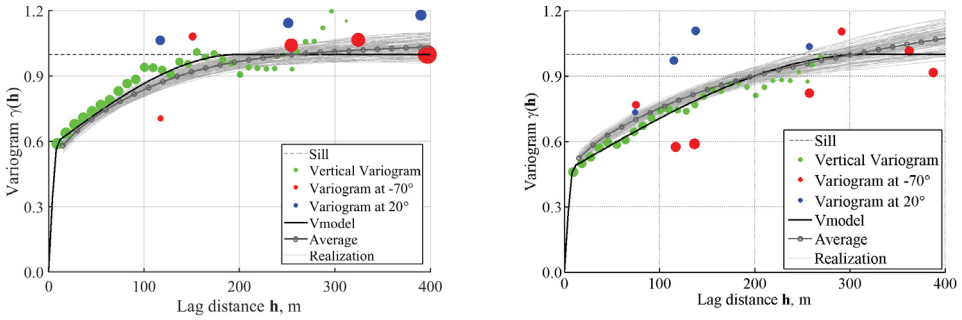
(a) Histogram reproduction of residuals **(b)** Histogram reproduction of the final simulated values

Figure 6.29: Histogram reproduction with the conventional non-stationary geostatistical modeling in original units

from 2,496 modeling residuals/data and the black lines are the histograms from all realizations. The realizations of the residuals in Figure 6.29a successfully reproduce the global mean, -0.005%, the global standard deviation, 0.197%, and the global distribution of the residuals in original units. Figure 6.29b shows the histogram reproduction of the final grades that shows an obvious bias. The mean of the realizations of the final model without negative values, 0.184%, is lower than the reference mean, 0.203%. The standard deviation of the final model, 0.212%, successful reproduces the conditioning standard deviation of the

regionalized data, 0.218%. However, the distributions of all realizations fail to reproduce the global distribution of the regionalized data. The histogram reproduction of the final simulated values performs poor.

Figure 6.30 shows the variogram reproduction of the residuals and the final model over 100 realizations. The red, blue and green markers represent the experimental variograms in the horizontal major, horizontal minor and the vertical directions. Light gray lines are the variograms of each realization and the dark gray line represents the average of all realizations. The variogram model with the residuals in Figure 6.30a performs reasonable reproduction that reproduces the variogram model from $2,496$ modeling residuals (black line). The variogram model after adding the trend back contains similar variability shown by a black line in Figure 6.30b. Overall, the variogram reproduction with the final simulated values is well performed.



(a) Variogram reproduction of residuals

(b) Variogram reproduction of final simulated values

Figure 6.30: Variogram reproductions with the conventional non-stationary geostatistical modeling in original units

6.5.2 Cross Validation

The first validation is to compare the true residuals with the simulated average residuals at 806 test locations in normal score units. The residuals are

obtained from the data that removes the optimal trend. These residuals are assumed known without uncertainty. 806 true residuals are transformed into normal score units with the reference distribution of $2,496$ modeling residuals. Figure 6.31 shows the uncertainty of the 806 true residuals and the simulated average residuals in normal score units. The dots in each intervals fall close to the 45° line so that the predicted residuals are accurate and precise. The local uncertainty with the trend dichotomy technique is 0.821 which shows 40% more variability than that from the conventional method.

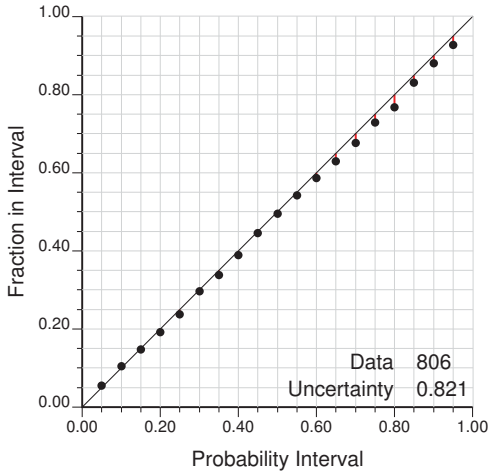
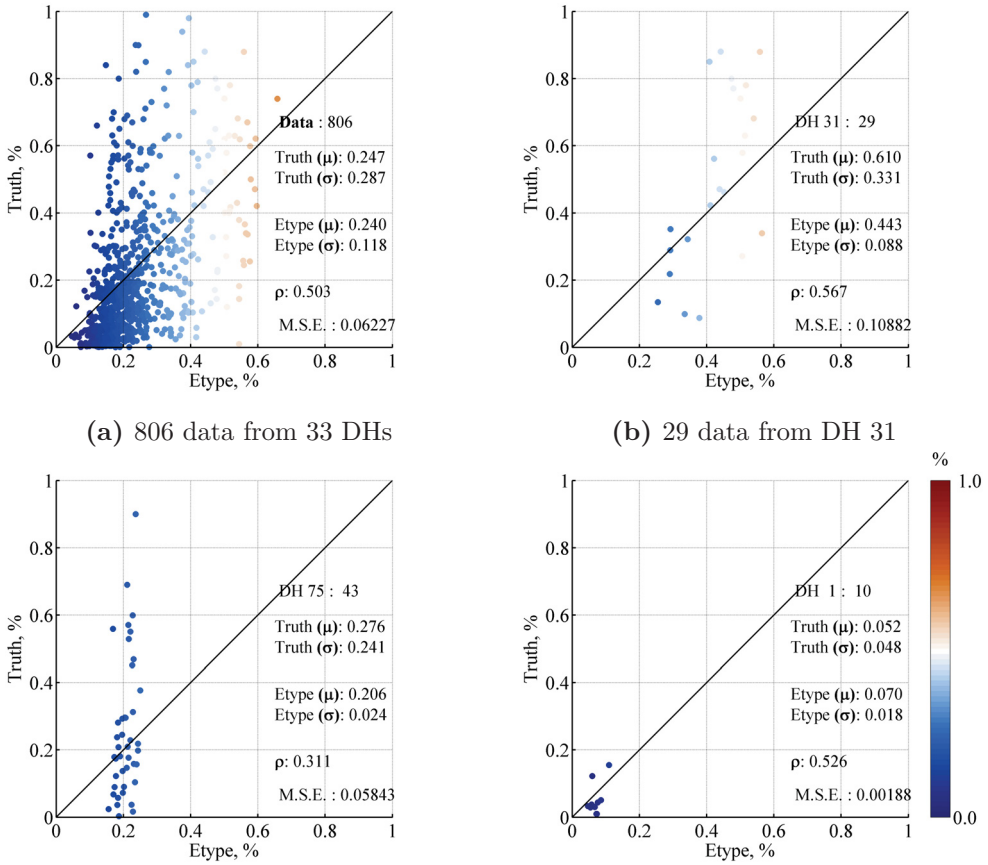


Figure 6.31: Cross validation from 806 test residuals with the conventional non-stationary geostatistical modeling in NS units

The second validation is to compare 806 real values with the simulated average values of the regionalized data in original units. Figure 6.32 shows the comparisons. The mean of the average values, 0.238%, is close to the true mean, 0.247%. The standard deviation of the average values, 0.118%, is lower than the true standard deviation, 0.287%, but is higher than that of the conventional method, 0.108%. The mean squared error value measures the difference between the truth and what is being estimated, and further, summarizes the prediction performance. The minimized mean squared error can be used to identify the

best method. The mean squared error values between the true values and the average values are 0.06229 with a mild improvement of 1.19% with the preceding model of sequential Gaussian simulation directly. 3 drillholes show the improvements of 8.43%, 5.64% and 85.80%, respectively, compared with the conventional stationary geostatistical modeling technique.



(a) 806 data from 33 DHs

(b) 29 data from DH 31

(c) 43 data from DH 75

(d) 10 data from DH 1

Figure 6.32: Mean squared error values from 806 test data with the conventional non-stationary geostatistical modeling in original units

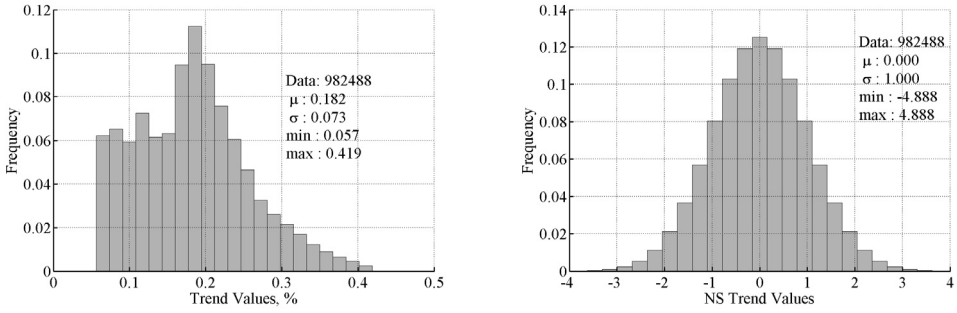
6.6 Developed Geostatistical Modeling

The stepwise conditional transformation with Gaussian mixture models is now considered. In the application of the transformation, 2,496 modeling data are transformed conditional to the trend component. The complex features in the data are removed so that the transformed variable stays stationary. All complex features should be reproduced in the back transformation. The major points of the developed method can be summarized as: (1) normal score transform the optimal trend model and the modeling data; (2) fit the Gaussian mixture model between the normal score transformed trend model and data; (3) transform the data in normal score units according to the Gaussian mixture model; (4) conduct sequential Gaussian simulation on the transformed data; (5) generate many realizations; (6) transfer all realizations back to normal score units, and then original units; and (7) check the histogram and variogram reproductions of the models over all realizations. Details with the developed simulation workflow will be described with the same example.

6.6.1 Gaussian Mixture Model

The trend model contains the large-scale variability and is shown in Figure 6.7d. Obvious trend extrapolations exist in the Southwest and Northeast areas where the trend values approach the declustered global mean of the regionalized variable. Figure 6.33a shows the histogram of the exhaustive trend mode with a mean of 0.182% and a standard deviation of 0.073%. There are high frequencies near the declustered global mean of the regionalized variable because of the trend model calculations in trend extrapolation regions. The scatter plot between the trend and the data in original units is shown in Figure 6.9a indicating a correlation of 0.49. The exhaustive trend model is transformed into normal score units, see Figure 6.33b, while 2,496 modeling data are transformed into normal score

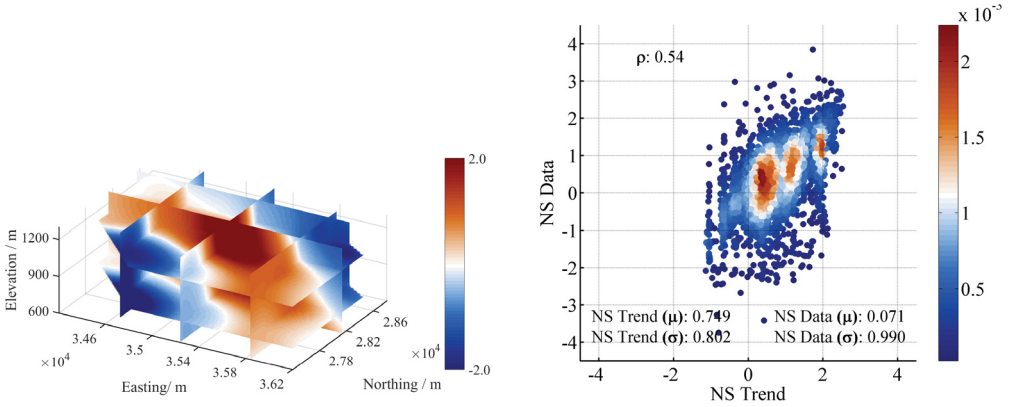
units with the declustering weights independently. Figure 6.34 shows the transformed trend model and the transformed scatter plot on a 2D density estimation showing a direct relationship with a correlation of 0.54.



