# GEOMETRIC DATA IMPUTATION OF TABULAR VEIN DEPOSITS

by

Jing Bai

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

Master of Science

in

Mining Engineering

Department of Civil & Environmental Engineering
University of Alberta

# Abstract

Multiple imputation is a method to quantify geometric data uncertainty to improve the performance of modeling the geological domain of tabular vein deposits. Having a stable geological domain is the prerequisite of applying other geostatistical methods to estimate mineral resources and reserve. There are many implicit and explicit modeling methods for simple tabular vein deposits, which are single-layered or multi-layered vein deposits with gentle folds. Applying multiple imputation can characterize the uncertainty in the geological domain and the workflow of modeling tabular vein deposits starts from geological data imputation in order to get the stable geological domain.

Geometric data include the position of the hangingwall and the position of the footwall as well as the thickness of the deposit. There is uncertainty in geometric data and this uncertainty increases with highly deviated drill holes. In order to have homotopic geometric data to model the deposits by having one position and one thickness, multiple imputation is used to quantify geometric uncertainty.

During the process of multiple imputation, the primary position distribution from spatial information and the secondary thickness distribution from spatial information are formed. The thickness distribution is transformed into the position. Then, these two distributions are merged together to characterize the local conditional distribution of the position.

The method can be applied to many scenarios. The original application is for single-layered deposits. The method can be extended to multiple-layered deposits. The difference is that there are more thickness values that can be transformed. Fur-

thermore, with assumptions about the distribution of elevation and thickness, the method can be extended to areas with highly deviated long drill holes. The method has an acceptable performance for all three scenarios.

There are many parameters that can influence the performance of the method. The first parameter in the angle tolerance. It is important to have an optimum angle tolerance to save enough samples to conduct distribution transformation, variogram calculation and modeling, as well as form kriging systems, and to eliminate inappropriate calculated thickness values. The imputation sequence should start from the most vertical drill holes to the most inclined drill holes. The merging method is another important parameters. The method of Bayesian updating performs slightly better if there are well known prior distributions. If the scenarios start to become complicated, the method of error ellipses perform better.

There are some limitations of the method. The most important limitation is that during the transformation, there is no guarantee that the imputed values would be in the original range of the geometric variables. Another important limitation is that imputation is mostly based on the geometric samples and the distributions of the imputed results tend to have lower variances than the true distributions of the geometric variables; furthermore, the calculated sample of thickness may not represent the true distribution of thickness.

The workflow of multiple imputation can be applied to the modeling of the geological domain and be integrated into the entire workflow of probabilistic modeling tabular vein deposits.

# Preface

This thesis is an original work by Jing Bai. No part of this thesis has been previously published.

*To Charles and Jimmy*

# Acknowledgements

I would like to express my great gratitude to Professor Clayton V. Deutsch, who is the affable oracle on geostatistics, for his authoritative and affectionate motivation and guidance, which serves as both support and illumination for me during a very hard time. Furthermore, I would like to say thank you to the Centre for Computational Geostatistics (CCG) and all the members at the University of Alberta for providing a wonderful place to study. At the end, I would like to say thank you to Dhaniel, who provides the real data used in the case study.

# Table of Contents

# List of Tables

# List of Figures

# Abbreviations

**BU** Bayesian updating.

**DF** Distance Function.

**DH** Drill Hole.

**EE** Error Ellipses.

**FW** Footwall.

**GMM** Gaussian Mixture Model.

**HW** Hangingwall.

**KDE** Kernel Density Estimation.

**MAR** Missing at Random.

**RMSE** Root Mean Square Error.

**SEDEX** Sedimentary-exhalative.

**SGS** Sequential Gaussian Simulation.

**TLS** Total Least Square.

**Tri** Triangular.

**Uni** Uniform.

# Chapter 1

# Introduction

*A stationary geological domain is a vital prerequisite to applying geostatistical methods. For tabular vein deposits, this geological domain can be defined by the position of the hangingwall and the footwall. Sparse sampling leads to geometric uncertainties in the domain boundaries. This uncertainty increases with highly deviated drill holes that are relative to the plane of continuity of the deposits. Multiple imputation is a method to quantify this uncertainty and to improve the modeling of the geological domain; however, the method is not fully tested and understood for geometric position variables. The method is developed and implemented for different scenarios to evaluate geometric uncertainties of tabular vein deposits.*

## 1.1 Modeling of Tabular Vein Deposits

Geostatistics is applied within clear geological domains. The first step is to create a geological model of the domain boundaries, and the second step is to estimate or simulate grades in the geological model. Practical methods of modeling include explicit modeling and implicit modeling (Cowan et al., 2003; McInerney et al., 2007; Vollgger et al., 2013). Explicit modeling uses manual digitisation to define the surfaces. Implicit modeling uses interpolation of volume functions or indicators to define the surfaces. Implicit modeling does not require manual interpretation and is widely applied because it is automated and thus fast (Cowan et al., 2003; McLennan and Deutsch, 2006). Furthermore, many improvements, like using hydrogeological data (D'Affonseca et al., 2020), training the interpolation functions with machine learning (Ítalo Gomes Gonçalves et al., 2017), and using a linear combination of interpolation functions (Yang et al., 2021), are made in implicit modeling, making implicit modeling a very flexible and popular method.

Explicit and implicit modeling produce a single geometry of the deposits and geometric uncertainties are not assessed. In order to quantify the geometric uncertainties, independent simulation of the hangingwall and the footwall using Sequential Gaussian Simulation (SGS) has been applied (Silva and Deutsch, 2015). Unrealistic simulations can be produced from independently simulated surfaces. In order to prevent unrealistic simulations, manual intervention could be included in the simulation. Developments in geometric modeling of irregular tabular deposits started to involve other improvements, like local coordinate systems (Ostenberg and Deutsch, 2017) and thickness as a geometric variable, to make realistic models (Manchuk and Deutsch, 2015). A recent framework of probabilistic resource modeling for tabular vein deposits was proposed (Carvalho and Deutsch, 2017a,b,c, 2018a,b,c). Simple tabular vein deposits with single or multiple layers and gentle folds are considered (Carvalho, 2018). These deposits are common in mining including certain hydrothermal vein deposits, magmatic deposits, sedimentary-exhalative (SEDEX) lead-zinc deposits, nickel laterite deposits, coal deposits, and uranium deposits (Carvalho, 2018).

The framework consists of quantifying several uncertainties including geometric uncertainty, boundary uncertainty, grade uncertainty, and parameter uncertainty. Several methods were proposed to quantify these uncertainties, as shown in the following list. Implementations of the workflow also include post-processing and sensitivity analysis (Carvalho, 2018).

1. Using data imputation to quantify geometric uncertainty.
2. Using distance function (DF) to quantify boundary uncertainty.
3. Using simulation in an unstructured grid of tetrahedrons to quantify grade uncertainty.
4. Using spatial bootstrap to quantify input parameter uncertainty.

This framework of modeling tabular vein deposits can provide a probabilistic evaluation of the resources and reserves and assess uncertainties by producing realizations of the deposit geometry. The framework is different from traditional deterministic methods and preliminary application has been applied to NexGen Energy Ltd.'s

Arrow Deposit (Batty and Boisvert, 2020a,b,c, 2021a,b). Multiple imputation can improve the performance of modeling of the geological domain.

## 1.2 Geometric Uncertainty

A tabular vein deposit is bounded by a hangingwall and a footwall. The drill holes provide geometric data consisting of the position of these surfaces. The thickness is computed as the difference between the two surfaces measured perpendicular to the plane of greatest continuity. The thickness must be considered in the modeling process in order to prevent the unrealistic crossing of the hangingwall and surfaces footwall (Manchuk and Deutsch, 2015). Then, the geological domain can be modeled by having one thickness of one position from either the hangingwall or the footwall. For example, it is possible to model the hangingwall by adding the thickness to the footwall position. A drill hole usually penetrates both the hangingwall and the footwall, giving two coordinates ($X_{HW}$, $Y_{HW}$, and $Z_{HW}$) and ($X_{FW}$, $Y_{FW}$, and $Z_{FW}$) of the piercing points. From these two coordinates, it is possible to get the thickness between the two surfaces if the drill hole penetrates the deposit perpendicular to the plane of the vein. The positions of the hangingwall and the footwall and thickness provide the geometric values for modeling geometries and surfaces of the deposit.

Geometric uncertainty, one of the uncertainties quantified in the workflow of probabilistic modeling of tabular vein deposits, arises from sparse drilling and increases with inclined drill holes. In Figure 1.1, the configuration of two drill holes is shown in a 2D schematic example. Drill holes do not always penetrate the tabular vein deposit perpendicular to the plane of greatest continuity. For the vertical drill hole 1, both the position of the hangingwall and the footwall are observed and the thickness can be calculated uniquely. The geometry and surfaces can be modeled by having one position and one thickness value. As a result, geometric data imputation is not required. However, for the inclined drill hole 2, the horizontal coordinates are not the same between the two pierce points with the deposit, so the opposing position of the hangingwall and footwall are not known and the thickness cannot be uniquely calculated. The red lines in Figure 1.1 represent one possible geometric configuration of the tabular vein deposit. There are many other possible geometric configurations.

Furthermore, the dashed lines in Figure 1.1 show that because the thickness values are now known, there are various possibilities of the position of the surfaces of the tabular vein deposits for drill hole 2. The modeling method of having one position and one thickness value for each drill hole must be modified in order to model these locations. Geometric data imputation is required for these locations.

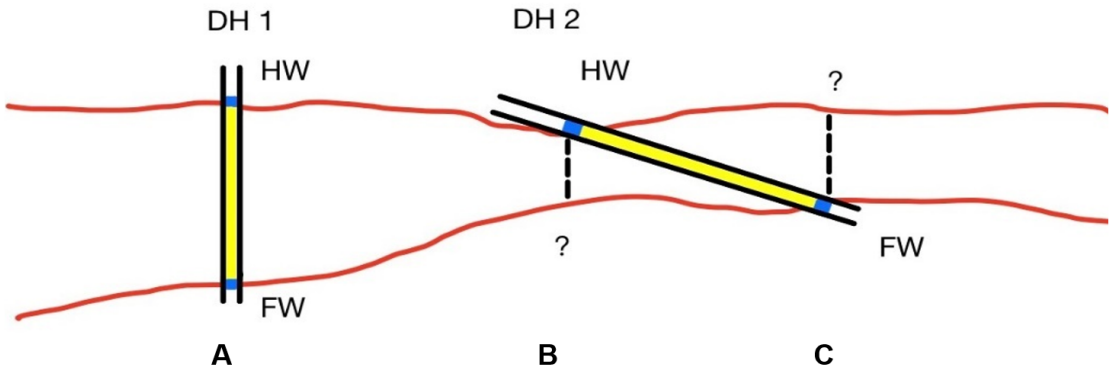



Figure 1.1: Configuration of Drill Holes of the Schematic Example

The schematic example shows the configuration of a tabular vein deposits with two drill holes. The first drill hole is vertical, so geometric data imputation is not required. The second drill hole is inclined, so geometric data imputation is required. (not to scale)

This context of missing thickness because of inclined drill holes motivates the imputation of geometric data at the locations of inclined drill holes to get the full geometric data in order to aid the assessment of geometric uncertainty and other uncertainties as well as the modeling of the geological domain.

## 1.3 Multiple Imputation

Highly inclined drill holes lead to a need for geometric data imputation at pierce points. Data imputation infers missing values at locations with partial data. Missing data are common in a geostatistical dataset. However, many methods, like multivariate transformation, require an equal valued or homotopic dataset (Barnett, 2015). Being homotopic means that each sampled location has all measurable variables. As a result, imputation of the missing data is necessary. There are many data imputation methods (Enders, 2010), like data exclusion, single imputation (mean imputation, regression imputation), and multiple imputation.

The theory of data missing mechanism (Enders, 2010) is considered in order to choose a method to conduct geometric data imputation for tabular vein deposits. The data set is considered as a data matrix $\mathbf{Z}$, and a missing data matrix $\mathbf{M}$ that has the same dimension as $\mathbf{Z}$ can be defined. Each component of $\mathbf{M}$ has values of 1 (observed value) or, 0 (missing value). The data missing mechanism investigates the conditional distribution of $\mathbf{M}$ given $\mathbf{Z}$, $F(\mathbf{M}|\mathbf{Z})$ (Little and Rubin, 2002). The missing matrix $\mathbf{M}$ can be shown in Table 1.1 for the 2D schematic example shown in Figure 1.1.