(a) Exhaustive trend model in original units

(b) Exhaustive trend model in NS units

Figure 6.33: Histograms of the trend models in original and NS units



(a) Trend model

(b) Scatter plot

Figure 6.34: The transformed trend model and crossplot from 2,496 data in NS units

The scatter plot in Figure 6.34 shows obvious artifacts. The gridding from the data sampling causes these artifacts. Figure 6.33b shows the histogram of the fully transformed trend model under the standard normal distribution and Figure 6.35 shows the histogram of the trend model only at data locations in

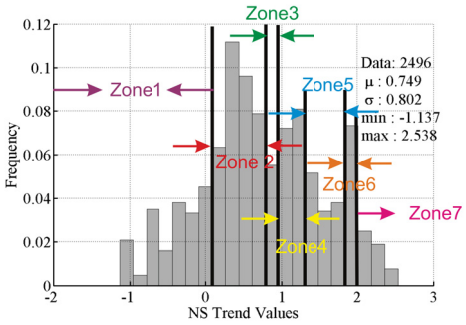
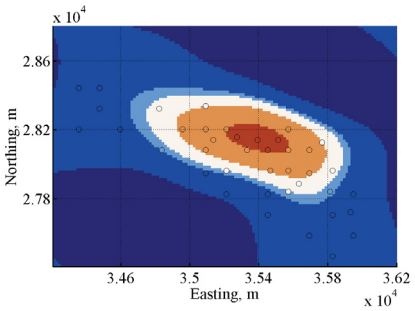


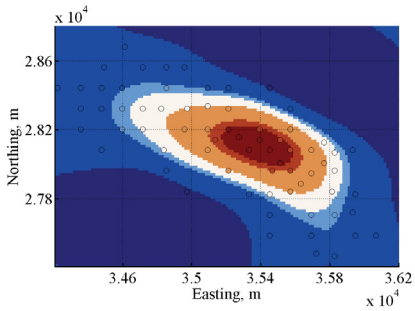
Figure 6.35: Histogram from the trend model at 2,496 modeling data locations in NS units

normal score units with a mean of 0.749% and a standard deviation of 0.802%. The trend values are divided into zones. These categories are defined by the frequencies from the probability distribution function in Figure 6.35. Zones 1, 3, 5 and 7 have relatively low probability and Zones 2, 4 and 6 have relatively high probability. Zones 3 and 6 are the stripe-like artifacts are located in the scatter plot. Figure 6.36 shows the contour maps of the normal score transformed trend values in different elevation levels and the dots represent the 2,496 modeling data within each category. Not many samples could be found in Zone 3 and 5 so that causes less points in that area of the scatter plot. Other examples with more random sampling have been tested and there are no such obvious strips.

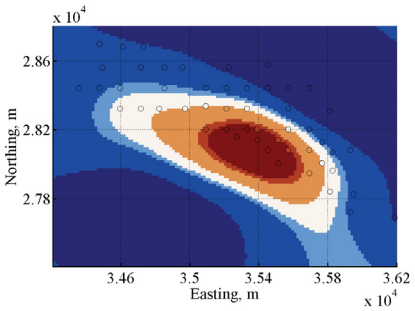
The stepwise conditional transformation with a Gaussian mixture model is considered to remove the complexity from the data. The decision of the number of mixture components is subjective. Too many components of the Gaussian mixture model could over-fit the relationship between the trend and the data, especially when the density issues in the scatter plot exist. In this case study, two components are determined by visual inspection to fit the scatter plot. Figure 6.37 shows the Gaussian mixture model. The univariate distributions of the trend model and the data are shown in Figure 6.37a and Figure 6.37d, respectively. The marginal distributions from the Gaussian mixture models are not



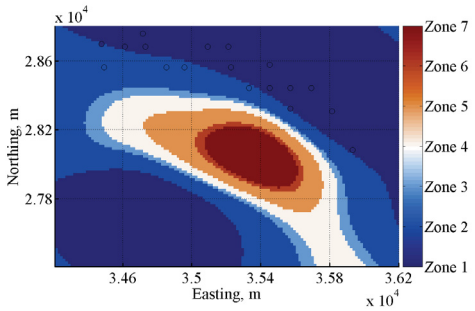
(a) Contour map at an elevation 730.5 meters



(b) Contour map at an elevation 865.5 meters



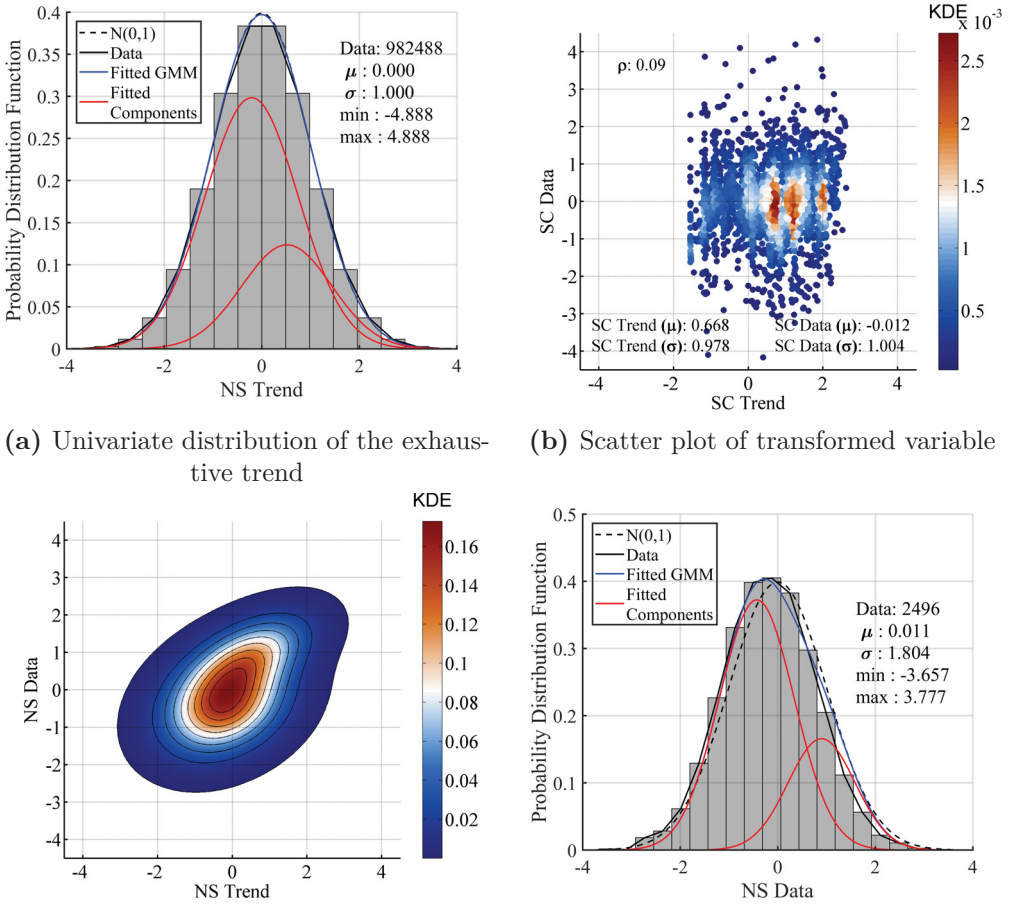
(c) Contour map at an elevation 1000.5 meters



(d) Contour map at an elevation 1135.5 meters

Figure 6.36: Contour maps of the trend model with 2,496 Modeling data in NS units

exactly normal; however, the deviation appears to be very small in Figure 6.37a where the combined mixture distribution and an exact normal distribution are almost perfectly overlapping. The bivariate distribution is shown on a 2D probability density plot in Figure 6.37c. The transformed variables are uncorrelated, 0.18, in Figure 6.37b. Note that the striping artifacts in the scatter plot come from the normal score transformation rather than the conditional transformation.



(a) Univariate distribution of the exhaustive trend

(b) Scatter plot of transformed variable

(c) Bivariate distribution

(d) Univariate distribution of data

Figure 6.37: Bivariate and univariate distributions of the Gaussian mixture model

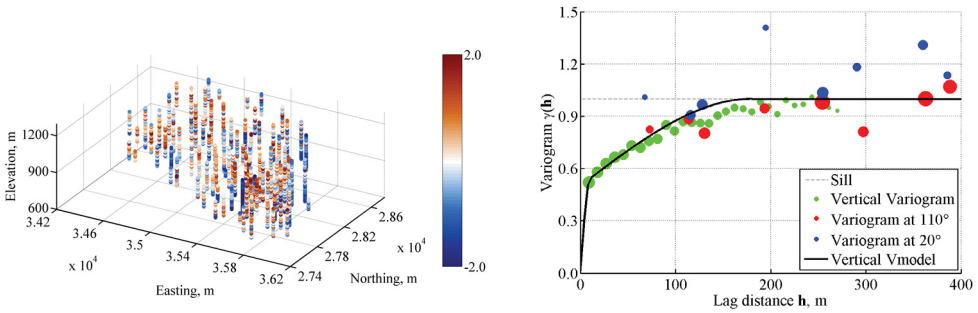
6.6.2 Stepwise Conditional Transformation

The data after the stepwise conditional transformation in Figure 6.38a show a randomness and the large-scale trend-like features are removed. The directional variograms in stepwise units are fitted with an isotropic variogram model. The

isotropic variogram model is shown in Figure 6.38b and the equation is given as:

$$\gamma(\mathbf{h}) = 0.50 \cdot \mathrm{Sph}(\mathbf{h})_{\substack{\alpha=0^\circ \\ a_{\mathrm{hmax}}=10\ \mathrm{m} \\ a_{\mathrm{hmin}}=10\ \mathrm{m} \\ a_{\mathrm{hvert}}=10\ \mathrm{m}}} + 0.50 \cdot \mathrm{Sph}(\mathbf{h})_{\substack{\alpha=0^\circ \\ a_{\mathrm{hmax}}=180\ \mathrm{m} \\ a_{\mathrm{hmin}}=180\ \mathrm{m} \\ a_{\mathrm{hvert}}=180\ \mathrm{m}}} \tag{6.5}$$

The sizes of the markers represent the relative number of pairs in each direction.