Table 1.1: Missing Matrix $\mathbf{M}$

| DH | Location | Wall | Horizontal Coordinate | Vertical Coordinate | Thickness |
|---|---|---|---|---|---|
| 1 | A | HW | 1 | 1 | 1 |
| 1 | A | FW | 1 | 1 | 1 |
| 2 | B | HW | 1 | 0 | 0 |
| 2 | B | FW | 0 | 1 | 0 |
| 2 | C | HW | 0 | 1 | 0 |
| 2 | C | FW | 1 | 0 | 0 |

Observed values of $\mathbf{Z}$ can be specified as $\mathbf{Z}_{obs}$, and missing values of $\mathbf{Z}$ can be specified as $\mathbf{Z}_{mis}$. In the case of geometric data imputation, if the observed data $\mathbf{Z}_{obs}$ is from the vertical drill hole 1, the $\mathbf{M}$ will be:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{1.1}$$

and if the observed data $\mathbf{Z}_{obs}$ is from the inclined drill hole 2, the $\mathbf{M}$ will be:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \tag{1.2}$$

As a result, the data missing mechanism is missing at random (MAR) because the conditional distribution of $\mathbf{M}$ are dependent on the value of $\mathbf{Z}_{obs}$ (Barnett, 2015).

$$F(\mathbf{M}|\mathbf{Z},\varphi) = F(\mathbf{M}|\mathbf{Z}_{obs},\varphi)\forall\mathbf{Z}_{mis},\varphi \tag{1.3}$$

where $\varphi$ is an unknown parameter describing the relationship between $\mathbf{M}$ given $\mathbf{Z}$ (Little and Rubin, 2002).

Because of this data missing mechanism, data exclusion and simple imputation (mean imputation, regression imputation) will include biases. As a result, multiple imputation is a preferred method to impute geometric data because this method does not introduce biases (Barnett, 2015). Multiple imputation, which has been adopted into geostatistics recently, can be seen as a derivative of SGS.

The process of multiple imputation is schematically illustrated in Figure 1.2. Multiple imputation starts from a heterotopic dataset and generates 100 to 200 realizations of homotopic dataset realizations based on this heterotopic dataset. The number of realizations match the number of geostatistical realizations that will be simulated in the subsequent workflow. These imputed data realizations are considered as the for each realization data and will be used for the modeling. After the modeling workflow, these geostatistical realizations are used for resources and reserves calculation.

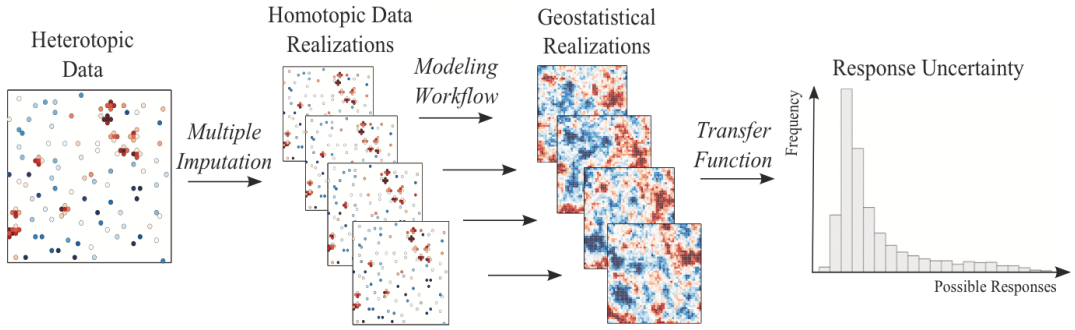



Figure 1.2: Schematic Illustration of Multiple Imputation (Barnett and Deutsch, 2015)

The workflow of multiple imputation is shown. Multiple imputation generates many homotopic data realizations from the heterotopic data. Geostatistical simulations are based on these homotopic data realizations. Transfer function is applied after generating geostatistical simulations.

The method of multiple imputation is modified to suit the problem of imputation of tabular vein geometric data. The position and the thickness of nearby perpendicular drill hole pierce points are used to constrain the unknown geometric data at specific locations having inclined drill holes. The position is used as the primary distribution,

and the thickness is used as the secondary distribution to make realistic models. After the geometric data imputation, the geological domain can be modeled by having one position and one thickness can be applied for the tabular vein deposits.

## 1.4 Problem Statement

Geometric uncertainty exists in geological domain modeling of tabular vein deposits due to sparse sampling and increases in areas with more inclined drill holes. Geometric uncertainty is assessed by geometric data imputation with multiple imputation. The method is not fully implemented and understood in current literature. Furthermore, there are alternative approaches to using multiple imputation to quantify the geometric uncertainties for tabular vein deposits, including single-layered deposits, multiple-layered deposits, and along long drill holes. There is a need to evaluate the performance of multiple imputation and conduct sensitivity analysis on parameters that can influence the performance of multiple imputation. Developing flexible workflows that can correctly conduct multiple imputation allows the method to be implemented in real case studies.

## 1.5 Thesis Outline

Chapter 2 gives the basic method and workflow of multiple imputation for single-layered deposits. This basic method and workflow serves as the foundation of all other chapters. Chapter 2 also demonstrates the method of multiple imputation step by step using a detailed 2D schematic example. Chapter 3 applies the method to 3D single-layered deposits. Examples with true datasets are demonstrated in order to compare the result of multiple imputation. Chapter 3 also discusses the parameters that can influence the performance of the method of multiple imputation as well as some limitations of the method. Chapter 4 extends the basic multiple imputation methods used for single-layered deposits to multi-layered deposits. It demonstrates the method with examples with true datasets and discusses the parameters that can influence the performance of the method. It also concludes with some limitations of the method. Chapter 5 develops multiple imputation to assess geometric uncertainty along long drill holes with both schematic example and examples with true datasets.

Parameters that can influence the performance of the method are also discussed. Chapter 6 illustrates the method with a real case study. The method and the workflow is shown. Chapter 7 discusses conclusions and future work.

# Chapter 2

# Workflow of Multiple Imputation of Geometric Data

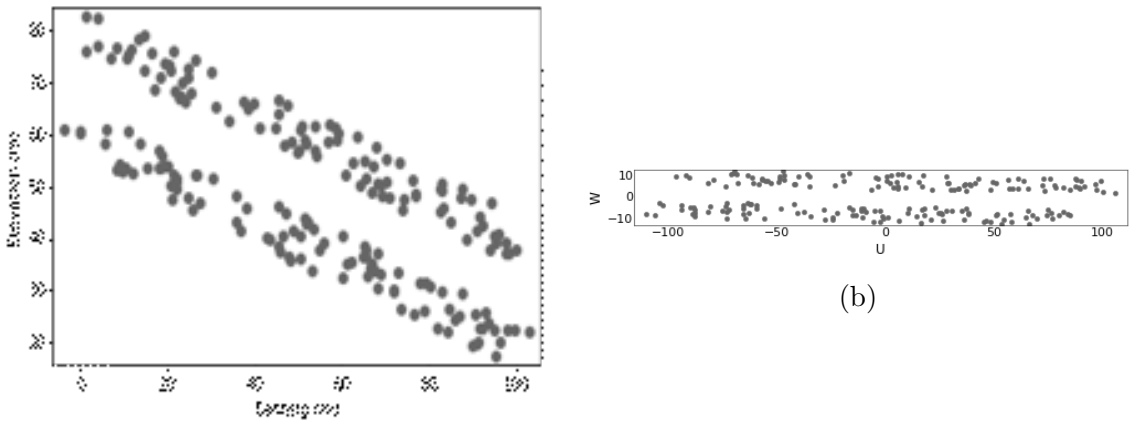
*The workflow of multiple imputation of geometric data serves as the basis for the thesis. The workflow consists of constructing a local coordinates system, angle and thickness calculation, variogram calculation and modeling, simulation of elevation and thickness in a sequential manner, transforming elevation and thickness distribution into the same units, merging elevation and thickness distributions, and sampling from the merged distribution to generate multiple imputation results. There are two merging methods considered: Bayesian updating and error ellipses. The chapter starts with the prerequisites of multiple imputation before proposing the framework of multiple imputation. A 2D schematic example is shown to demonstrate the method. The chapter concludes with a discussion on the two merging methods.*

## 2.1 Local Coordinates System

The scope of multiple imputation is about simple single-layered and multi-layered tabular vein deposits with gentle folds. Some non-linear transformations can be used to constrain the deposits into this scope. These transformations include applying a stratigraphic coordinates system, and unfolding straightening function (Rossi and Deutsch, 2014).

The first prerequisite of multiple imputation is to build a local coordinate system. Tabular vein deposits can be inclined with different dips. Using the original real world coordinate system can be inefficient and may not represent the true anisotropy correctly. Furthermore, it may not be possible to represent the thickness in an arbitrary coordinate system. As a result, a local coordinate system is constructed by Total Least Squares (TLS) (Ostenberg and Deutsch, 2017). The method, which is a

linear regression fitting process, minimizes the error between the intercepts and the fitted plane (Carvalho and Deutsch, 2017a,b; Carvalho, 2018). The result of TLS is transforming the original Cartesian coordinates system with $X$, $Y$, and $Z$ into a local coordinates system of $U$, $V$, and $W$ where these new coordinates are aligned with the dip, strike, and thickness, respectively, of the tabular vein deposit. Figure 2.1 shows the original coordinate system and the local coordinate system transformed after applying TLS. TLS allows an efficient use of the coordinate system and the calculation of thickness values along the plane of continuity. The detailed step by step calculation of TLS (Carvalho and Deutsch, 2018a, 2017b, 2018d; Carvalho, 2018) can be found in these references.




(a) (b)

Figure 2.1: Total Least Square for Single-Layered Deposits

The coordinates of the pierce points of the drill holes with an inclined deposit are shown in (a). After TLS, the coordinates are shown in (b), making the calculation of thickness easier.

The $U$, $V$, and $W$ coordinates are required for geometric data imputation, and the multiple imputation for the elevation ($W$) is executed in the $U$ and $V$ 2D plane for the hangingwall and the footwall. After the transformation, the elevation ($W$) of the hangingwall and or the footwall, and the thickness ($TH$), as a function of the $U$ and $V$ coordinates, are variables in a 2D space.

## 2.2 Other Prerequisites

After TLS, the local coordinate system allows the thickness to be calculated. There are other prerequisites for multiple imputation, including angle and thickness cal-

culation, normal score transform and variogram calculation and modeling. These prerequisites enable the application of multiple imputation. Angle and thickness calculation decides the locations that are necessary to be imputed and construct the thickness distribution, and normal score transform and variogram calculation and modeling allows the implementation of SGS.

### 2.2.1 Angle and Thickness Calculation

The geometric uncertainty increases with highly deviated drill holes. As a result, the calculated angle for each drill hole is used to determine whether it is necessary to conduct imputation to quantify the geometric uncertainty. The drill hole angle is calculated by the following equation:

$$\theta = cos^{-1}\left(\frac{W_{HW} - W_{FW}}{\sqrt{(U_{HW} - U_{FW})^2 + (V_{HW} - V_{FW})^2 + (W_{HW} - W_{FW})^2}}\right) \tag{2.1}$$

where $\theta$ is the angle between the drill hole and the tabular vein deposit from 0 degree (perpendicular to the reference plane) to 90 degrees (parallel to reference plane), $U_{HW}$, $U_{FW}$, $V_{HW}$, $V_{FW}$, $W_{HW}$, and $W_{FW}$ are the $U$, $V$, and $W$ coordinates of the pierce points of the hangingwall and the footwall.