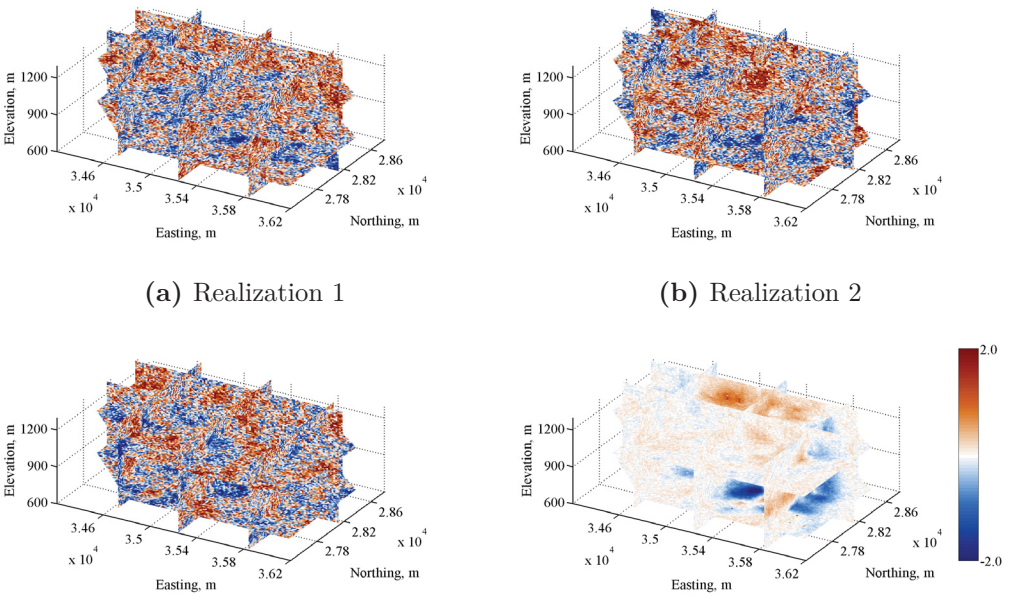


(a) Location map **(b)** Variogram model

Figure 6.38: Location map and variogram model from 2,496 modeling data in SC units

6.6.3 Sequential Gaussian simulation

Sequential Gaussian simulation is conducted on the transformed variable. The search radii and ellipsoid are defined based on variogram ranges in Equation 6.5. Many realizations are generated. Figure 6.39 shows the first three realizations and the average over 100 realizations in stepwise units. No trend-like features exist in the simulated results. A stepwise conditional back transformation with the Gaussian mixture model is performed. Figure 6.40 shows the first three realizations and the average of 100 realizations in normal score units. The simulated results show that the trend-like features are restored in the back transformation. The initial normal score transformation should be also reversed. Figure 6.41 shows the first three realizations and the average of 100 realizations in original



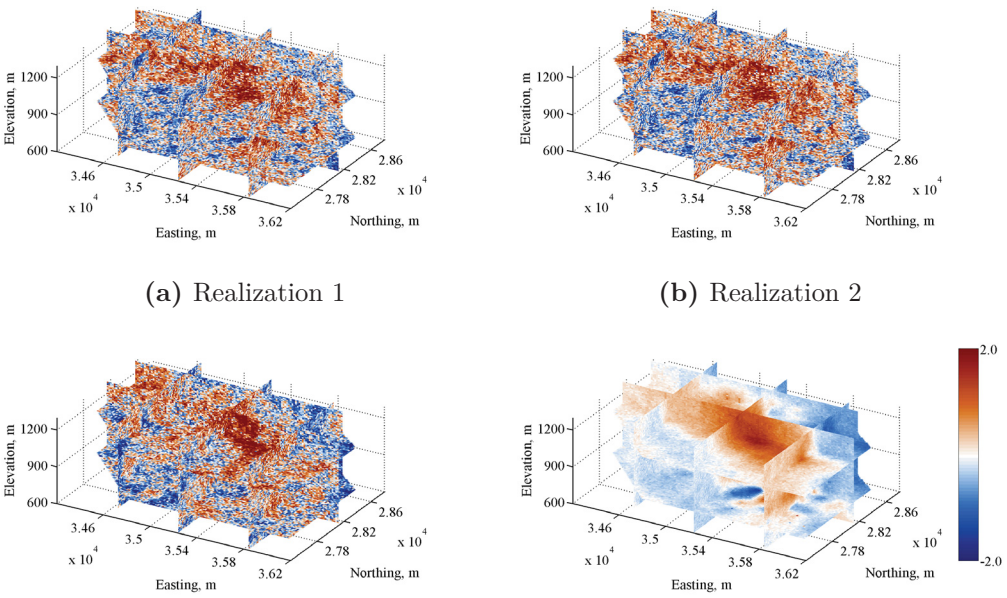
(a) Realization 1 **(b)** Realization 2

(c) Realization 3 **(d)** Average values over 100 realizations

Figure 6.39: First three realizations and the averages over 100 realizations in SC units

units. The local variance is calculated and shown in Figure 6.42. The map shows the high variance in the central region due to the low variance at the margins.

The histogram of realizations should be consistent with the histogram from 2,496 conditioning data. The realizations over all locations are considered. The histogram is reasonably reproduced in original units, see in Figure 6.43a. The mean over 100 realizations is 0.193% that is slightly lower than the conditioning mean, 0.203%. The standard deviation is 0.202% that is reasonably close to the standard deviation from the conditioning data, 0.218%. The histograms of all realizations are reproduced well. Figure 6.43b shows the variogram reproduction. The overall variogram reproduction from realizations is better than that from the conventional (non-)stationary methods in Figure 6.21b and Figure 6.30.



(a) Realization 1

(b) Realization 2

(c) Realization 3

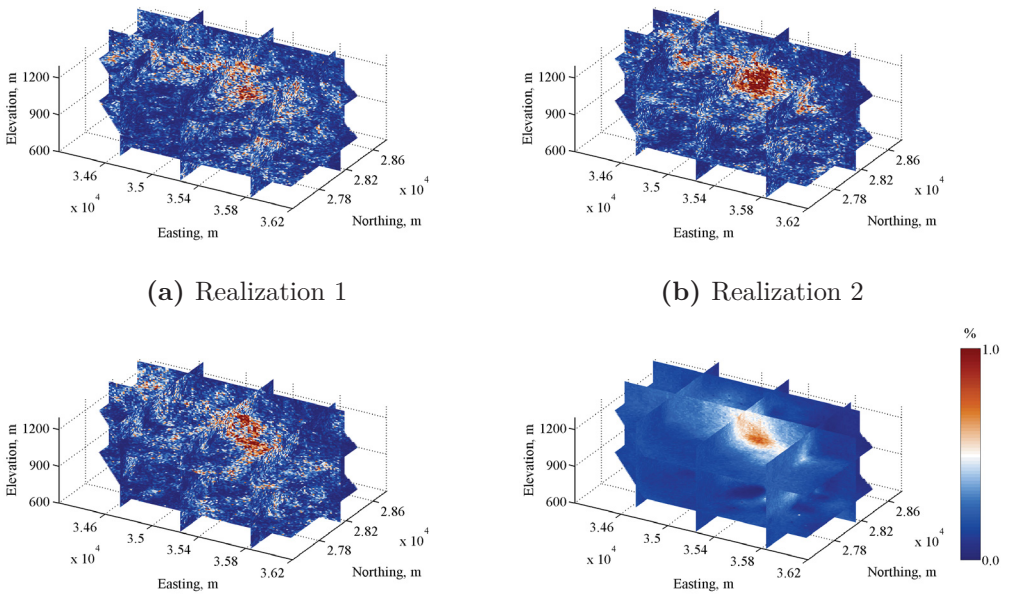
(d) Average values over 100 realizations

Figure 6.40: First three realizations and the averages over 100 realizations in NS units

6.6.4 Cross Validation

806 true values and the simulated average values in normal score units are compared. The local uncertainty appears accurate and the average uncertainty, 0.653, with the developed method in Figure 6.44. Although higher than the conventional method, the predicted uncertainty values are precise. Meanwhile, the numerical model by the developed method contains less variability than the model by the conventional non-stationary method due to the trend subtraction and addition are required by the conventional non-stationary method. The accuracy and precision of the developed method outperforms the conventional (non-)stationary geostatistical modeling methods.

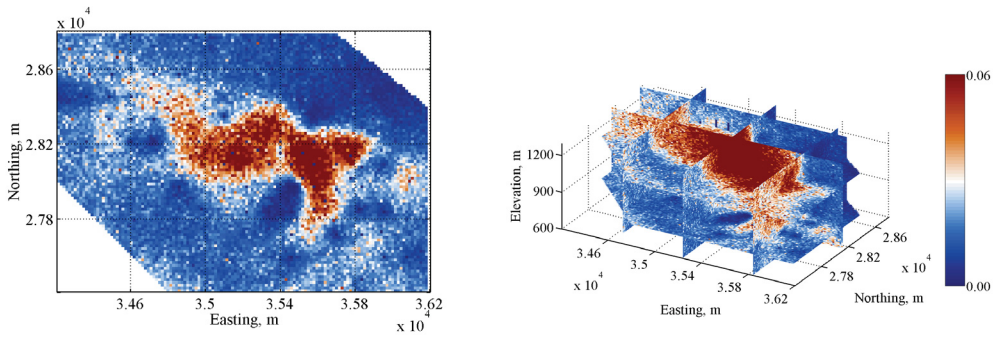
806 true values and the simulated average values are compared in original units. The mean of the developed method, 0.238%, is close to the true mean,



(a) Realization 1 **(b)** Realization 2

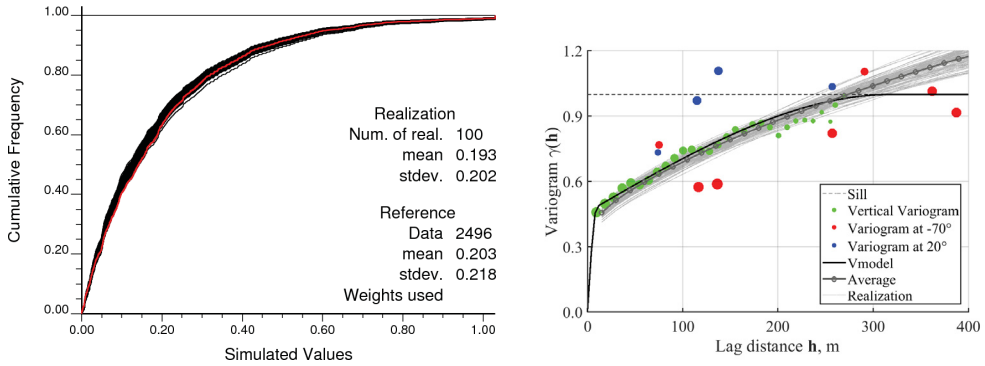
(c) Realization 3 **(d)** Average values over 100 realizations

Figure 6.41: First three realizations and the averages over 100 realizations in original units



(a) A slice at an elevation 775.5 meters **(b)** A volume in a 3-D view

Figure 6.42: Local variance over 100 realizations in original units



(a) Histogram reproduction **(b)** Variogram reproduction

Figure 6.43: Histogram and variogram reproduction from 2,496 modeling data in original units

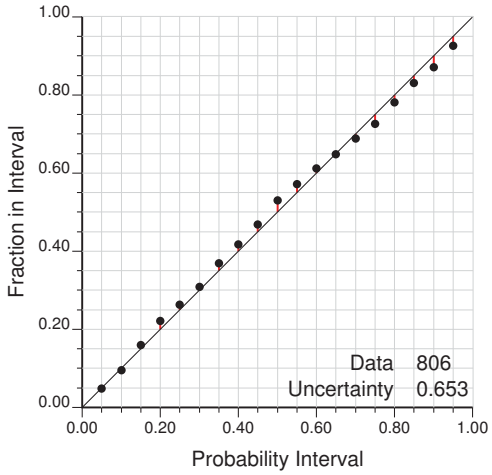
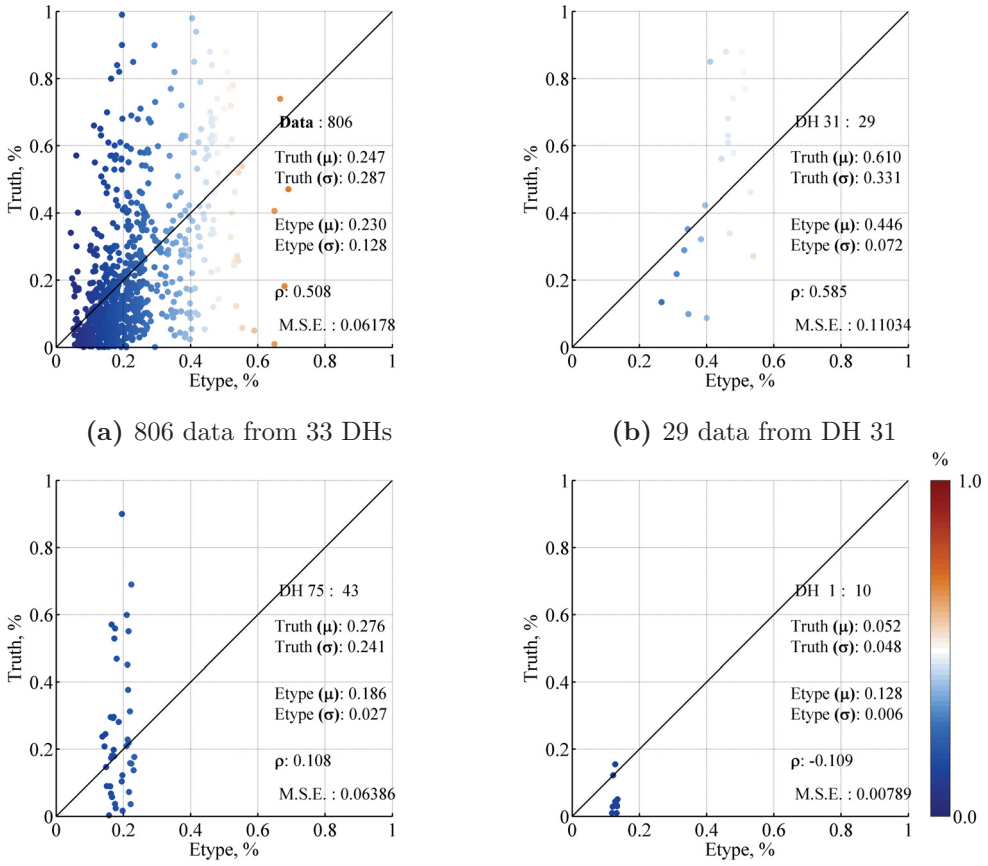


Figure 6.44: Cross validation from 806 test data with the proposed geostatistical modeling in NS units

0.247%. The standard deviation of the average values with the developed method is 0.130%. The result by the proposed method contains the most variance than those by the conventional methods, 0.108% and 0.118%. The mean squared error values between the true values and the average values is 0.06118. It shows a 2.95% improvement in the developed method than the conventional stationary method. 3 drillholes are compared and shown in Figure 6.45 showing 8.37%,

4.49%, and 87.11% improvements, respectively. The developed method shows a significant improvement, especially near the edges of the model.



(a) 806 data from 33 DHs

(b) 29 data from DH 31

(c) 43 data from DH 75

(d) 10 data from DH 1

Figure 6.45: Mean squared error values from 806 test data with the proposed geostatistical modeling in original units

6.7 Remarks

A real porphyry copper deposit with an obvious trend is used to demonstrate the practical framework for non-stationary geostatistical techniques. The data are divided into a modeling set and a test set. The modeling set is used for

geostatistical modeling and the test data is used for checking the results. The performance of the numerical models, the reproduction of geological characterizations and the analysis of the local uncertainty are compared.

The assumption of stationarity is made in the conventional (non-)stationary geostatistical prediction. The map shows that the simulated results might be underestimated by the conventional stationary method. The trend dichotomy approach is also implemented. The optimal trend is removed prior to the simulation. Then the residual model is applied into the numerical model for the prediction. Finally, the optimal trend model is added to the simulated values. The constraint of non-negative simulated values is considered for the final simulated attributes. A significant improvement is found by using this detrending approach; however, the model contains more variability and high uncertainty.

A strategy of the stepwise conditional transformation with a Gaussian mixture model is proposed to improve the prediction accuracy. The assumption of stationarity is relaxed in the proposed method. The variable is transformed conditional to the trend. The proposed method is more accurate, but with greater uncertainty than the conventional stationary geostatistical modeling and less uncertainty than the conventional non-stationary modeling. The mean squared error comparisons show a modest yet important 2.95% improvement in the proposed method. Drillholes in the margins of the deposit show the greatest improvement.

The trend dichotomy approach could also be applied with the stepwise conditional transformation. The approach of modeling with residuals using Gaussian mixture models and adding the trend model $m^*(\boldsymbol{u})$ back in final models, that is, $r(\boldsymbol{u}) = z(\boldsymbol{u}) - m^*(\boldsymbol{u})$ then $z(\boldsymbol{u}) = m^*(\boldsymbol{u}) + r(\mathbf{u})$, is also implemented, but it is not demonstrated in this chapter. The local uncertainty is 0.861 that is higher than the proposed technique of modeling with data directly. The mean squared error value between the truth and the simulated results is 0.06159, indicating a 0.03%

loss. The results of 3 drillholes show 0.78%, 0.07%, and 12.29% worse than the proposed method of modeling with the data. The performance of modeling with the trend dichotomy approach using Gaussian mixture models is not as good as the proposed method, which models the data more accurately. In addition, the constraint of the non-negative simulated values $z(\boldsymbol{u}) \geq 0.0$ is required. Thus, the Gaussian mixture model approach should be applied to the grade values directly.

Chapter 7

Concluding Remarks

A methodology for geostatistical simulation in the presence of a trend has been developed and demonstrated with numerous 2-D synthetic examples and two real case studies. The contributions and future work for the research and practical application are summarized in this chapter.

7.1 Summary of Contributions

Geostatistics is used for predicting spatial variability. Conventional geostatistical techniques begin with a set of assumptions such as stationarity. Stationarity refers to the properties at all locations in the statistics of the region of interest, that is, the regionalized variable does not change over the domain. Real geologic data often exhibit trend-like features that represent the large-scale variability of the regionalized variable. The assumption of stationarity is not satisfied in the presence of a trend. The trend should be treated to lead more accurate estimates.

The research focuses on developing and enhancing geostatistical algorithms suited to modeling in the presence of non-stationary variables. The first main contribution is to develop a function to model the deterministic component that

accounts for all available data, anisotropy and other considerations in an objective manner. The second main contribution is to establish a framework for geostatistical techniques using the modeled non-stationary component. Furthermore, in addition to the small 2-D illustrative examples, the research is applied to realistic case studies.

The contributions of this thesis can be highlighted as follows:

- **Placing emphasis on understanding and modeling the trend features that characterize the large-scale features of a regionalized variable**

  Geological explanations highly depend on observed data. These observations are collected for understanding, for example, establishing a geological model, identifying an estimate of a geological variable and analyzing the spatial relationship between variables (Sarma, 2009). A numerical model that is built with geostatistics can be used to quantify and predict the spatial variation of these observations.

  Most spatial data show some spatial dependence, for example, the mineralization displays the preference that high grades concentrate in some areas and low grades are in other areas. This locally varying spatial dependence features should be taken into consideration.

  Trend modeling analysis is a procedure for separating the relatively large-scale changes from the essential small-scale variations. The approach of dealing with a non-stationary variable is the data decomposition into a trend component with large-scale features and a stochastic component with small-scale variations in Equation 2.17. The scales of the large-scale features refers to the data spacing. The large-scale features represent the deterministic trend $m(\boldsymbol{u})$ that conveys some important geological structures

at the large-scale, while the residual data reflects some spatial geological features and randomized geological events.

- **The decision on the trend modeling function and the optimal selection of trend parameters to avoid excessive subjectivity in trend modeling**

  The trend function using the moving averages with a Gaussian kernel weighting function is determined to be free of artifacts. The weighted averaged trend reflects smooth large-scale features of the regionalized variable, and the remaining component is a fluctuation. The benefit of the moving average function is simplicity, fast calculation, ease of programming and reasonable estimates for different types of geologic data.

  The optimal selection of parameters for trend modeling are considered including anisotropy, background value, smoothing parameter and declustering weights in Chapter 3.2. The anisotropy includes the continuity directions and the length scales of the regionalized variable in the primary directions. A stable anisotropy factor in trend modeling is proposed and implemented. A small amount of background value is considered to smooth the local anomalies. In this study, data within the defined length scales are taken into consideration. The requirement of a background value makes it possible to smooth data by using successively neighboring points around each location and to ensure computational stability. The smoothing parameter is introduced to control the variance of the Gaussian kernel function and simplify the length scale settings. Different trend values can be constructed by changing the smoothing parameters. In real mining or petroleum applications, geologic data are often sampled unevenly. Declustering and debiasing techniques are required in trend modeling to generate

a representative distribution of the regionalized variable. All trend parameters have been calibrated with a 2-D synthetic example where a true trend model is available and where the cross validation is established to validate the parameters. The trend parameters can be calculated in an objective and repeatable manner.

- **Establishment of the trend modeling function in terms of trend extrapolation. The trend values in the extrapolation regions follow the continuity from the interpolation regions.**

  The trend prediction using the standard moving averages algorithm may not work as well in some scenarios, for example, the trend model is unreliable with less or even no controlling samples. The trend model with the standard non-negative Gaussian weights could not be inferred properly in regions of beyond the data locations.

  Extrapolation is an estimation value by extending a sequence of known values beyond the last known value in the attribute of interest. The moving averages method with positive and negative weights is proposed in Chapter 3.4 so that the trend model could be extrapolated in a more reasonable way. The non-positive weights are only assigned in the extrapolation regions. The assumption underlying the trend extrapolation method is that the trend values in interpolation regions continue to be extended in extrapolation regions. The extrapolation regions must be preliminarily informed before modeling. The implementation on a 1-D example is demonstrated. The number of samples that are assigned to the positive weights and the negative weights leads to different results in trend modeling.

- **An objective function for optimum trend modeling. A smoothing parameter that minimizes the absolute correlation between the**

trend and residuals is the basis for inferring a reasonable trend model.

The trend optimization is an important part of trend modeling. The goal is to infer a general objective function that could find an optimal smoothing parameter, and further, generate an optimal trend associated with that optimum smoothing parameter.

The criterion of the objective trend modeling considers both the deterministic and stochastic components of the regionalized variable. The first numerical experiment in Chapter 3.3.2 is to determine a relationship function between the deterministic trend and stochastic residuals. $|\rho\{m(\boldsymbol{u}), R(\boldsymbol{u})\}|$ is a conventional way to summarize the relationship. The second numerical experiment is to find an appropriate adjustment factor associated with the proposed correlation function. A factor of 2 works consistently. The optimal smoothing parameter is then determined in Equation 3.15. This objective function is a robust and reliable measurement so that it leads to construct a reasonable trend model and improve predictions at unsampled locations.

- **Construction of conditional distributions using Gaussian mixture models with different fitting data.**

  The Gaussian mixture model is introduced to reduce binning artifacts in the forward conditional transformation. The benefit of the Gaussian mixtures allows the easy assessment of any conditional distribution and suits high dimensional data spending significantly lower computational cost.