It is important to include the thickness of the tabular vein deposit in the model in order to create realistic surfaces. All thickness values must be greater than 0, otherwise the hangingwall and the footwall surfaces will cross. After the calculation of the angles, a maximum possible angle tolerance will decide whether the thickness values can be calculated or should be imputed. If the angle of the drill hole is smaller than angle tolerance, the drill hole will be considered as perpendicular to the plane of continuity and the drill hole can provide information about the hangingwall and the footwall surfaces and a thickness value. If the angle of the drill hole is larger than the angle tolerance, the drill hole is consider as inclined and the thickness will be imputed. The geometric data derived from the drill holes having small angles that do not require imputation can be used to form local thickness values and the global thickness distribution. The geometric data including the elevation and the thickness of other drill holes will be imputed.

After the TLS transformation, the elevation ($W$) and the thickness ($TH$) are a function of $U$ and $V$ coordinates. As a result, the thickness can be calculated by the following equation, and is added into the raw data:

$$TH = W_{HW} - W_{FW} \tag{2.2}$$

where $TH$ is the calculated thickness value, $W_{HW}$ is the elevation of the hangingwall, and $W_{FW}$ is the elevation of the footwall. The calculated thickness approximates the true thickness if the drill hole is perpendicular to the hangingwall and the footwall.

After the calculation of the angles and the thickness, the geometric variables including the elevation and the thickness used for multiple imputation can be normal score transformed.

### 2.2.2 Normal Score Transform

It is necessary to normal score transform the three geometric variables, $W_{HW}$, $W_{FW}$, and $TH$, because multiple imputation is conducted in Gaussian units. The transformation tables should be saved for further processing and simulation. The three geometric variables should be transformed independently by the following equations:

$$y_{W_{HW}} = G^{-1}\left(F_{W_{HW}}\left(W_{HW}\right)\right) \tag{2.3}$$

$$y_{W_{FW}} = G^{-1}\left(F_{W_{FW}}\left(W_{FW}\right)\right) \tag{2.4}$$

$$y_{TH} = G^{-1}\left(F_{TH}\left(TH\right)\right) \tag{2.5}$$

where $F_{W_{HW}}$, $F_{W_{FW}}$, and $F_{W_{TH}}$ are the cumulative distribution functions, $G^{-1}$ is the quantile function of the standard univariate Gaussian distribution. $W_{HW}$, $W_{FW}$, and $TH$ are variables in their original units, and $y_{W_{HW}}$, $y_{W_{FW}}$, and $y_{TH}$ are the values in the standard Gaussian distribution, respectively. Manual interpretation after declustering can derive the cumulative distribution functions. $y_{W_{HW}}$, $y_{W_{FW}}$, and $y_{TH}$ are used to form the kriging system and the distribution for imputation.

### 2.2.3 Variograms Calculation and Modeling

Variograms of the normal score variables are required to infer the covariances used to calculate conditional distributions at unsampled locations. $y_{W_{HW}}$, $y_{W_{FW}}$, and $y_{TH}$ data are used to calculate the variograms. Variograms of the elevation of the hangingwall and the footwall as well as the thickness are calculated in the $U$ and $V$ 2D plane. If there are insufficient perpendicular data to calculate and model a thickness variogram, then the variogram of the thickness can be set to either the variogram of the elevation of the hangingwall or the variogram of the elevation of the footwall. It is recommended to use a Gaussian variogram model and a small nugget effect to model the variogram because the elevation ($W$) and thickness ($TH$) are expected to be quite continuous and the short scale variability is expected to be quite low (Isaaks and Srivastava, 1989).

## 2.3 Multiple Imputation

After applying the prerequisites of building a local coordinate system, calculation of the angle and the thickness, normal score transformation, and variogram calculation and modeling, it is possible to conduct multiple imputation. Multiple imputation consists of forming local distributions using kriging, transforming distributions of thickness into elevation, merging these two distributions and drawing from the merged distribution.

Multiple imputation considers the elevation ($W$) as the primary distribution and the thickness ($TH$) as the secondary distribution. The thickness ($TH$) distribution is transformed into the elevation distribution ($W$) by considering the opposing surface position, allowing the merging of these two distributions to constrain the unknown pierce points' locations.

### 2.3.1 Local Distributions by Kriging

Multiple imputation uses SGS to form the distribution of the elevation of the hangingwall ($W_{HW}$), footwall ($W_{FW}$) and the thickness ($TH$), and samples from the merged distribution. It is assumed that the local distribution can be fully characterized by

the conditional mean and conditional variance in Gaussian units and can be defined under a multivariate Gaussian distribution.

The distribution of the elevation ($W$), which is considered as the primary data, is characterized by the conditioned mean $y_{\bar{W}}(\mathbf{u})$, and the conditioned variance $\sigma^2_{y_W}(\mathbf{u})$. Simple kriging is applied here to incorporate the spatially correlated data, which can be calculated by the following equations (Barnett, 2015):

$$y_{\bar{W}}(\mathbf{u}) = \sum_{\alpha=1}^{n} \lambda_\alpha y_W(\mathbf{u}_\alpha) \tag{2.6}$$

$$\sigma^2_{y_W}(\mathbf{u}) = 1 - \sum_{\alpha=1}^{n} \lambda_\alpha C_{y_W}(\mathbf{u}, \mathbf{u}_\alpha) \tag{2.7}$$

where $C_{y_W}(\mathbf{u}, \mathbf{u}_\alpha)$ is the covariance of the elevation values between the missing data value at $\mathbf{u}$ and the observed data values at $\mathbf{u}_\alpha$. The covariance comes directly from the variogram model $(1-\gamma)$. The weights, $\lambda_\alpha$, are calculated by the normal equations:

$$\sum_{\beta=1}^{n} \lambda_\beta C_{y_W}(\mathbf{u}_\alpha, \mathbf{u}_\beta) = C_{y_W}(\mathbf{u}, \mathbf{u}_\alpha) \qquad \alpha = 1, \cdots, n \tag{2.8}$$

The distribution of the thickness ($TH$), which is considered as the secondary data, is characterized by the conditioned mean $y_{\bar{TH}}(u)$, and the conditioned variance $\sigma^2_{y_{TH}}(u)$. Simple kriging is applied as well, which is shown in the following equations:

$$y_{\bar{TH}}(\mathbf{u}) = \sum_{\alpha=1}^{n} \lambda_\alpha y_{TH}(\mathbf{u}_\alpha) \tag{2.9}$$

$$\sigma^2_{y_{TH}}(\mathbf{u}) = 1 - \sum_{\alpha=1}^{n} \lambda_\alpha C_{y_{TH}}(\mathbf{u}, \mathbf{u}_\alpha) \tag{2.10}$$

where $C_{y_{TH}}(\mathbf{u}, \mathbf{u}_\alpha)$ is the covariance of the thickness between the missing data value at $\mathbf{u}$ and the observed data value at $\mathbf{u}_\alpha$. The weights, $\lambda_\alpha$, are calculated by the normal equations and are different from the weights calculated above in Equation 2.8:

$$\sum_{\beta=1}^{n} \lambda_\beta C_{y_{TH}}(\mathbf{u}_\alpha, \mathbf{u}_\beta) = C_{y_{TH}}(\mathbf{u}, \mathbf{u}_\alpha) \qquad \alpha = 1, \cdots, n \tag{2.11}$$

The number of data ($n$) used in the kriging system in the same search radius for the elevation ($W$) and the thickness ($TH$) are likely different because one vertical drill hole can provide both the elevation and the thickness, while one inclined drill hole can provide one elevation of the hangingwall and one elevation of the footwall, but no thickness value can be provided. As a result, the data number of elevation ($n_W$) is always larger than or equal to the data number of thickness ($n_{TH}$).

The two local distributions are both formed by simple kriging using surrounding data, which is different from the original multiple imputation method. The original method uses the covariance between the primary and many collocated secondary variables to form the local distribution and it is similar to linear regression which gives a distribution of the primary variable. The calculations are shown in the following equations (Barnett, 2015; Barnett and Deutsch, 2015):

$$\bar{y}(\mathbf{u}) = \sum_{i=1}^{K} \lambda_i y_i(\mathbf{u}) \tag{2.12}$$

$$\sigma^2(\mathbf{u}) = 1 - \sum_{i=1}^{K} \lambda_i C_{p,i} \tag{2.13}$$

where $K$ is the number of secondary data, and $C_{p,i}$ is the covariance between the primary variable and the $i^{th}$ secondary variable. The weights, $\lambda_i$, are calculated by the following equation:

$$\sum_{j=1}^{K} \lambda_j C_{i,j} = C_{p,i} \qquad i = 1, \cdots, K \tag{2.14}$$

where $C_{i,j}$ is the covariance between the two secondary variables.

In geometric data imputation, the thickness, as the secondary data, is used as a reference to convert the uncertainty to elevation uncertainty. The thickness distribution is fully determined by kriging Equations 2.9, 2.10, and 2.11 and thickness values are not used to form a distribution of elevation directly by regression Equations 2.12, 2.13, and 2.14. Consequently, there are two different distributions, the distribution of elevation and the distribution of thickness. So, it is necessary to transform the distribution of thickness into the distribution of elevation.

### 2.3.2 Transforming Thickness to Elevation

The distributions of elevation and thickness are not in the same units. As a result, it is necessary to transform the distribution of the thickness before merging them together (Batty and Boisvert, 2020c). Because the elevation is the final imputed value, the thickness distribution is transformed into the elevation distribution by the following procedure:

1. Generate many random $y_{TH}$ values from the thickness distribution characterized by the local mean $y_{\bar{TH}}(\mathbf{u})$ and local variance $\sigma^2_{y_{TH}}(\mathbf{u})$.

2. Back transform the thickness values $y_{TH}$ into original units by the following equation:

$$TH_l = {F_{TH}}^{-1}(G(y_{TH,l})) \qquad l = 1, \cdots, L \tag{2.15}$$

   where $G$ is the cumulative distribution function of the standard univariate Gaussian distribution, ${F_{TH}}^{-1}$ is the quantile function of the thickness, and $TH$ is the thickness in the original unit.

3. Calculate the opposite elevation using the back transformed thickness value in the original unit and the known elevation value. For an inclined drill holes, there are four geometric data, including two hangingwall and two footwall values, and one hangingwall and one footwall are known and the other two are not known. If the elevation of the hangingwall is being imputed, get the transformed distribution of the hangingwall by adding the thickness and the footwall shown in the following equation:

$$W_{HW} = W_{FW} + TH \tag{2.16}$$

   and if the elevation of the footwall is being imputed, get the transformed distribution of the footwall by subtracting the thickness from the hangingwall shown in the following equation:

$$W_{FW} = W_{HW} - TH \tag{2.17}$$

where $W_{HW}$ is the elevation of the hangingwall and $W_{FW}$ is the elevation of the footwall.

4. Normal score transform the calculated elevation values using the corresponding distribution. If the elevation of the hangingwall is being imputed, use equation 2.3 and if the elevation of the footwall is being imputed, use equation 2.4.

5. Calculate the mean and the variance of the transformed elevation distribution. The transformed local distribution could be assumed as Gaussian characterized by the mean and the variance.

The generation of random thickness values in step 1 follows the following steps. $L$ is usually set to 100 and $l$ is randomly generated. $y_{TH}$ is used to back transform the thickness values into original units in Step 2.

$$P_l = \frac{l}{L+1} \qquad l = 1, \cdots, L \tag{2.18}$$

$$y_{TH,l} = \bar{y_{TH}}(\mathbf{u}) + \sigma_{y_{TH}}(\mathbf{u}) G^{-1}(P_l) \qquad l = 1, \cdots, L \tag{2.19}$$

After the transformation, the primary distribution of the elevation is not changed and the secondary distribution of the thickness is changed into the distribution of elevation unit.

### 2.3.3 Merging Two Distributions

After the transformation of the thickness distribution to elevation, the distributions can be merged into one distribution.