  The co-located trend values with the data are considered in the construction of mixture models; however, some important structures in the trend are not captured. Gaussian mixtures must be modeled properly. The spatial trend structures in the data are decomposed into several Gaussian

mixtures. Discarding some available trend values may introduce the conditional bias in estimates. Accounting for the trend values at unevenly sampled locations is significant. It helps to form a representative conditional transformation. The marginalized distributions of the observed data and the exhaustive trend must be in a standard normal score units. The test shows that Gaussian mixture model fitted with all trend values performance the best.

- **A geostatistical modeling framework using a parametric conditional transformation to account for non-stationary features of regionalized variables in an artifact-free fashion**

  A modified stepwise conditional transformation for geostatistical modeling is implemented. The stepwise conditional transformation attempts to remove the trend-like features from the regionalized variable and de-correlate the trend values with the regionalized variable at a zero lag distance to form an uncorrelated Gaussian variable. The standard geostatistical algorithm is attributed to the stationarity of the transformed variable. The back transformation reintroduces the original trend-like features. The use of the Gaussian mixtures eliminates the artifacts from data binning in the conventional stepwise conditional transformation. This stepwise conditional transformation with Gaussian mixture models is suitable for a high dimensional dataset and no artifacts are introduced.

  The example addresses the application of the parametric stepwise conditional transformation. 2-D synthetic data with an apparent trend are transformed conditionally by a parametric transformation. This parametric conditional transformation brings a low uncertainty and more accuracy to the numerical models. The correlation structures, the data distribution and the spatial relationship are well reproduced.

- **Applications on the real data with an obvious trend. The proposed algorithm improves the performance of non-stationary geostatistical modeling.**

  A practical framework for non-stationary geostatistical techniques using the parametric conditional transformation is established. The comparison with the standard geostatistical technique is made. The performance of numerical models is compared.

  The data are divided into a modeling set and a test set. The modeling set is used for proceeding the geostatistical modeling and the test data is used for checking the results. The assumption of stationarity is made in the conventional geostatistical prediction and loose in the developed method. The conventional modeling includes the normal score transformation, variogram fitting, sequential Gaussian simulation and normal score back transformation. The developed modeling requires the construction and the optimization of the trend model. A conditional transformation is considered to transform the data in the presence of the modeled trend and transformed data reach to a stationary regionalized variable. Standard Gaussian simulation is applied. The back transformation brings all trend features into original units.

  The validations are made by comparing the simulated results with the test data. The results show significant improvements if the trend model is used for the prediction instead of considering the original observed values directly, especially in the margins of the domain.

The programs related to the trend modeling and modeling with a trend are written in FORTRAN 90 following a GSLIB-style format and shown in Appendix B.

7.2 Limitations and Future Work

Limitations still remain and further research is required. A numbers of implementation details could be considered to improve performance of non-stationary geostatistical modeling.

- **The implementation with other trend modeling functions**

  It should be noted that the approach of the moving averages may not always provide the best trend function. The calculation of the trend model requires relative high computational time when a large amount of data is incorporated. It is worth trying other trend functions that are more straightforward and efficient.

  The trend model is difficult to construct beyond the range of available data. Exhaustive and auxiliary secondary data cound be introduced and merged with the primary non-stationary data together to generate more stable trend models in the regions with less primary data. In this case, the trend extrapolation is also solved. The purpose is to gain the confidence in the trend modeling with less primary data and improve geostatistical modeling.

- **The implementation with other possible objective functions for trend modeling**

  An objective function associated with the correlation at data locations is used for determining the optimal trend model. However, the objective function does not work well in obvious trend extrapolations. The difficulty is that the model could not be achieved by an optimal removal of the unwanted trend. A new objective function would be proposed to account for the trend model in the interpolation regions, as well as the extrapolation regions.

- **The optimal components for Gaussian mixture model fitting**

  A visual inspection of the optimal components of Gaussian mixtures is a common approach. 2-5 components for the Gaussian mixture modeling are recommended in this thesis. The decision on the number of components is subjective and depends on the practitioner. Chen and Kalbfleisch (1996) have demonstrated the importance of choosing an optimal number of Gaussian components. The mixture model may over-fit the data if too many components are used, while the mixture models may not be enough to capture the true underlying data structure if too few components are considered. The selection for the number of Gaussian mixtures is not only for theoretical interest, but also contains significant useful in the practical applications.

  The decision on the number of Gaussian components should be determined objectively. The re-sampling technique, for example, the spatial bootstrap technique implemented by Deutsch (2004) and uncertainty measurement proposed by Yamamoto et al. (2014) seems a promising start.

- **Data in the presence of trend after the parametric stepwise conditional transformation should be multi-Gaussian at all lag distances**

  The covariance after the stepwise conditional transformation is zero at a lag distance $\boldsymbol{h} = 0$ and may not be zero at other lag distances that could affect the result (Leuangthong and Deutsch, 2003). The use of Minimum/-Maximum Autocorrelation Factors (MAF) (Desbarats and Dimitrakopoulos, 2000) may be considered on the transformed variables if remnant cross spatial correlation is present. There are challenges related to the unequal sampling of the data.

- **The parametric stepwise conditional transformation with multivariate variables in the presence of different trends**

  More than one variable is often considered in mineral deposits. There may be multiple valuable elements and contaminants in the presence of trends. Accounting for the multivariate non-stationary data significantly increases the complexity of geostatistical modeling.

  Multivariate geologic data in the presence of trends should be considered simultaneously in numerical models used for resource management. The specification of the direct- and cross-covariance functions in the presence of trends refers to the spatially varying linear model of coregionalization (Gelfand et al., 2004). The misrepresentation of the multivariate non-stationary spatial features may lead to incorrect decision-making. Multivariate non-stationary regionalized variables could be considered simultaneously in a hierarchical workflow when using the parametric stepwise conditional transformation. Each variable could be processed according to the proposed workflow in Figure 4.4, and then another Gaussian mixture models could be fitted to the de-trended multivariate variables. A second parametric stepwise conditional transform would remove the dependency between the multivariate variables. Gaussian simulation of the independent factors would proceed, then the back transformation would be performed in reverse order to account for multivariate dependencies and the non-stationary trend features.

  Another challenge on the applications with non-stationary multivariate data arises because multiple data sources may come from different scales, for example, the geophysical data are much larger than the drill core data

(Kupfersberger et al., 1998; Ren, 2007). Geologic data also exhibit deterministic trends at different scales. Understanding multivariate non-stationary variables with scale differences will better describe the heterogeneity at all relevant scales and improve the prediction of resource management. The grid-free simulation approach proposed by Zagayevskiy (2015) provides a flexible implementation of the multivariate conditional simulation that addresses the multi-scale problem. One promising method is to represent all available data including the trends, secondary variables and primary variable at the same scale using the point-scale block value representation (Zagayevskiy, 2015). This will simplify the representation of the data and will not involve an explicit downscaling or upscaling. The research will be required to understand the limitations and applications.

More research should be applied to increase the robustness and reliability of geostatistical modeling algorithms and reduce spatial uncertainty in the presence of non-stationary features. The proposed research should also be implemented in realistic sets of data and validated to show the improvement brought by the proposed methods.

Bibliography

Agterberg, F. P. (1970). Autocorrelation functions in geology. In Merriam, D. F., editor, *Geostatistics: A Colloquium*, pages 113–141. Springer.

Armstrong, M. (1984). Problems with universal kriging. *Journal of the International Association for Mathematical Geology*, 16(1):101–108.

Arthur, D. and Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035. Society for Industrial and Applied Mathematics.

Barnett, R. M., Manchuk, J. G., and Deutsch, C. V. (2014). Projection pursuit multivariate transform. *Mathematical Geosciences*, 46(3):337–359.

Bartier, P. M. and Keller, C. P. (1996). Multivariate interpolation to incorporate thematic surface data using inverse distance weighting (IDW). *Computers & Geosciences*, 22(7):795–799.

Boisvert, J. B., Rossi, M. E., Ehrig, K., and Deutsch, C. V. (2013). Geometallurgical modeling at Olympic Dam Mine, South Australia. *Mathematical Geosciences*, 45(8):901–925.

Brunsdon, C., Fotheringham, A. S., and Charlton, M. (2002). Geographically weighted summary statistics: a framework for localised exploratory data analysis. *Computers, Environment and Urban Systems*, 26(6):501–524.

Brunsdon, C., Fotheringham, S., and Charlton, M. (1998). Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(3):431–443.

Burrough, P. A., McDonnell, R., and Lloyd, C. D. (2015). *Principles of Geographical Information Systems*. Oxford: Oxford University Press, 3 edition.

Chatfield, C. (2004). *The Analysis of Time Series: an Introduction*. Boca Raton: CRC, 6 edition.

Chen, J. and Kalbfleisch, J. D. (1996). Penalized minimum-distance estimates in finite mixture models. *Canadian Journal of Statistics*, 24(2):167–175.

Chiles, J. P. and Delfiner, P. (2012). *Geostatistics: Modeling Spatial Uncertainty*. New York: Wiley, 2 edition.

Choquette, A. L. (1955). The exploration desk: the blue ridge member of the graminia formation. *Journal of the Alberta Society of Petroleum Geologists*, 3(5):70–73.

Crain, I. K. and Bhattacharyya, B. K. (1967). Treatment of non-equispaced two-dimensional data with a digital computer. *Geoexploration*, 5(4):173–194.

Davis, J. C. (2002). *Statistics and Data Analysis in Geology*. New York: Wiley, 3 edition.

Delalleau, O., Courville, A., and Bengio, Y. (2012). Efficient EM training of Gaussian mixtures with missing data. *arXiv*, pages 1–6.

Delfiner, P. (1976). Linear estimation of non stationary spatial phenomena. In M., G., M., D., and C., H., editors, *Advanced Geostatistics in the Mining Industry*, volume 24, pages 49–68. Springer.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38.

Desbarats, A. J. and Dimitrakopoulos, R. (2000). Geostatistical simulation of regionalized pore-size distributions using min/max autocorrelation factors. *Mathematical Geology*, 32(8):919–942.

Deutsch, C. V. (2000). MIN E 650: Special topic in geostatistics. Technical report, University of Alberta, Edmonton, Canada.

Deutsch, C. V. (2004). A statistical resampling program for spatially correlated data: Spatial Boostrap. *Centre for Computational Geostatistics Annual Report*, 06(401):1–9.

Deutsch, C. V. (2010a). Citation program in applied geostatistics: Trend modeling. Technical report, Centre for Computational Geostatistics, Edmonton, Canada.

Deutsch, C. V. (2010b). Display of cross validation/ jackknife results. *Centre for Computational Geostatistics Annual Report*, 12(406):1–4.

Deutsch, C. V. and Journel, A. G. (1998). *GSLIB: Geostatistical Software Library and User's Guide*. New York: Oxford University Press, 2 edition.

Deutsch, M. V. (2015). Geostatistics lessons: The angle specification for GSLIB software. Technical report, Centre for Computational Geostatistics, Edmonton, Canada, http://www.geostatisticslessons.com/lessons/anglespecification.

Fotheringham, A. S., Brunsdon, C., and Charlton, M. (2002). *Geographically Weighted Regression: the Analysis of Spatially Varying Relationships*. Chichester: Wiley.

Friedman, J. H. (1987). Exploratory projection pursuit. *Journal of the American Statistical Association*, 82(397):249–266.

Gelfand, A. E., Schmidt, A. M., Banerjee, S., and Sirmans, C. F. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. *Test*, 13(2):263–312.

Gilardi, N., Bengio, S., and Kanevski, M. (2002). Conditional Gaussian mixture models for environmental risk mapping. In *Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing*, pages 2507–2514.

Gonzalez, E., McLennan, J. A., and Deutsch, C. V. (2006). A new approach to SGS with a trend: Nonstationary transformation tables. *Centre for Computational Geostatistics Annual Report*, 08(120):1–13.

Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. New York: Oxford University Press.

Goovaerts, P. (2000). Estimation or simulation of soil properties? An optimization problem with conflicting criteria. *Geoderma*, 97(3):165–186.

Grant, F. (1957). A problem in the analysis of geophysical data. *Geophysics*, 22(2):309–344.

Gray, A. G. and Moore, A. W. (2003). Nonparametric density estimation: toward computational tractability. In *Proceedings of the SIAM International Conference on Data Mining*, pages 203–211.

Harbaugh, J. W. and Merriam, D. F. (1968). *Computer Applications in Stratigraphic Analysis*. New York: Wiley.

Heaton, M. J., Christensen, W. F., and Terres, M. A. (2017). Nonstationary Gaussian process models using spatial hierarchical clustering from finite differences. *Technometrics*, 59(1):93–101.

Hong, S. and Deutsch, C. V. (2009). 3D trend modeling by combining lower order trends. *Centre for Computational Geostatistics Annual Report*, 11(130):1–14.

Howarth, R. J. (2001). A history of regression and related model-fitting in the earth sciences (1636?-2000). *Natural Resources Research*, 10(4):241–286.

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

Journel, A. G. (1974). Geostatistics for conditional simulation of ore bodies. *Economic Geology*, 69(5):673–687.

Journel, A. G. (1986). Geostatistics: models and tools for the earth sciences. *Mathematical Geology*, 18(1):119–140.

Journel, A. G. and Huijbregts, C. J. (1978). *Mining Geostatistics*. London: Academic Press.

Journel, A. G. and Rossi, M. E. (1989). When do we need a trend model in kriging? *Mathematical Geology*, 21(7):715–739.

Kariya, T. and Kurata, H. (2004). *Generalized Least Squares*. Chichester: Wiley.

Kim, H. M., Mallick, B. K., and Holmes, C. C. (2005). Analyzing nonstationary spatial data using piecewise Gaussian processes. *Journal of the American Statistical Association*, 100(470):653–668.

Knotters, M., Brus, D. J., and Voshaar, J. H. O. (1995). A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. *Geoderma*, 67(3-4):227–246.

Krige, D. G. (1976). A review of the development of geostatistics in south africa. In M., G., M., D., and C., H., editors, *Advanced Geostatistics in the Mining Industry*, pages 279–293. Springer.

Krumbein, W. C. (1937). Sediments and exponential curves. *The Journal of Geology*, 45(6):577–601.

Krumbein, W. C. (1959). Trend surface analysis of contour-type maps with irregular control-point spacing. *Journal of Geophysical Research*, 64(7):823–834.

Kupfersberger, H., Deutsch, C. V., and Journel, A. (1998). Deriving constraints on small-scale variograms due to variograms of large-scale data. *Mathematical Geology*, 30(7):837–852.

Lavrenko, V. (2014). Expectation maximization: how it works. https://www.youtube.com/watch?v=iQoXFmbXRJA.

Leuangthong, O. (2003). *Stepwise Conditional Transformation for Multivariate Geostatistical Simulation*. Doctoral dissertation, University of Alberta.

Leuangthong, O. and Deutsch, C. V. (2003). Stepwise conditional transformation for simulation of multiple variables. *Mathematical Geology*, 35(2):155–173.

Leuangthong, O. and Deutsch, C. V. (2004). Transformation of residuals to avoid artifacts in geostatistical modelling with a trend. *Mathematical Geology*, 36(3):287–305.

Leuangthong, O., Khan, K. D., and Deutsch, C. V. (2011). *Solved Problems in Geostatistics*. New York: Wiley.

Little, R. J. A. and Rubin, D. B. (2014). *Statistical Analysis with Missing Data*. New Jersey: Wiley, 2 edition.

Lloyd, C. D. (2011). *Local Models for Spatial Analysis*. Boca Raton: CRC, 2 edition.

Machuca-Mory, D. F. (2010). *Geostatistics with Location-Dependent Statistics*. Doctoral dissertation, University of Alberta.

Machuca-Mory, D. F. and Deutsch, C. V. (2013). Non-stationary geostatistical modeling based on distance weighted statistics and distributions. *Mathematical Geosciences*, 45(1):31–48.

Manchuk, J. G. and Deutsch, C. V. (2011). A short note on trend modeling using moving windows. *Centre for Computational Geostatistics Annual Report*, 13(403):1–8.

Mandelbaum, H. (1963). Statistical and geological implications of trend mapping with nonorthogonal polynomials. *Journal of Geophysical Research*, 68(2):505–519.

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

Matheron, G. (1971). Theory of regionalized variables and its applications. *Cah. Centre Morrphol. Math.*, 5:1–211.

Matheron, G. (1973). The intrinsic random functions and their applications. *Advances in Applied Probability*, pages 439–468.

McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*. New York: Wiley.

McLennan, J. A. (2007). *The Decision of Stationarity*. Doctoral dissertation, University of Alberta.

McLennan, J. A., Leuangthong, O., and Deutsch, C. V. (2005). Trend modeling techniques and guidelines. *Centre for Computational Geostatistics Annual Report*, 07(104):1–23.

Metropolis, N. and Ulam, S. (1949). The monte carlo method. *Journal of the American Statistical Association*, 44(247):335–341.

Miller, R. L. (1956). Trend surfaces: their application to analysis and description of environments of sedimentation. *The Journal of Geology*, 64(5):425–446.

Myers, D. E. (1989). To be or not to be... stationary? That is the question. *Mathematical Geology*, 21(3):347–362.

Olea, R. A. (1974). Optimal contour mapping using universal kriging. *Journal of Geophysical Research*, 79(5):695–702.

Oliver, M. A. and Webster, R. (2015). *Basic Steps in Geostatistics: the Variogram and Kriging*. Switzerland: Springer.

Opsomer, J., Wang, Y., and Yang, Y. (2001). Nonparametric regression with correlated errors. *Statistical Science*, pages 134–153.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076.

Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. A*, 185:71–110.

Pyrcz, M. J. and Deutsch, C. V. (2014). *Geostatistical Reservoir Modeling*. Oxford: Oxford University Press, 2 edition.

Pyrcz, M. J., Gringarten, E., Frykman, P., and Deutsch, C. V. (2003). Representative input parameters for geostatistical simulation. *Centre for Computational Geostatistics Annual Report*, 05(118):1–25.

Qu, J. and Deutsch, C. (2014). Estimation and simulation in presence of non-stationary mean and standard deviation. *Centre for Computational Geostatistics Annual Report*, 16(111):1–9.

Ren, W. (2005). Short note on conditioning turning bands realizations. *Centre for Computational Geostatistics Annual Report*, 07(405):1–11.

Ren, W. (2007). *Exact Downscaling in Reservoir Modeling*. Doctoral dissertation, University of Alberta.

Ren, W., Cunha, L., and C.V., D. (2004). Preservation of multiple point structure when conditioning by kriging. *Centre for Computational Geostatistics Annual Report*, 06(108):1–21.

Rivoirard, J. (1987). Two key parameters when choosing the kriging neighborhood. *Mathematical Geology*, 19(8):851–856.

Roberts, W. J. (2010). Application of a Gaussian, missing-data model to product recommendation. *IEEE Signal Processing Letters*, 17(5):509–512.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, 23(3):470–472.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837.

Rossi, M. E. and Deutsch, C. V. (2014). *Mineral Resource Estimation*. Dordrecht: Springer.

Sadler, P. M. (1981). Sediment accumulation rates and the completeness of stratigraphic sections. *The Journal of Geology*, 89(5):569–584.

Sarma, D. D. (2009). *Geostatistics with Applications in Earth Sciences*. Dordrecht: Springer, 2 edition.

Sheiner, L. B. and Beal, S. L. (1981). Some suggestions for measuring predictive performance. *Journal of Pharmacokinetics and Pharmacodynamics*, 9(4):503–512.

Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. *Proceedings of the 1968 23rd ACM National Conference*, pages 517–524.

Silva, D. S. F. and Deutsch, C. V. (2015). Tool for univariate KDE with optimal bandwidth based on bootstrap. *Centre for Computational Geostatistics Annual Report*, 17(133):1–8.

Silva, D. S. F. and Deutsch, C. V. (2016). Multivariate data imputation using Gaussian mixture models. *Spatial Statistics*.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Stoer, J. and Bulirsch, R. (2002). *Introduction to Numerical Analysis*. New York: Springer, 3 edition.

Switzer, S. B., Holland, W. G., Christie, D. S., Graf, G. C., Hedinger, A. S., McAuley, R. J., Wierzbicki, R. A., and Packard, J. J. (1994). Devonian woodbend-winterburn strata of the western canada sedimentary basin. *Geological Atlas of the Western Canada Sedimentary Basin: Canadian Society of Petroleum Geologists and Alberta Research Council*, pages 165–202.

Tily, R. and Brace, C. J. (2006). A study of natural neighbour interpolation and its application to automotive engine test data. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 220(7):1003–1017.

Villalba, M. E. and Deutsch, C. V. (2011). A conditional finite domain implementation with LU simulation. *Centre for Computational Geostatistics Annual Report*, 13(125):1–10.

Wackernagel, H. (2003). *Multivariate Geostatistics*. Berlin: Springer, 3 edition.

Wang, G., Carranza, E. J. M., Zuo, R., Hao, Y., Du, Y., Pang, Z., Sun, Y., and Qu, J. (2012). Mapping of district-scale potential targets using fractal models. *Journal of Geochemical Exploration*, 122:34–46.

Watson, G. S. (1972). Trend surface analysis and spatial correlation. *Geological Society of America Special Papers*, 146:39–46.

Webster, R. and Oliver, M. A. (2008). *Geostatistics for Environmental Scientists*. Chichester: Wiley, 2 edition.

Wren, A. E. (1973). Trend surface analysis: a review. *Canadian Journal of Exploration Geophysics*, 9(1):39–44.

Yamamoto, J. K., Koike, K., Kikuda, A. T., and Campanha G. A. da C. C., E. A. (2014). Post-processing for uncertainty reduction in computed 3d geological models. *Tectonophysics*, 633:232–245.

Zagayevskiy, Y. (2015). *Multivariate Geostatistical Grid-Free Simulation of Natural Phenomena*. Doctoral dissertation, University of Alberta.

Appendix A

Appendix - Description of Trend Extrapolation

The problem of the extrapolation in trend modeling at an unsampled location using the available samples is considered, see the 1D schematic illustration in Figure 3.25a. $\boldsymbol{u}_0$ represents the location to be modeled, while black dots are the known observed samples. These ten samples are labeled with numbers (p_1 to p_{10}) by the distances from the unsampled location $\boldsymbol{u}_0$. A novel weighting function in Figure 3.25c that accounts for the positive weights and negative weights is considered. The derivation of the expression is illustrated below.