There are two methods applied in geostatistics to merge compatible conditional distributions. Bayesian updating is a widely used method and should be considered for geometric data imputation (Batty and Boisvert, 2020c). The method can be summarized by the following equations (Zhang et al., 2020):

$$y_{\bar{W(m)}BU} = \frac{\sigma^2_{y_{W(p)}} y_{\bar{W}(t)} + \sigma^2_{y_{W(t)}} y_{\bar{W}(p)}}{\sigma^2_{y_{W(p)}} - \sigma^2_{y_{W(p)}} \sigma^2_{y_{W(t)}} + \sigma^2_{y_{W(t)}}} \tag{2.20}$$

$$\sigma_{y_{W(m)^2_{BU}}} = \frac{\sigma^2_{y_{W(p)}}\sigma^2_{y_{W(t)}}}{\sigma^2_{y_{W(p)}} - \sigma^2_{y_{W(p)}}\sigma^2_{y_{W(t)}} + \sigma^2_{y_{W(t)}}} \tag{2.21}$$

The error ellipses method is the proposed method for the original workflow to quantify geometric uncertainties (Carvalho, 2018). The method can be summarized by the following equations (Erten and Deutsch, 2020):

$$y_{\bar{W(m)}_{EE}} = \sigma^2_{y_{W(m)}} \left( \frac{y_{\bar{W}(p)}}{\sigma^2_{y_{W(p)}}} + \frac{y_{\bar{W}(t)}}{\sigma^2_{y_{W(t)}}} \right) \tag{2.22}$$

$$\sigma_{y_{W(m)^2_{EE}}} = \left( \frac{1}{\sigma^2_{y_{W(p)}}} + \frac{1}{\sigma^2_{y_{W(t)}}} \right)^{-1} \tag{2.23}$$

where $y_{\bar{W(m)}}$ is the merged mean and $\sigma^2_{y_{W(m)}}$ is the merged variance and $y_{\bar{W}(p)}$, $\sigma^2_{y_{W(p)}}$ is the primary mean and variance of the elevation distribution, $y_{\bar{W}(t)}$, $\sigma^2_{y_{W(t)}}$ is the transformed mean and variance of the elevation distribution from the thickness distribution, respectively.

Both of the methods can combine probability distributions but there are some differences. Bayesian updating assumes a known prior distribution (Zhang et al., 2020), and the method of error ellipses assumes a infinite variance prior (Erten and Deutsch, 2020). The merging method is one of the determining factors of the performance of multiple imputation and will be investigated in the following chapters.

At the end, it is possible to draw a value from the merged elevation distribution to form one realization of the missing geometric data at the pierce point. Once both the elevation of the hangingwall and the elevation of the footwall are known, the thickness can be uniquely calculated. These imputed values can be added to the dataset and are used in the future as known data when forming new elevation and thickness distributions. The imputation can proceed to the next location in an organized sequence which will be discussed in the following chapters. After all missing data locations are imputed, a homotopic imputation geometric data file will be generated.

## 2.4 Surface and Solid Modeling

The number of homotopic geometric data files generated by multiple imputation matches the number of realizations required by the following simulation and transfer functions. Surface and solid modeling can proceed using the generated imputed geometric data. There will be many different surface and solid models at the end of surface simulation. The geometric uncertainty can be quantified from these simulated geological domains. Other sources of uncertainty, like boundary uncertainty and grade and tonnage uncertainty, can be quantified using these generated solids (Carvalho and Deutsch, 2017a,b,c; Carvalho, 2018; Carvalho and Deutsch, 2018d).

## 2.5 Schematic Example

A 2D schematic example is shown in this section to demonstrate the method (Bai and Deutsch, 2020b; Batty and Boisvert, 2020b). The simplified prerequisites of multiple imputation are given and the detailed workflow of multiple imputation is illustrated. Imputation results and local distributions after multiple imputation are shown and the performance of merging method is discussed at the end.

### 2.5.1 Data and Prerequisites of the Schematic Example

The figure in Section 1.2 is extended to create this 2D schematic example. The merged 1D $U$ and $V$ coordinates of the pierce points of the deposit and drill holes are added to the original Figure 1.1. The elevation ($W$) is also added to the figure. The configuration of the drill holes of the 2D schematic example with coordinates is shown in Figure 2.2.

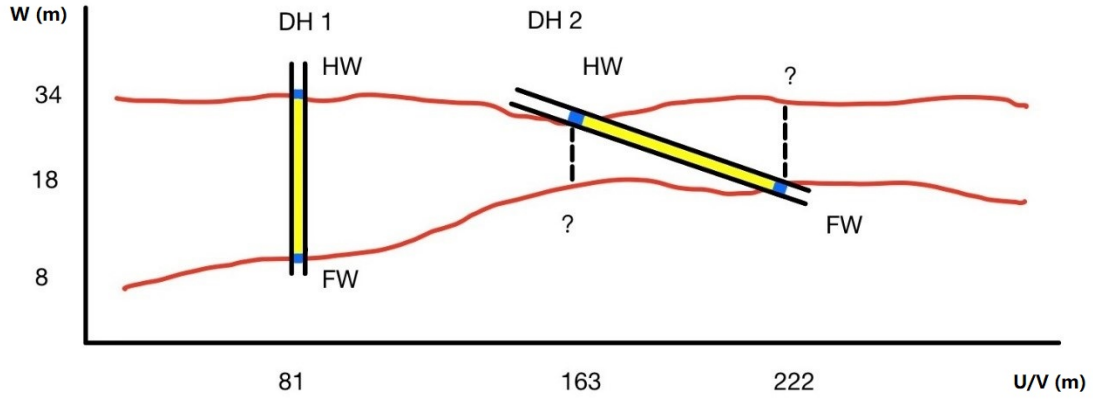



Figure 2.2: Configuration of Drill Holes for the Schematic Example

1D schematic example with the coordinates $U/V$ and $W$ of the pierce points between the drill holes and the deposits is shown. (not to scale)

The geometric data shown in Figure 2.2, including drill hole (DH), wall type (Wall) and the coordinates of the pierce points can be summarized into Table 2.1. For vertical drill hole 1, both the elevation of the hangingwall and the elevation of the footwall are known; for inclined drill hole 2, one elevation of the hangingwall and one elevation of the footwall are not known.

Table 2.1: Geometric Data of the Schematic Example

| DH | Wall | U/V | W |
|---|---|---|---|
| 1 | HW | 81 | 34 |
| 1 | FW | 81 | 34 |
| 2 | HW | 163 | 31 |
| 2 | FW | 222 | 18 |

The thickness can be calculated for drill hole 1 and the thickness cannot be calculated for drill hole 2. The calculated value of the thickness can be added into the raw geometric data set in order to form the data set for imputation, which is summarized in Table 2.2. The question marks in Figure 2.2 represent the pierce points that are not known and need to be imputed.

Table 2.2: Geometric Data with Thickness of the Schematic Example

| DH | Wall | U/V | Hangingwall | Footwall | Thickness |
|---|---|---|---|---|---|
| 1 | HW | 81 | 34 | 8 | 34 - 8 = 26 |
| 2 | HW | 163 | 31 | ? | ? |
| 2 | FW | 222 | ? | 18 | ? |

For the schematic example, the distribution of the three geometric variables and the variograms are known. It is assumed that the elevation of the hangingwall follows a triangular distribution of Tri $(20, 30, 40)$, the elevation of the footwall follows a triangular distribution of Tri $(0, 10, 20)$, and the thickness follows a uniform distribution of Uni $(5, 30)$. For a random variable $X$ following a triangular distribution $X \sim$ Tri $(a, c, b)$, $a$ is the lower limit, $b$ is the upper limit, and $c$ is the mode. For a random variable $X$ following a uniform distribution $X \sim$ Uni $(a, b)$, $a$ is the lower limit, and $b$ is the upper limit. All these three variables can be normal score transformed based on the distribution functions. The transformed data values are added in Table 2.3.

Table 2.3: Geometric Data with NS Values of the Schematic Example

| DH | Wall | X | $W_{HW}$ | $W_{FW}$ | $TH$ | $y_{W_{HW}}$ | $y_{W_{FW}}$ | $y_{TH}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | HW | 81 | 34 | 8 | 26 | 0.9154 | -0.4677 | 0.9945 |
| 2 | HW | 163 | 31 | ? | ? | 0.2404 | ? | ? |
| 2 | FW | 222 | ? | 18 | ? | ? | 2.0537 | ? |

It is also assumed that the variograms of the elevation of the hangingwall, the elevation of the footwall, and the thickness can be all modeled as Gaussian with a nugget effect of 0.01 and a range of 200. The following equation shows this variogram model.

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ 0.01 + 0.99 \times \left(1 - exp\left(-3 \times \frac{h^2}{200^2}\right)\right) & h \neq 0 \end{cases} \tag{2.24}$$

where $h$ is the 1D distance between data points. The covariance values used in the

kriging system come from the variogram model $(1-\gamma(h))$.

After these prerequisites of thickness calculation, normal score transform, and variogram modeling, distributions can be derived for multiple imputation.

### 2.5.2 Multiple Imputation of Schematic Example

It is possible to form the distributions using the geometric data shown in Table 2.3 and the variogram shown in equation 2.24 and it is assumed that the sequence of imputation starts from the footwall at 163 and goes to the hangingwall at 222.

The distribution of elevation of the footwall can be formed by kriging using the data at 81 and 222. The variogram and covariance based on the distance and the variogram model can be calculated. The kriging equations can be constructed and solved. The mean and variance of the local distribution can be derived.

$$\begin{bmatrix} 1 & 1-\gamma(222-81) \\ 1-\gamma(222-81) & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 1-\gamma(163-81) \\ 1-\gamma(222-163) \end{bmatrix} \tag{2.25}$$

$$\mathbf{C} = \begin{bmatrix} 1 & 0.2229 \\ 0.2229 & 1 \end{bmatrix} \mathbf{D} = \begin{bmatrix} 0.5979 \\ 0.7625 \end{bmatrix} \tag{2.26}$$

$$\lambda = \mathbf{C}^{-1}\mathbf{D} = \begin{bmatrix} 0.4503 \\ 0.6621 \end{bmatrix} \tag{2.27}$$

$$y_{\bar{W(p)}} = 0.4503 \times -0.4677 + 0.6621 \times 2.0537 = 1.1493 \tag{2.28}$$

$$\sigma^2_{y_{W(p)}} = 1 - 0.4503 \times 0.5979 - 0.6621 \times 0.7625 = 0.2259 \tag{2.29}$$

So, the mean and variance of the primary distribution of the elevation of the footwall is 1.1493 and 0.2259. The distribution of the thickness can be formed by kriging using the data at 81.

$$\begin{bmatrix} 1 \end{bmatrix} \begin{bmatrix} \lambda_1 \end{bmatrix} = \begin{bmatrix} 1-\gamma(163-81) \end{bmatrix} \tag{2.30}$$

$$\mathbf{C} = \begin{bmatrix} 1 \end{bmatrix} \mathbf{D} = \begin{bmatrix} 0.5979 \end{bmatrix} \tag{2.31}$$

$$\lambda = \mathbf{C}^{-1}\mathbf{D} = \begin{bmatrix} 0.5979 \end{bmatrix} \tag{2.32}$$

$$\bar{y_{TH(s)}} = 0.5979 \times 0.9945 = 0.5946 \tag{2.33}$$

$$\sigma^2_{y_{TH(s)}} = 1 - 0.5979 \times 0.5979 = 0.6425 \tag{2.34}$$

So, the mean and variance of the secondary distribution of the thickness of the footwall is 0.5946 and 0.6425. The thickness distribution can be transformed into the elevation of the footwall at 163.