Suppose a quadratic equation for expressing the weighting function $\{f(\boldsymbol{u}_i), i = 0, 1, \cdots, np\}$ with np samples given of the positive weights. The function is considered as:

$$\begin{aligned} f(\boldsymbol{u}_i) = a_0 + a_1 \cdot \boldsymbol{u}_i + a_2 \cdot \boldsymbol{u}_a^2 \\ \boldsymbol{u}_i \in (0, D_1) \text{ and } \boldsymbol{u}_i \in A \end{aligned} \tag{A.1}$$

here $\{\boldsymbol{u}_i, i = 0, 1, \cdots, np\}$ represent the data locations will be given with positive weights to the estimate in the domain A. D_1 is the distance that the weight stays

non-negative. a_0, a_1 and a_2 are constant numbers that will be decided by some constraints.

Three conditions are considered to generate the final positive weighting function $\{f(\boldsymbol{u}_i), i = 0, 1, \cdots, np\}$: (1) the total weight 1.0 is given at the location to be estimated, that is, $f(0) = a_0 + a_1 \cdot 0 + a_2 \cdot 0^2 = 1.0$; (2) the maximum weight is given at the location to be modeled, that is, $f^{'}(0) = a_1 + 2a_2 \cdot 0 = 0.0$; and (3) the weight equals to zero at the location of D_1, that is, $f(D_1) = a_0 + a_2 \cdot D_1^2 = 0.0$. Thus, the final equation for expressing the positive weights is written as:

$$f(\boldsymbol{u}_i) = 1.0 - \frac{1}{D_1^2} \cdot \boldsymbol{u}_i^2 \qquad \boldsymbol{u}_i \in (0, D_1) \text{ and } \boldsymbol{u}_i \in A \tag{A.2}$$

The area of the positive weighting zone (dark gray zone in Figure 3.25c) is expressed as:

$$\int_0^{D_1} f(\boldsymbol{u}_i) d(\boldsymbol{u}_i) = 1 + \Delta \tag{A.3}$$

here Δ is a factor that accounts for the sum of the negative weights. This integral is solved as follows:

$$\begin{aligned} \int_0^{D_1} (1 - \frac{1}{D_1^2} \cdot \boldsymbol{u}_i^2) d(\boldsymbol{u}_i) &= 1 + \Delta \\ 1 \cdot \boldsymbol{u}_i \Big|_0^{D_1} - \frac{1}{3} \cdot \frac{1}{D_1^2} \cdot \boldsymbol{u}_i^3 \Big|_0^{D_1} &= 1 + \Delta \\ 1 \cdot D_1 - \frac{1}{3} \cdot \frac{1}{D_1^2} \cdot D_1^3 &= 1 + \Delta \\ \frac{2 \cdot D_1}{3} &= 1 + \Delta \\ \Delta &= \frac{2 \cdot D_1}{3} - 1 \end{aligned} \tag{A.4}$$

Let us suppose an other equation for expressing the weighting function $\{g(\boldsymbol{u}_j), j = 0, 1, \cdots, nn\}$ with nn samples given of the negative weights. The

equation is given as:

$$g(\boldsymbol{u}_j) = b_0 + b_1 \cdot \boldsymbol{u}_j + b_2 \cdot \boldsymbol{u}_j^2 \qquad \boldsymbol{u}_j \in (D_1, D_2) \text{ and } \boldsymbol{u}_j \in A \tag{A.5}$$

here $\{\boldsymbol{u}_j, j = 0, 1, \cdots, nn\}$ represent the data locations will be given with negative weights to the estimate. D_1 is the distance that the weight stays non-negative, and D_2 is the distance between the data with the last negative weight and the next unused data. b_0, b_1 and b_2 are the constant numbers that will be considered by:

$$\begin{aligned} g(D_1) &= b_0 + b_1 \cdot D_1 + b_2 \cdot D_1^2 = 0 \\ g(D_2) &= b_0 + b_1 \cdot D_2 + b_2 \cdot D_2^2 = 0 \end{aligned} \tag{A.6}$$

A new system would be obtained by $g(D_2)$ subtracting $g(D_1)$, and get

$$\begin{aligned} b_1 \cdot (D_2 - D_1) + b_2 \cdot (D_2^2 - D_1^2) &= 0 \\ (D_2 - D_1) \cdot [b_1 + b_2 \cdot (D_2 + D_1)] &= 0 \end{aligned}$$

where $D_2 \neq D_1$, and then $b_1 = -b_2 \cdot (D_2 + D_1)$. Consider to put b_1 into $g(D_1)$, and obtain

$$\begin{aligned} b_0 - b_2 \cdot (D_2 + D_1) \cdot D_1 + b_2 \cdot D_1^2 &= 0 \\ b_0 - b_2 \cdot D_1 \cdot D_2 - b_2 \cdot D_1^2 + b_2 \cdot D_1^2 &= 0 \\ b_0 &= b_2 \cdot D_1 \cdot D_2 \end{aligned} \tag{A.7}$$

Thus, the final equation for expressing the negative weights is written as:

$$g(\boldsymbol{u}_j) = b_2 \cdot D_1 \cdot D_2 - b_2 \cdot (D_2 + D_1) \cdot \boldsymbol{u}_j + b_2 \cdot \boldsymbol{u}_j^2 \qquad \boldsymbol{u}_j \in (D_1, D_2) \text{ and } \boldsymbol{u}_j \in A \tag{A.8}$$

The area of the negative weighting zone (light gray zone in Figure 3.25c) is expressed as:

$$\int_{D_1}^{D_2} g(\boldsymbol{u}_j) d(\boldsymbol{u}_j) = -\Delta \tag{A.9}$$

where Δ standardizes the total weights. Then, this expression is solved as:

$$\begin{aligned}
\int_{D_1}^{D_2} (b_2 \cdot D_1 \cdot D_2 - b_2 \cdot (D_2 + D_1) \cdot \boldsymbol{u}_j + b_2 \cdot \boldsymbol{u}_j^2) d(\boldsymbol{u}_j) &= -\Delta \\
b_2 \cdot D_1 \cdot D_2 \cdot \boldsymbol{u}_j \Big|_{D_1}^{D_2} - \frac{1}{2} \cdot b_2 \cdot (D_2 + D_1) \cdot \boldsymbol{u}_j^2 \Big|_{D_1}^{D_2} + \frac{1}{3} \cdot b_2 \cdot \boldsymbol{u}_j^3 \Big|_{D_1}^{D_2} &= -\Delta \\
b_2 \cdot D_1 \cdot D_2 \cdot (D_2 - D_1) - \frac{1}{2} \cdot b_2 \cdot (D_2 + D_1) \cdot (D_2^2 - D_1^2) + \frac{1}{3} \cdot b_2 \cdot (D_2^3 - D_1^3) &= -\Delta \\
\frac{1}{6} \cdot b_2 \cdot (D_2 - D_1) \cdot [6 \cdot D_1 \cdot D_2 - 3 \cdot (D_2 + D_1)^2 - 2 \cdot (D_2^2 + D_1^2 + D_1 \cdot D_2) &= -\Delta \\
-\frac{1}{6} \cdot b_2 \cdot (D_2 - D_1)^3 &= -\Delta \\
b_2 = \frac{6}{(D_2 - D_1)^3} \cdot \Delta &
\end{aligned} \tag{A.10}$$

The final equation for expressing the negative weight is written as:

$$\begin{aligned}
g(\boldsymbol{u}_j) = \frac{6}{(D_2 - D_1)^3} \cdot \Delta \cdot D_1 \cdot D_2 - \frac{6}{(D_2 - D_1)^3} \cdot \Delta \cdot (D_2 + D_1) \cdot \boldsymbol{u}_j + \\
\frac{6}{(D_2 - D_1)^3} \cdot \Delta \cdot \boldsymbol{u}_j^2 \\
\boldsymbol{u}_j \in (D_1, D_2) \text{ and } \boldsymbol{u}_j \in A
\end{aligned} \tag{A.11}$$

Appendix B

Appendix - Description of Fortran Programs

The descriptions of the Fortran coded programs are provided in this Appendix. The presented programs are `VMODEL_AVG` for the calculation of the average anisotropy ratios, `TMODEL` for the construction of the smooth trend models, `TFUNCTION` for the selection of an optimal trend model among input trends, `TREND_EXTRP` for the creation of the trend models with the proposed extrapolation algorithm, and `SMP_ADD` for the correction of the bivariate distribution between the trend values and data values.

B.1 Program `VMODEL_AVG`

The provided `VMODEL_AVG` program is written in FORTRAN 90 and followed a GSLIB-style functionality which controlled by a text-based parameter file (Deutsch and Journel, 1998) and is presented in Section 3.1.4 of Chapter 3. A default parameter file of the program is shown in Table B.1. The descriptions are provided as follows.

Table B.1: Default parameter file of program `VMODEL_AVG`

```
1-                Parameters for VMODEL_AVG Program
2-                       ********************
3-
4-   START OF PARAMETERS:
5-   vmodel_avg.var                - file for modeled variogram
6-   vmodel_avg.dbg                - file for debugging output
7-   256 0.5 1.0                   -    nx, xmin, xsize
8-   128 0.5 1.0                   -    ny, ymin, ysize
9-    64 0.5 1.0                   -    nz, zmin, zsize
10-  3                             - number of directions
11-  135.0 0.0 1.0                 -    azm, dip, lag interval
12-   45.0 0.0 1.0                 -    azm, dip, lag interval
13-    0.0 0.0 1.0                 -    azm, dip, lag interval
14-  2  0.00                       - variogram: nst, nugget
15-  1  0.75    135.0   0.0   0.0  -    it, cc, ang1, ang2, ang3
16-              30.0  30.0  30.0  -    a1, a1, a3
17-  1  0.75    135.0   0.0   0.0  -    it, cc, ang1, ang2, ang3
18-           10000.0 700.0 200.0  -    a1, a1, a3
19-
20-  Additional Notes:
21-  - Line 11 should be specified with the major direction.
```

- **Line 1-4**: The headers of the parameter file. All input and output data files are in GeoEAS formats with the first four lines being a title.

- **Line 5**: The output file contains the conventional semi-variogram model values. Two additional columns, the area under the variogram model curves in each lag distance and the total areas from the origin to the specific lag distance, are added to the output file.

- **Line 6**: The debugging file contains some useful information about the semi-variogram models, such as the domain size, the area under the variogram models in each direction and the calculation of the recommended anisotropy ratios.

- **Line 7-9**: The grid parameters of X, Y and Z directions. The X, Y, Z axis are associated with the directions of East, North and Elevation. The coordinate system is established by specifying the center of the first grid (xmin, ymin, zmin), the number of grid nodes (nx, ny, nz), and the size of the grid nodes (xsize, ysize, zsize).
- **Line 10**: The number of directions for the variogram models that are needed to be considered.
- **Line 11-13**: The directions of the anisotropy are specified by the azimuth and dip. A unit lag offset must be also specified.
- **Line 14**: The number of the nested structures and the nugget effect.
- **Line 15-18**: The types of the structures, the sill contributions, the principal directions of the anisotropy, and the variogram ranges are required to set up the variogram models.
- **Line 21**: Additional Notes. It illustrates some useful information to help practitioners better understand some parameters in this program. Line 11 must be specified with the major direction. Line 12-13 can be either the minor direction or the vertical direction.