1. Generate random values and corresponding quantiles of the distribution characterized by (0.5946, 0.6425).

$$\begin{bmatrix} -0.9002 & -0.7250 & \cdots & 0.5785 & 0.5946 & 0.6107 & \cdots & 1.8030 & 1.9142 \end{bmatrix}$$

2. Transform the quantile values to the original unit of the thickness.

$$TH = F^{-1}_{TH}\left(G\left(y_{TH}\right)\right)$$

$$\begin{bmatrix} 9.6004 & 10.956 & \cdots & 22.963 & 23.098 & 23.232 & \cdots & 29.108 & 29.305 \end{bmatrix}$$

3. Calculate the footwall based on the thickness. It is possible to find that the hangingwall is 31 at 163. Because the elevation of the footwall follow a triangular distribution Tri $\sim (0, 10, 20)$, some values that are higher than 20 are labeled as -999.

$$W_{FW} = W_{HW} - TH$$

$$\begin{bmatrix} -999 & -999 & \cdots & 8.0368 & 7.0916 & 7.7676 & \cdots & 1.8923 & 1.6950 \end{bmatrix}$$

4. Normal score transform the elevation of the footwall.

$$y_{W_{FW}} = G^{-1}\left(F_{W_{FW}}\left(W_{FW}\right)\right)$$

$$\begin{bmatrix} -999 & -999 & \cdots & -0.4594 & -0.4897 & -0.5196 & \cdots & -2.0991 & -2.1872 \end{bmatrix}$$

5. Calculate the mean and the variance of the new distribution, which is -0.3502 and 1.1746.

After the transformation, the values of $y_{W_{FW}^{-}(p)}$, $\sigma^2_{y_{W_{FW}(p)}}$, $y_{W_{FW}^{-}(t)}$, and $\sigma^2_{y_{W_{FW}(t)}}$ are known. The merged distribution can be calculated by the following equation, assuming the method of error ellipses is used.

$$y_{\bar{W(m)}} = \sigma^2_{y_{W(m)}} \left( \frac{1.1493}{0.2259} + \frac{-0.3502}{1.1746} \right) \tag{2.35}$$

$$\sigma^2_{y_{W(m)}} = \frac{1}{\frac{1}{0.2259} + \frac{1}{1.1746}} \tag{2.36}$$

The final merged distribution is characterized by (1.2701, 0.2653) and a value of the elevation of the footwall can be drawn from this distribution. The thickness can be calculated using the known hangingwall value. The dataset with the imputed result is shown in Table 2.4.

Table 2.4: Imputation Result at 163 of the Schematic Example

| DH | Wall | Horizontal Coordinate | Hangingwall | Footwall | Thickness |
|---|---|---|---|---|---|
| 1 | HW | 81 | 34 | 8 | 26 |
| 2 | HW | 163 | 31 | 16.341 | 14.659 |
| 2 | FW | 222 | ? | 18 | ? |

After imputing the elevation of the footwall at 163, the imputation can proceed to the hangingwall at 222. The previously imputed value at 163 is used for future imputation. The distribution of the elevation of the hangingwall can be formed by kriging using the data at 81 and 163 and the thickness distribution can be formed by kriging using the data at 81 and 163. The thickness value of 14.659 is imputed before. The thickness distribution can be transformed into the elevation of the hangingwall at 222 by adding 18. The final distribution can be formed by merging these two distributions using the same method as illustrated above. The imputed dataset is shown in Table 2.5.

The drill hole plot is shown in black in Figure 2.3 at the top and one realization of the imputed result of multiple imputation is shown in blue in Figure 2.3. The

Table 2.5: Imputed Result of the Schematic Example

| DH | Wall | Horizontal Coordinate | Hangingwall | Footwall | Thickness |
|---|---|---|---|---|---|
| 1 | HW | 81 | 34 | 8 | 26 |
| 2 | HW | 163 | 31 | 16.341 | 14.659 |
| 2 | FW | 222 | 29.078 | 18 | 11.078 |

elevation of the footwall at 163 is 16.341 and the elevation of the hangingwall at 222 is 29.078.

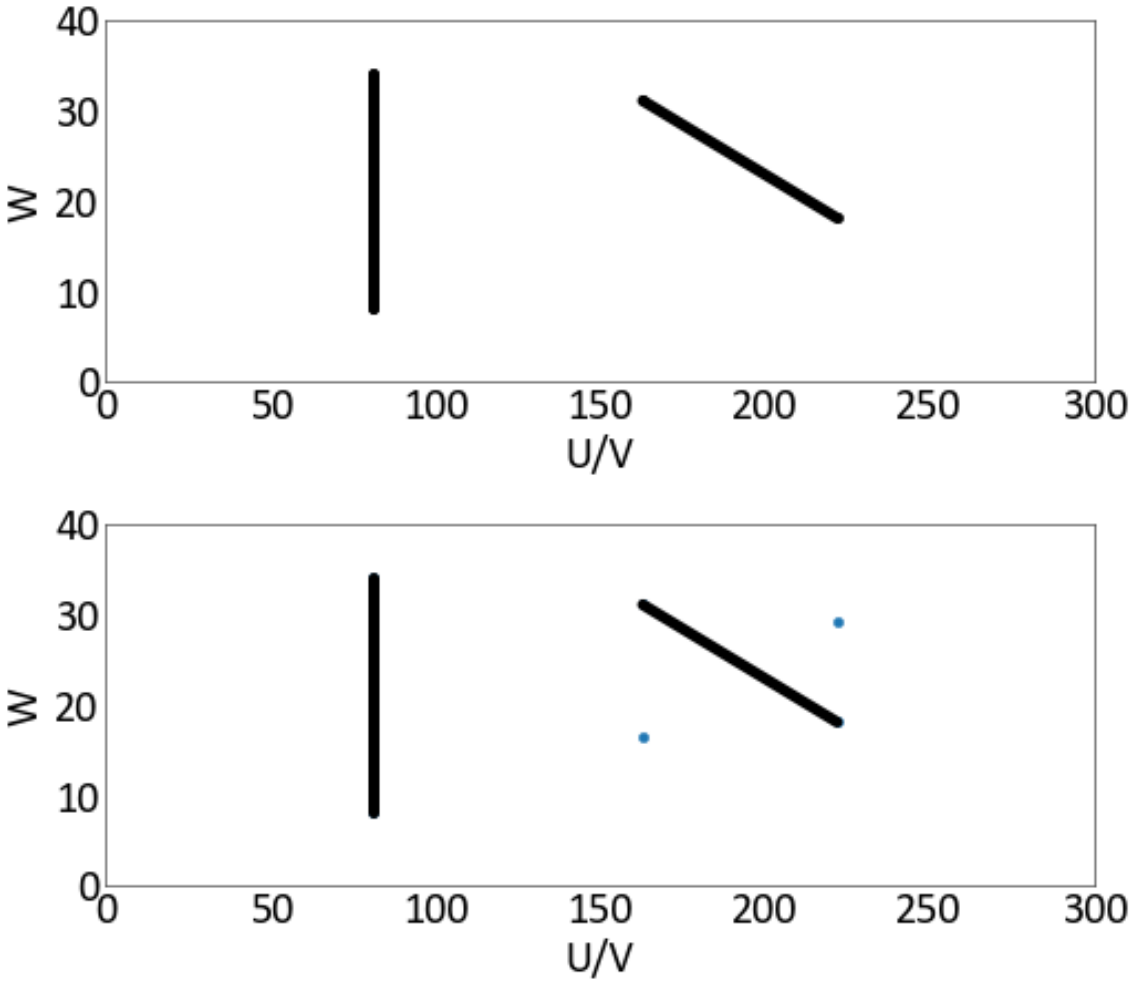



Figure 2.3: Drill Plot and One Imputation of the Schematic Example

The geometric data with two drill holes before imputation are shown at the top. One realization of the imputed result is shown at the bottom.

### 2.5.3 Result of Schematic Example

The imputation can be repeated many times in order to match the final number required by future geostatistical simulations. The imputed values with 200 realizations are shown in Figure 2.4. The two bolded blue areas show the distribution of the elevation and it is possible to assess the geometric uncertainties at these two locations.

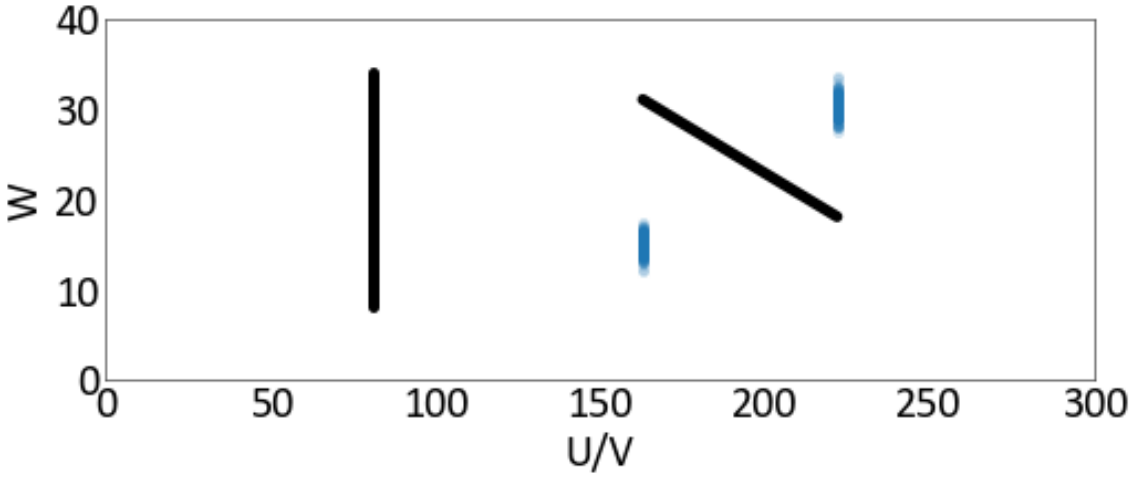
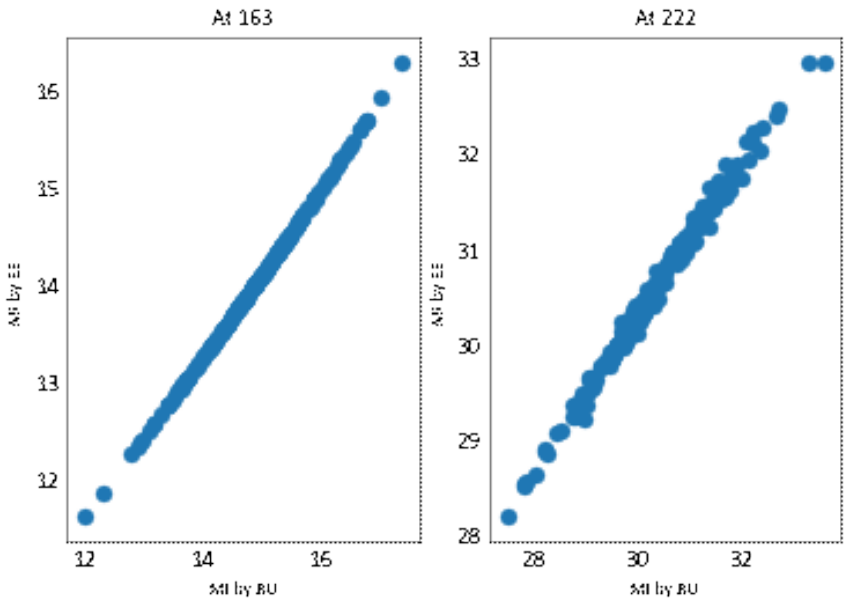



Figure 2.4: 200 Imputed Realizations of the Schematic Example

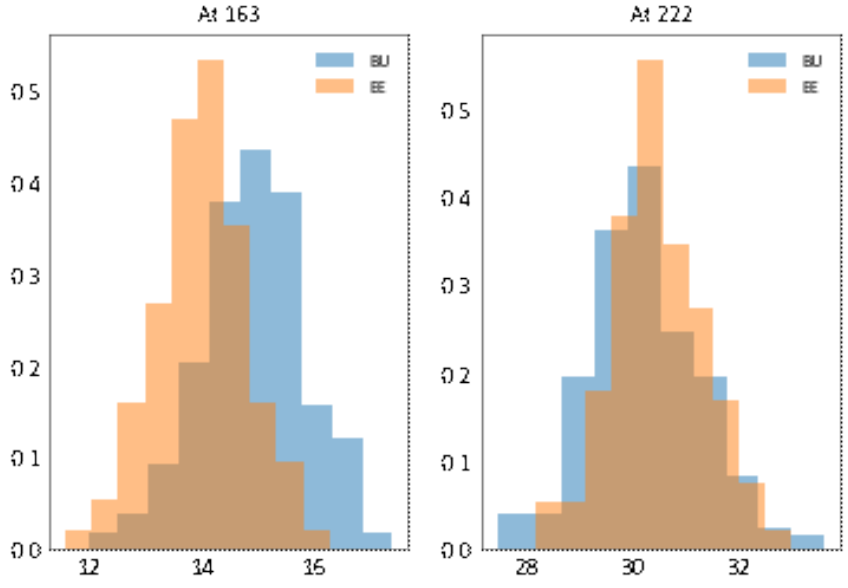
200 imputed realizations show the possible distribution of the elevation of the hangingwall and the footwall.

A simple comparison between the method of Bayesian updating and error ellipses is presented here. The method of error ellipses is chosen as the method in the original workflow because it always gives a distribution in the middle of two merging distributions (Carvalho, 2018). However, the method of Bayesian updating is also a robust and simple method widely used, which should be tried in multiple imputation. In Figure 2.5, the distributions formed by Bayesian updating and error ellipses are shown. The means are similar, and the variances of Bayesian updating shown in orange are slightly larger than the ones of error ellipses shown in blue. A more complete comparison will be shown in the next chapter.

Building a local coordinates system, angle and thickness calculation, normal score transform, and variogram calculation and modeling serve as the prerequisites of multiple imputation. Multiple imputation consists of forming distributions of elevation and thickness, transforming thickness distribution to elevation, merging elevation dis-




(a)




(b)

Figure 2.5: Bayesian Updating and Error Ellipses Results of the Schematic Example

In (a), the results of the Bayesian updating and error ellipses are plotted in a scatter plot. The results are similar. In (b), the local distribution of the imputed values are shown. The results by the two methods are still similar. (BU for Bayesian updating and EE for error ellipses)

tribution and thickness distribution. This workflow of multiple imputation serves as the basis to assess geometric uncertainty and many applications will be demonstrated in the following chapters.

# Chapter 3

# Single-Layered Deposits

*The application of multiple imputation for geometric data imputation in single-layered deposits is developed. This chapter starts with the demonstration of the workflow with two synthetic examples. The distributions and the variograms of the variables in the first example are assumed known, while this assumption is relaxed in the second example. The imputed results are compared with the true values in order to validate the method. Through a sensitivity analysis investigating various influencing parameters that can affect the performance of the method, it is found that choosing an appropriate angle tolerance and transformation method is important and the best sequence is to start from the most vertical drill holes. Also, error ellipses performs slightly better than Bayesian updating. This chapter concludes with some limitations that include that the calculated thickness distribution may not reflect the entire distribution and there is no guarantee that the imputed values will be in the range of the distribution.*

## 3.1 Demonstration of Single-Layered Deposits

Two examples demonstrate the major steps and the results of multiple imputation for single-layered deposits. One full synthetic true dataset is generated for both examples. The same workflow is applied to the two examples, but the transformation methods and variograms are different. It is assumed that the true distributions and variograms of all the geometric variables in the first example are known, while in the second example, the distributions and the variograms of all the geometric variables are not known and are derived in a non-parametric way from the drill hole samples. The validation of the method of multiple imputation is achieved through comparing the true synthetic values with the imputed values. Several statistics, including error rate, Pearson correlation coefficient ($\rho$), and root mean square error (RMSE) are

calculated. Cumulative distributions and variograms are accessed to compare global distribution and spatial continuity. A comparison of the results from the two examples concludes the section.

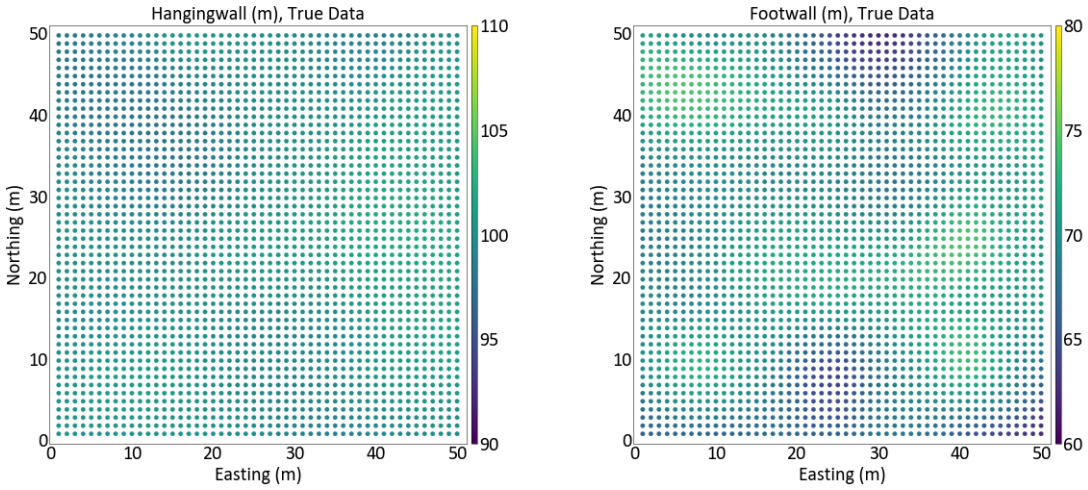
### 3.1.1 True Data for Demonstration of Single-Layered Deposits

A full geometric dataset, including the elevation of the hangingwall, the elevation of the footwall and the thickness is generated by unconditional simulation in order to draw geometric samples to conduct multiple imputation and check the results of multiple imputation.

The elevation of the hangingwall and the elevation of the footwall are simulated unconditionally and independently. The two sets of simulated normal values are transformed into original units. The elevation of the hangingwall follows a normal distribution with a mean of 100 and a variance of 5, and the elevation of the footwall follows a normal distribution with a mean of 70 and a variance of 5. Figure 3.1 shows 2-D location maps of the elevation of the hangingwall and the elevation of the footwall in the simulated true dataset.

Because the elevation of the hangingwall and the elevation of the footwall are simulated independently, there exists a possibility that the two surfaces would cross each other; however, no crossing of the hangingwall and the footwall is observed.

Then, the thickness values in the true dataset can be calculated by subtracting the elevation of the footwall from the elevation of the hangingwall. The thickness follows a normal distribution with a mean of 30 $(100 - 70)$ and a variance of 7.07 $(\sqrt{5^2 + 5^2})$. Other types of distributions could be considered. Figure 3.2 displays a map of the calculated 2-D locations and values of thickness. These distributions are used to build the cumulative distribution function and inverse cumulative distribution function for data transformation.




(a) (b)

Figure 3.1: Location Map of the Simulated Elevation Data for Single-Layered Deposits

In (a), the simulated values of the elevation of the hangingwall are shown in the location map. In (b), the simulated values of the elevation of the footwall are shown in the location map.

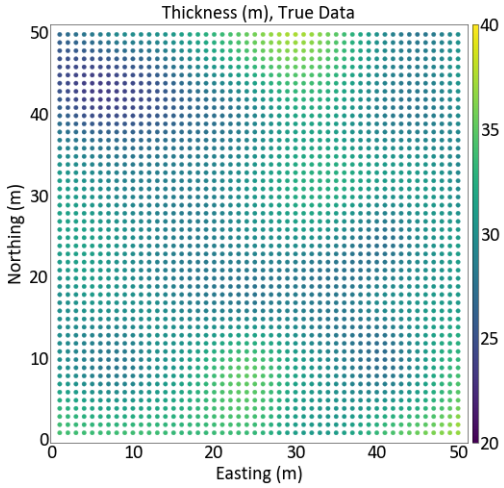
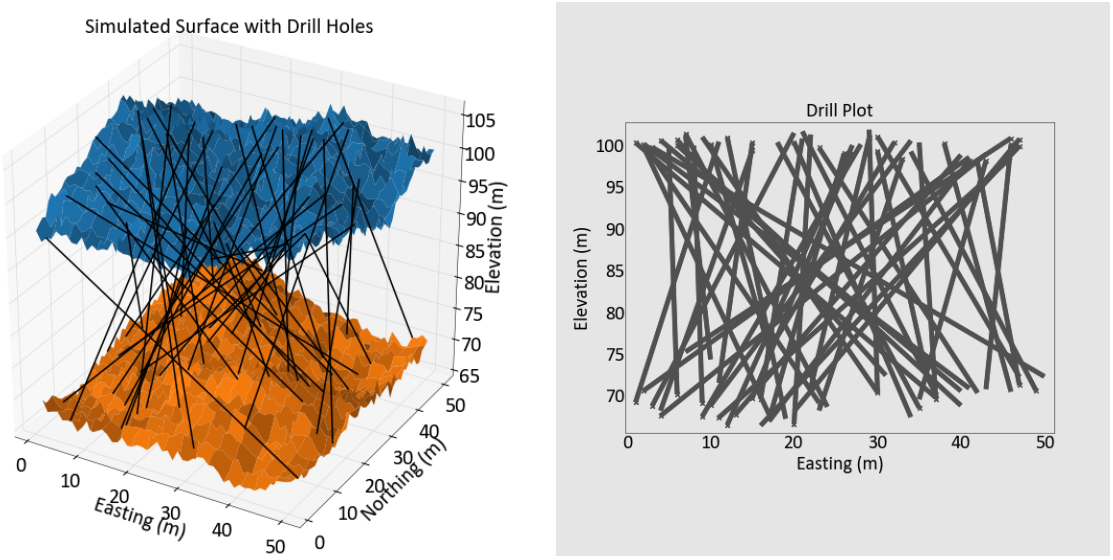



Figure 3.2: Location Map of the Calculated Thickness Data for Single-Layered Deposits

The thickness that is derived from the simulated elevation of the hangingwall and the elevation of the footwall is shown in the location map.

After the generation of the true dataset, it is possible to draw samples from the dataset. A set of synthetic drill holes are drawn randomly to mimic a drill hole campaign by selecting 50 pairs of $U$ and $V$ coordinates at both the hangingwall

wall and the footwall. Straight lines are used to connect these pairs of pierce points with the surfaces to form drill holes samples. It is assumed that all drill holes are straight. A triangulated surface plot of the true simulated data with drill holes is shown in Figure 3.3a. The blue surface is the hangingwall, and the orange surface is the footwall. The black straight lines represent drill holes. Figure 3.3b shows a cross-sectional drill plot.




(a) (b)

Figure 3.3: Simulated Surfaces with Drill Holes

In (a), the sample is synthesized by unconditional simulation of the hangingwall and the footwall and drawing drill holes that pierce both the hangingwall and the footwall. In (b), a cross-sectional view of the drill holes is shown.

Although random generation can provide unrealistic highly deviated drill holes, it can provide challenging data for testing. And the unrealistically small distance (0m to 50m) of easting and northing in the true dataset is for demonstration only.

### 3.1.2 Single-Layered Deposit Example 1

The workflow of multiple imputation is applied. The prerequisites of the method of multiple imputation include angle and thickness calculation, normal score transform, and variograms calculation and modeling. The first step is to calculate the angles of the 50 drill holes and the thickness values based on the angle tolerance. An angle tolerance of 30 degrees is used. All drill holes with an angle larger than 30 degrees

will be imputed in the workflow. This empirical limit is decided by trial and error in order to improve the performance of the method. Other angle tolerances are tried and will be discussed in Section 3.2.1. A distribution plot of the angles is shown in Figure 3.4. The angles of the 50 drill holes have a range from 10 degrees to 60 degrees. The vertical dashed line shows the angle tolerance of 30 degrees, keeping 11 drill holes as perpendicular to the plane of the continuity. Other drill holes are thus considered as inclined drill holes and will be imputed.