B.2 Program `TMODEL`

This program is prepared to model the trend, which is presented in Section 3.1 of Chapter 3. Trend models which are built through the smoothing parameter are defined by the GSLIB conventions. Different parameters depending on the practitioners could lead to a variety of trend modeling results. The default parameter file is shown in Table B.2 and detailed demonstrations are explained below.

Table B.2: Default parameter file of program `TMODEL`

```
1-                    Parameters for TMODEL Program
2-                         ********************
3-
4-   START OF PARAMETERS:
5-   ../data/data.dat       - file with data
6-   1 2 3                  -   columns for X, Y, Z
7-   4 5                    -   columns for variable and weight
8-   -998 1.0e21            -   trimming limits
9-   1.0 0.001              - smoothing parameter, background value
10-  50 0.5 1.0             -   nx, xmin, xsize
11-  50 0.5 1.0             -   ny, ymin, ysize
12-  50 0.5 1.0             -   nz, zmin, zsize
13-  2.0 1.0                - angles for three directions
14-  30.0 10.0 0.0          - anisotropy length scale
15-  ./trend.out            - file for trend output
16-  ./datapoints.out       - file for data output
17-  0                      - if export the weight at each location
18-  ./weight.out           - file for weight at each location
19-
20-  Additional Notes:
21-  - The weighting function is a Gaussian-like distribution and
22-    scaled by the smoothing parameter.
23-  - Two outputs in Line 15 and 16 containing the grided file and
24-    the point file are generated, respectively.
25-  - The weights that are assigned from each datum are exported
26-    if the option 1 in Line 17 is chosen.
```

- **Line 1-4**: The headers of the parameter file. All input and output data files are in GeoEAS formats with the first four lines being a title.

- **Line 5-8**: Specify the input reference distribution including X, Y, Z coordinate, the predictive variable and the declustering weight if applicable. The trimming limits apply to the input file.

- **Line 9**: The smoothing parameter controls the smoothness of the trend models. The defaulted parameter sets 1.0 that would be chosen reasonably by

practitioners. The smoothing parameter scales the range in the major direction, and further, scales the shape of Gaussian distribution. A small background is set for the stability of the algorithm.

- **Line 10-12**: The grid parameters of X, Y and Z directions. The X, Y, Z axis are associated with the direction of East, North and Elevation. The coordinate system is established by specifying the center of the first grid (xmin, ymin, zmin), the number of grid nodes (nx, ny, nz), and the size of the grid nodes (xsize, ysize, zsize).

- **Line 13-14**: Set the anisotropy ratios and directions. The stable anisotropy ratios are generated by the program `VMODEL_AVG`. Three anisotropy directions that are associated with the azimuth, dip and third arbitrary direction should be provided.

- **Line 15-16**: Output grid file and point file. The output grid file appends the gridded X, Y, Z coordinates, trend values, standard deviation of the trend. The output point file includes the original information from the input data file, and adds the predictive trend values, trend standard deviation, residuals and the standardized residuals to the file.

- **Line 17-18**: The weights that are assigned from each datum are exported (option = 1). Note that the program will run slowly if the option 1 is chosen.

- **Line 20-26**: Additional Notes. It illustrates some useful information to help practitioners better understand some parameters in this program. A moving average statistics with the Gaussian weighting function in this software is considered. Ideally, this software is not designed for the trend extrapolation.

B.3 Program TFUNCTION

This program is used to find an optimal trend model from several given modeled trends in program TMODEL, which is presented in Section 3.3.2 of Chapter 3. The parameters are shown in Table B.3 with some explanations.

Table B.3: Default parameter file of program TFUNCTION

```
1-                Parameters for TFUNCTION Program
2-                      ********************
3-
4-   START OF PARAMETERS:
5-   10                          - number of files
6-   ../data/datapoints_1.out    - file with data
7-   ../data/datapoints_2.out    - file with data
8-   ../data/datapoints_3.out    - file with data
9-   ../data/datapoints_4.out    - file with data
10-  ../data/datapoints_5.out    - file with data
11-  ../data/datapoints_6.out    - file with data
12-  ../data/datapoints_7.out    - file with data
13-  ../data/datapoints_8.out    - file with data
14-  ../data/datapoints_9.out    - file with data
15-  ../data/datapoints_10.out   - file with data
16-  9 12                        -    column for tmodel, tresidual
17-  2                           - adjust factor
18-  title                       - title
19-  tfunction.ps                - plot of the optimal function
```

- **Line 1-4**: The headers of the parameter file. All input and output data files are in GeoEAS formats with the first four lines being a title.
- **Line 5**: Specify the input number of trend model files. The number of files must match with the following settings.
- **Line 6-15**: Specify the names of the input trend model files. The number of the files should be consistent with Line 5. The trend model contains the predicted trend values and the residual values that is generated from the program TMODEL.

- **Line 16**: Specify the input reference distribution including the predicted trend column and the residual column from the input trend model files.
- **Line 17**: The setting for the adjustment factor which is presented in Section 3.3.2 of Chapter 3. The default parameter is $f = 2.0$. The users could change this adjustment factor with different scenarios.
- **Line 18**: The title of the PostScript plot. The default program includes a 40-character title in the top of the plot.
- **Line 19**: The file for the PostScript output.

B.4 Program `TREND_EXTRP`

The program `TREND_EXTRP` is written by the text-based parameter file. The theoretic algorithm is presented in Section 3.4 of Chapter 3. A default parameter file of the program is shown in Table B.4. The descriptions are provided as follows.

- **Line 1-4**: The headers of the parameter file. All input and output data files are in GeoEAS formats with the first four lines being a title.
- **Line 5-6**: Specify the input data file including X, Y, Z coordinate and data values.
- **Line 7**: The trimming limits apply to the predictive data values. The value beyond these limits will be discarded.
- **Line 8**: The option for the trend extrapolation. The program only works on a 1D case (option = 0). If 0 is chosen, the interpolation region ("from"and "to") must be given. If other options are chosen, no result will be obtained in this version.

Table B.4: Default parameter file of program `TREND_EXTRP`

```
1-                Parameters for TREND_EXTRP Program
2-                       ********************
3-
4-   START OF PARAMETERS:
5-   ../data/data.dat      - file with data
6-   0 0 1 7               -   columns for X, Y, Z, variable
7-   -998 1.0e21           -   trimming limits
8-   0 0.0 1000.0          - option: 0=1D, Interpolation Region
9-   20 20                 -   number of postive,negative weights
10-  1    0.5 1.0          -   nx, xmin, xsize
11-  1    0.5 1.0          -   ny, ymin, ysize
12-  1280 0.5 1.0          -   nz, zmin, zsize
13-  0 15                  - min and max value
14-  1                     - flag of debug file? (0=no;1=yes)
15-  trend_extrp.dbg       - file for debug
16-  data_extrp.out        - file for data output
17-  trend_extrp.out       - file for trend ouput
18-
19-  Additional Notes:
20-  - Program only works on 1D trend extrapolation.
```

- **Line 9**: The number of data that would be participated in the trend modeling. The first column requires the number of data that are assigned by the positive weights, and the second column requires the number of data that would be given by the negative weights. Different numbers for the positive and negative weights could lead to a variety of trend modeling results.

- **Line 10-12**: The X, Y, Z axis are associated with the direction of East, North and Elevation. The coordinate system is established by specifying the center of the first grid (xmin, ymin, zmin), the number of grid nodes (nx, ny, nz), and the size of the grid nodes (xsize, ysize, zsize).

- **Line 13**: The lower and upper limits that apply to trend values are given in this line. If the trend values to be modeled are lower or greater than the defined tails, the trend values will be given with the defined limits.

- **Line 14-15**: Some detailed information containing both the positive and the negative weighs are exported to the debugging output file. It can be indicated here using the value of 1. If the user is not interested in this information, the value should be set to 0.
- **Line 16**: Output data file contains input data file appending the trend values and the residuals at the data locations.
- **Line 17**: Output grid file appends the gridded X, Y, Z coordinates and the trend values.
- **Line 19-20**: Additional Notes. The current program only works on a 1D trend extrapolation.

B.5 Program `SMP_ADD`

This program `SMP_ADD` in Section 4.4.1 of Chapter 4 allows for generating trend values from the exhaustive trend model and matching the target trend distribution. Data values are updated with the iteration. The bivariate distribution between the trend values and data values is corrected by this program. The parameters are shown in Table B.5 with some explanations.

- **Line 1-4**: The headers of the parameter file. All input and output data files are in GeoEAS formats with the first four lines being a title.
- **Line 5-6**: Specify the input data file only including the exhaustive trend model and data values.
- **Line 7**: The trimming limits apply to both the exhaustive trend values and data values. Variables with values beyond these limits will be discarded.
- **Line 8**: The user defined bandwidth value is provided here. The distribution will be more smoother if a large kernel bandwidth is set.

Table B.5: Default parameter file of program `SMP_ADD`

```
1-                    Parameters for SMP_ADD Program
2-                         ********************
3-
4-   START OF PARAMETERS:
5-   ../data/data.dat       - file with data
6-   1 2                    -   columns for exhaustive variable
7-   -998 1.0e21            -   trimming limits
8-   0.01                   - kernel bandwidth
9-   2000                   - number of discretization
10-  10                     - number of classes
11-  1                      - variance correction: 0=constant, 1=linear
12-  0.90                   - degree of matching with exhaustive trend
13-  69069                  - seed number (if automatic)
14-  0 15                   - min and max value of drawing
15-  0                      - output distributions? (0=no;1=yes)
16-  ./smp_dist/           - directory for updated trend distribution
17-  smp_add.dbg           - file for debug
18-  smp_add.out           - file for data output
```

- **Line 9**: Number of the discretization that is applied in both the exhaustive trend distribution and the data distribution. The probability distribution function of variables will be calculated in each discretization point.

- **Line 10**: Number of the classes that is applied in calculating the conditional distribution of the residuals given the trend model.

- **Line 11**: The conditional variance of the residuals given the trend model $\{\varphi(\sigma_{R(\boldsymbol{u}_i)}, m(\boldsymbol{u}_p)), i = 1, \cdots, n \text{ and } p = 1, \cdots, N\}$ in Equation 4.10. There are two options for fitting the conditional variance of the residuals: (1) a constant fitting and (2) a linear fitting.

- **Line 12**: The similarity between the exhaustive trend distribution and the updated trend distribution. The trend continues to be updated until it reaches to the defined similarity.

- **Line 13**: The random seed is provided for drawing the trend values randomly.
- **Line 14**: The lower and upper limits that apply to update data values are given in this line. If the updated trend value is lower/greater than the lower/upper tail, the update trend value will be discarded and draw it again.
- **Line 15-16**: All probability distributions including the updated trend value and the exhaustive trend are calculated by the algorithm. It can be indicated here using the value of 1. If the user is not interested in this information, the value should be set to 0. If the distributions are required to be exported (option=1), the user should provide the path to a existing folder for all outputs. Files with the target true trend will be named as: {path}/trend.out. Files with updated trend distributions will be named as: {path}/distribution_{i}.out where {i} is replaced by the number indicating the $\left\{i^{\text{th}}, i = 0, 1, \cdots\right\}$ updated procedure of the updated trend model. Note that the initial distribution of the trend at the data location is provided with an indicator of $i = 0$.
- **Line 17-18**: The debugging file contains some necessary information (option=1). The information includes the conditional mean, the conditional variance of the residuals and the fitting regression parameters.
- **Line 19**: The output file contains the updated trend and updated data values.