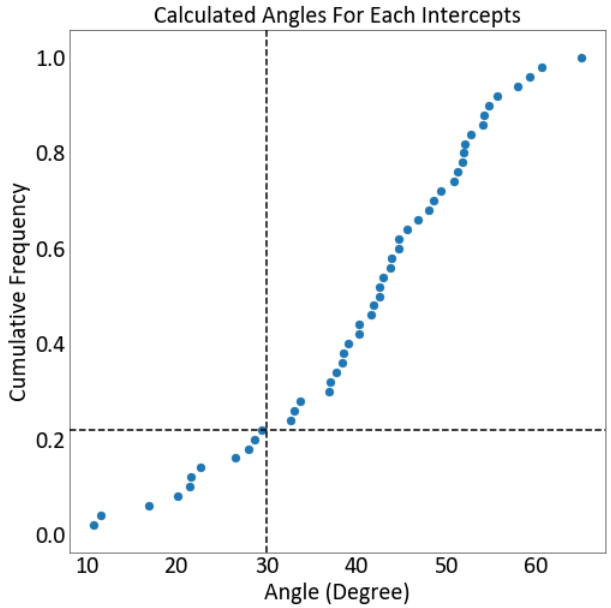
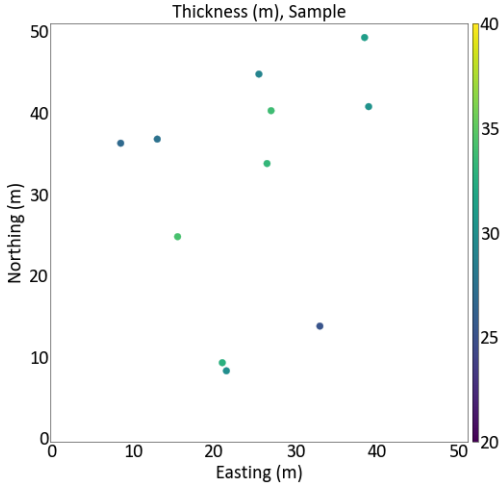



Figure 3.4: Cumulative Distribution of the Angles of the Drill Holes for Single-Layered Deposit Example 1

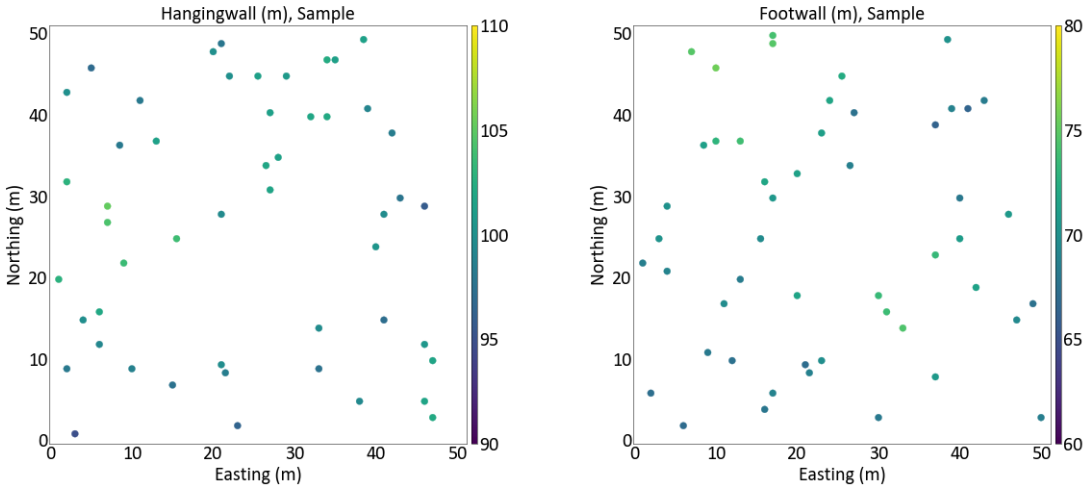
The cumulative distribution function of the angles of the drill holes are plotted and an angle tolerance of 30 degrees is used.

The geometric data samples used in the method of multiple imputation can be formed after calculating the angles and the thicknesses. Figure 3.5 displays a 2-D location map of the elevation of the hangingwall, the elevation of the footwall, and the thickness. The elevation values are derived directly from the coordinates of the pierce points, and the thickness values are calculated. The values in normal score units at these locations are used to conduct multiple imputation. There are 50 drill holes, so there are 50 elevation values of the hangingwall and 50 elevation values of the footwall. Based on the drill holes and the angle tolerance, there are 11 drill holes

having angles less than 30 degrees. As a result, there are only 11 thickness values that can be calculated by Equation 2.2. Other drill holes only provide information about the elevation of the hangingwall and the elevation of the footwall. Consequently, the sample size of the thickness is always equal or smaller than the sample size of the elevation. The total sample size after imputation will be 89 $(50 + 50 - 11)$.




(a)

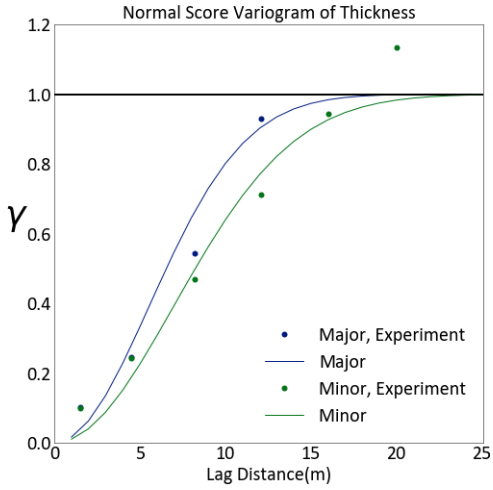



(b) (c)

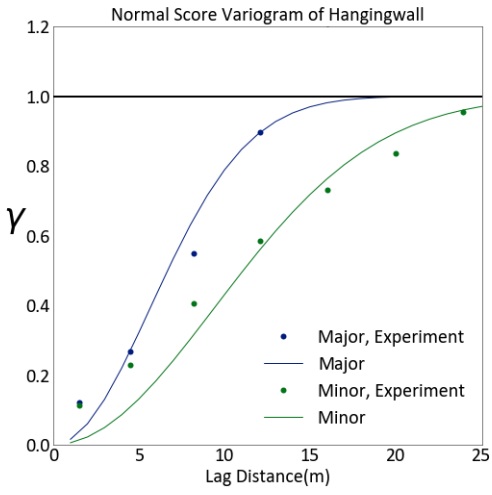
Figure 3.5: Location Maps of the Sample for Single-Layered Deposit Example 1

In (a), the location map of the thickness in the sample is shown. In (b), the location map of the elevation of the hangingwall in the sample is shown. In (c), the location map of the elevation of the footwall in the sample is shown.

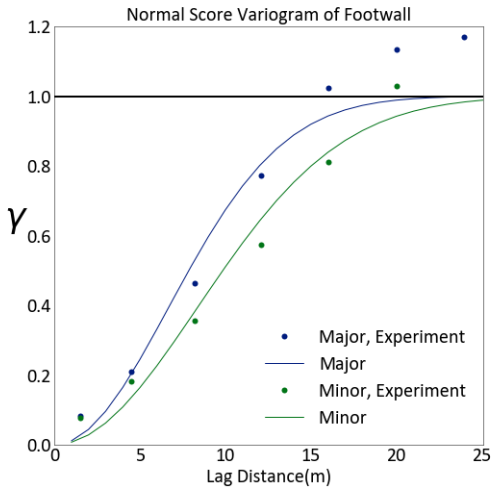
In this example, it is assumed that the variograms of geometric variables are available. The variograms are calculated and modeled from the true data. Figure 3.6 shows the modeled variograms. If there are few thickness values and it is hard to model the thickness variogram, it is possible to assign the variogram of one of the surfaces as the variogram of the thickness.




(a)




(b)




(c)

Figure 3.6: Modeled Variograms for Single-Layered Deposit Example 1

In (a), the calculated and modeled variogram of the thickness is shown. In (b), the variogram of the elevation of the hangingwall is shown. In (c), the variogram of the elevation of the footwall is shown.

The imputation includes forming local distribution of elevations and thicknesses by kriging, transforming the distribution of thickness into the distribution of elevation, merging the distribution of thickness and the distribution of elevation, and drawing an imputed value from this merged distribution. The imputation proceeds from the most constrained area to the least constrained area.

It is assumed that all the distributions of geometric variables are available in this example. As a result, the parametric cumulative distribution functions and inverse cumulative distribution functions can be used to conduct transformations.

After imputation, every location will have one value for all the geometric variables. Figure 3.7 shows the location maps of samples before imputation as well as one imputed realization. By comparing these location plots, it is possible to see that multiple imputation transforms the heterotopic geometric data samples into homotopic geometric data samples. The homotopic geometric data samples includes all the pierce points from the hangingwall and footwall, as well as all the thickness values. The number of imputed locations of the elevation of the hangingwall and the elevation of the footwall are the same. The number of the imputed locations of the thickness is bigger. After imputation, it is possible to model the deposit by having one elevation of the surface and one corresponding thickness value.

After imputing 100 realizations, the imputed values are validated against the true values. E-type values, which are the mean values of all the imputed realizations, are used to compared with the true values. Three summary statistics: error rate, Pearson correlation coefficient ($\rho$) and root mean square error (RMSE) are summarized in Table 3.1. The error rate is defined by the following formula and the unit is percentage.

$$\text{Error rate} = \frac{|\text{Imputed Value} - \text{True Value}|}{\text{True Value}} * 100\% \tag{3.1}$$

The $\rho$ is defined by the following formula and has no unit.

$$\rho = \frac{\sum_{i=1}^{n} \left(\text{True Value}_i - \overline{\text{True Value}}\right)\left(\text{Imputed Value}_i - \overline{\text{Imputed Value}}\right)}{\sqrt{\sum_{i=1}^{n} \left(\text{True Value}_i - \overline{\text{True Value}}\right)^2 \sum_{i=1}^{n} \left(\text{Imputed Value}_i - \overline{\text{Imputed Value}}\right)^2}} \tag{3.2}$$

The RMSE is defined by the following formula and the unit is meter (m).

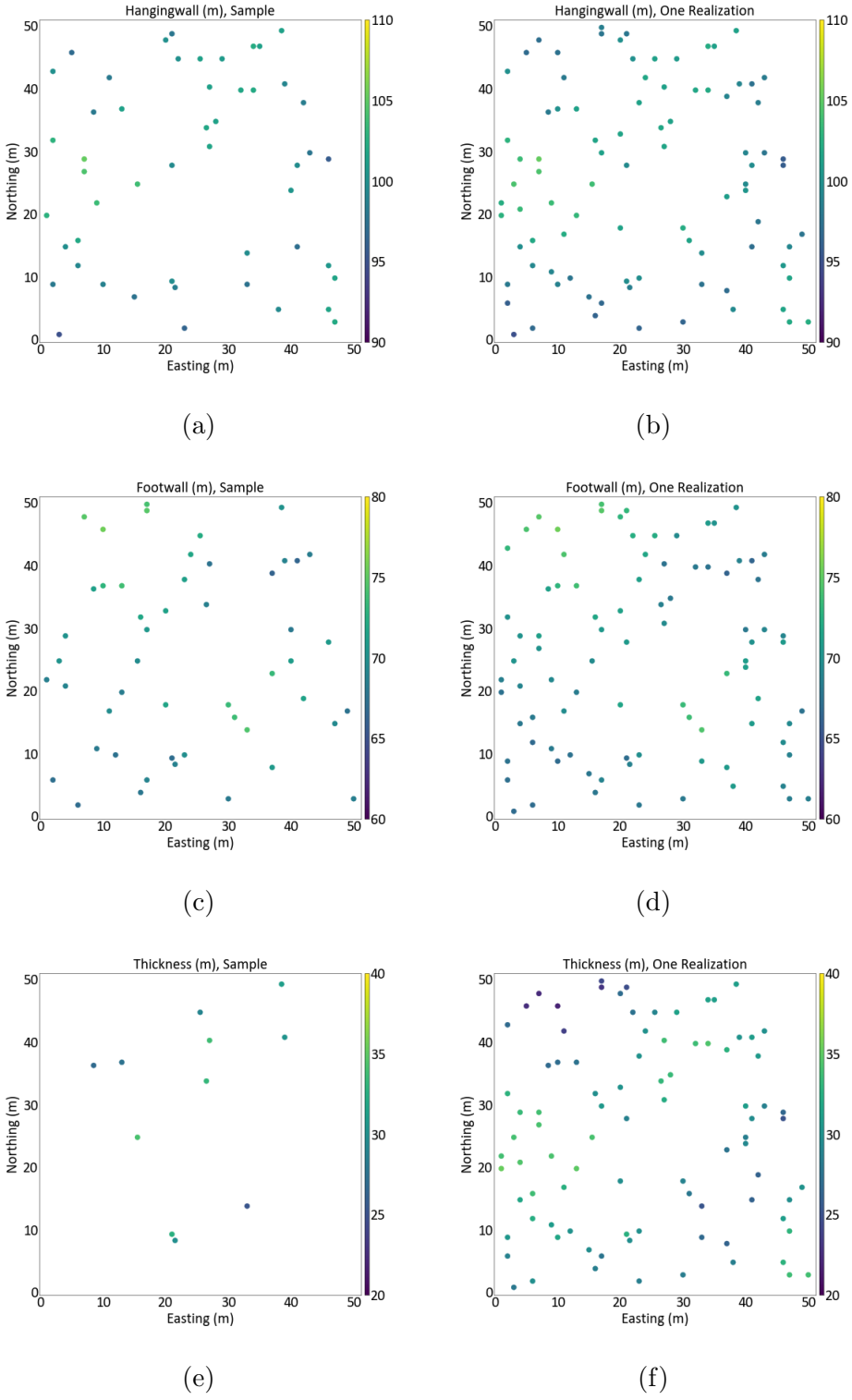



Figure 3.7: One Realization after Imputation for Single-Layered Deposit Example 1

In (a), (c), (e), the samples of the elevation of the hangingwall, the elevation of the footwall, and the thickness before imputation are shown respectively. In (b), (d), (f), one realization of the elevation of the hangingwall, the elevation of the footwall, and the thickness after imputation is shown.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\text{True Value}_i - \text{Imputed Value}_i\right)^2} \tag{3.3}$$

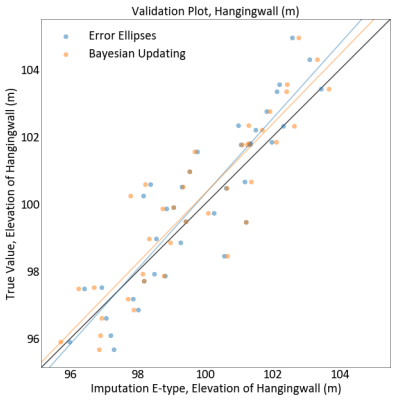
The error rate of the elevation (around 1) is smaller than the error rate of the thickness (around 3), however, the correlation of the thickness (around 0.95) is higher than the elevation (around 0.9). 0.9 means a strong correlation. The RMSE for the elevation and the thickness are more consistent (around 1). The better result is bolded in the table in order to compare the two merging methods. From the results, the method of Bayesian updating outperforms very slightly the method of error ellipses.

Table 3.1: Summary Statistics for Single-Layered Deposit Example 1

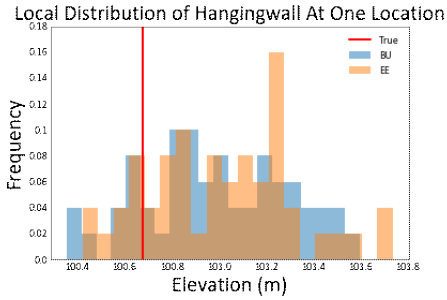
| Variable | Merging Method | Error Rate | Correlation | RMSE |
|---|---|---|---|---|
| $W_{HW}$ | EE | 0.926 | 0.899 | 1.127 |
| $W_{HW}$ | BU | **0.890** | **0.901** | **1.104** |
| $W_{FW}$ | EE | 1.196 | 0.870 | **1.017** |
| $W_{FW}$ | BU | **1.161** | **0.871** | 1.035 |
| TH | EE | 2.931 | **0.955** | 1.073 |
| TH | BU | **2.840** | 0.953 | **1.070** |

Accuracy and bias of the results are also compared in Figure 3.8a, 3.8c, and 3.8e. In these scatter plots, e-type estimates are plotted on the x-axis against the true values that are plotted on the y-axis. The overall regression line shows a strong correlation, which is consistent with the calculated correlation. The regression line illustrates a very slight conditional bias. The method tends to over-estimate the low values and under-estimate the high values. The two merging methods are also plotted and again similar results are produced.

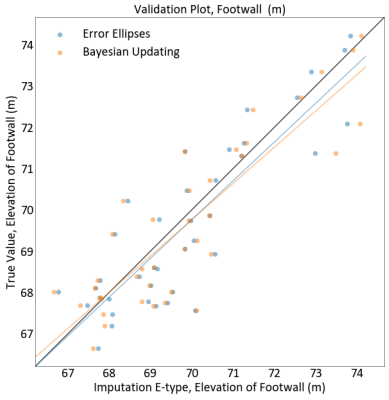
Histograms of local distributions of the elevation of the hangingwall, the elevation of the footwall, and the thickness at one random location are plotted in Figure 3.8b, 3.8d, and 3.8f, respectively. The local distributions from 100 realizations are plotted in blue and orange, and the true values are plotted in a red line. The local estimates can be slightly away from the true values. Again, the two merging methods show similar results.



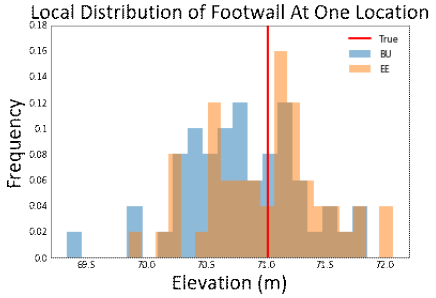

(a)



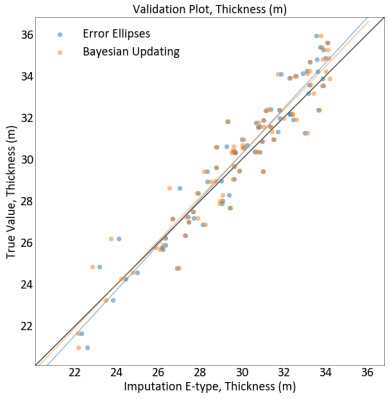

(b)



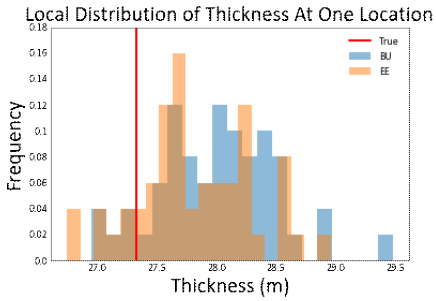

(c)




(d)




(e)




(f)

Figure 3.8: Validation Plot for Single-Layered Deposit Example 1

In scatter plots (a), (c), (e), the imputed values of the elevation of the hangingwall, the elevation of the footwall, and the thickness are plotted against the true values, respectively. In histograms (b), (d), (f), one local distribution of the three geometric variables is plotted.

Reproduction of the cumulative distribution of the imputed one hundred realizations is shown in Figure 3.9. The imputed distributions are plotted in gray. The true distributions in red and the sample distributions in blue are plotted for comparison. The imputed distributions overlap with the true distributions as well as the sample distributions. The imputed distributions have higher variance than the sample distributions, but lower variance than the true distributions. The method reproduces the cumulative distribution quite well overall, but it slightly reduces the variance for all three geometric variables. The reproduction of the thickness seems better than the elevation. After comparing Figure 3.9a with Figure 3.9d, Figure 3.9b with Figure 3.9e, and Figure 3.9c with Figure 3.9f, it seems that the two merging methods perform similarly.

Reproduction of the variograms of the imputed one hundred realizations are displayed against the true variograms in Figure 3.10. The true variograms are plotted in red, and the variograms from the realizations are plotted in blue and green. The method can reproduce the overall spatial relationship, but the anisotropy does not appear to be reproduced very well. After comparing Figure 3.10a with Figure 3.10b, Figure 3.10c with Figure 3.10d, and Figure 3.10e with Figure 3.10f, it seems that the two merging methods perform similarly.

The overall performance of the method of multiple imputation appears reasonable in this case when all the distributions and the variograms of the variables are known. The method has an acceptable correlation, and can reproduce the distribution and variogram.

### 3.1.3 Single-Layered Deposit Example 2

The same workflow of multiple imputation and the same samples are used in this example; however, it is assumed that the distribution of all three geometric variables, and the variograms models are not known. So, the transformation methods are now non-parametric and the variograms need to be calculated and modeled from the samples shown in Figure 3.5.

The prerequisites for the method are the same. The angle and thickness calculation can be done without having the distributions and variograms. As a result, the same samples can be derived. However, the normal score transform and variograms

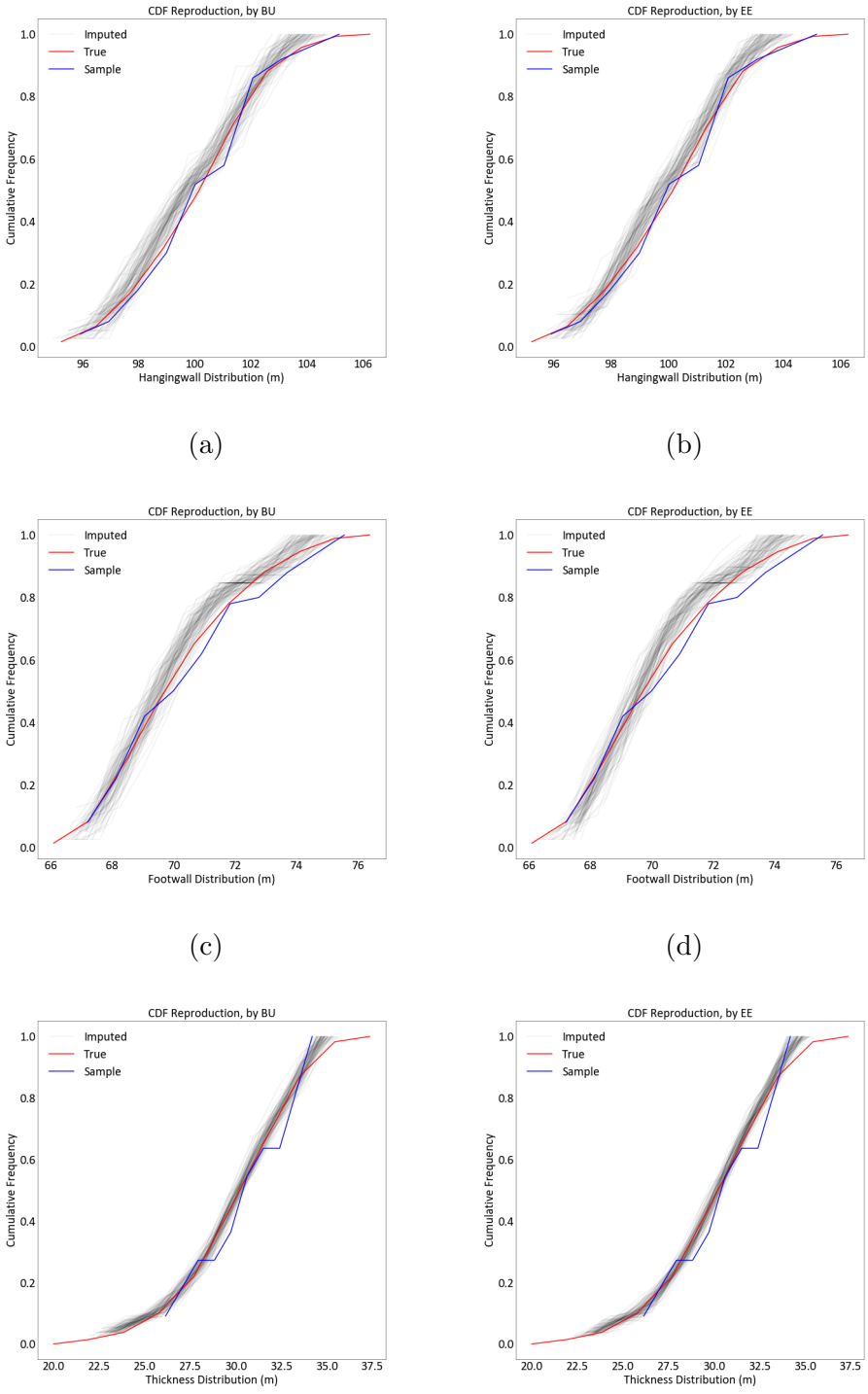



Figure 3.9: Distribution Reproduction of Single-Layered Deposit Example 1

In plot (a), (c), (e), the distribution of the imputed realizations by Bayesian updating of the elevation of the hangingwall, the footwall, and the thickness are plotted, respectively. In plot (b), (d), (f), the results by error ellipses are plotted.

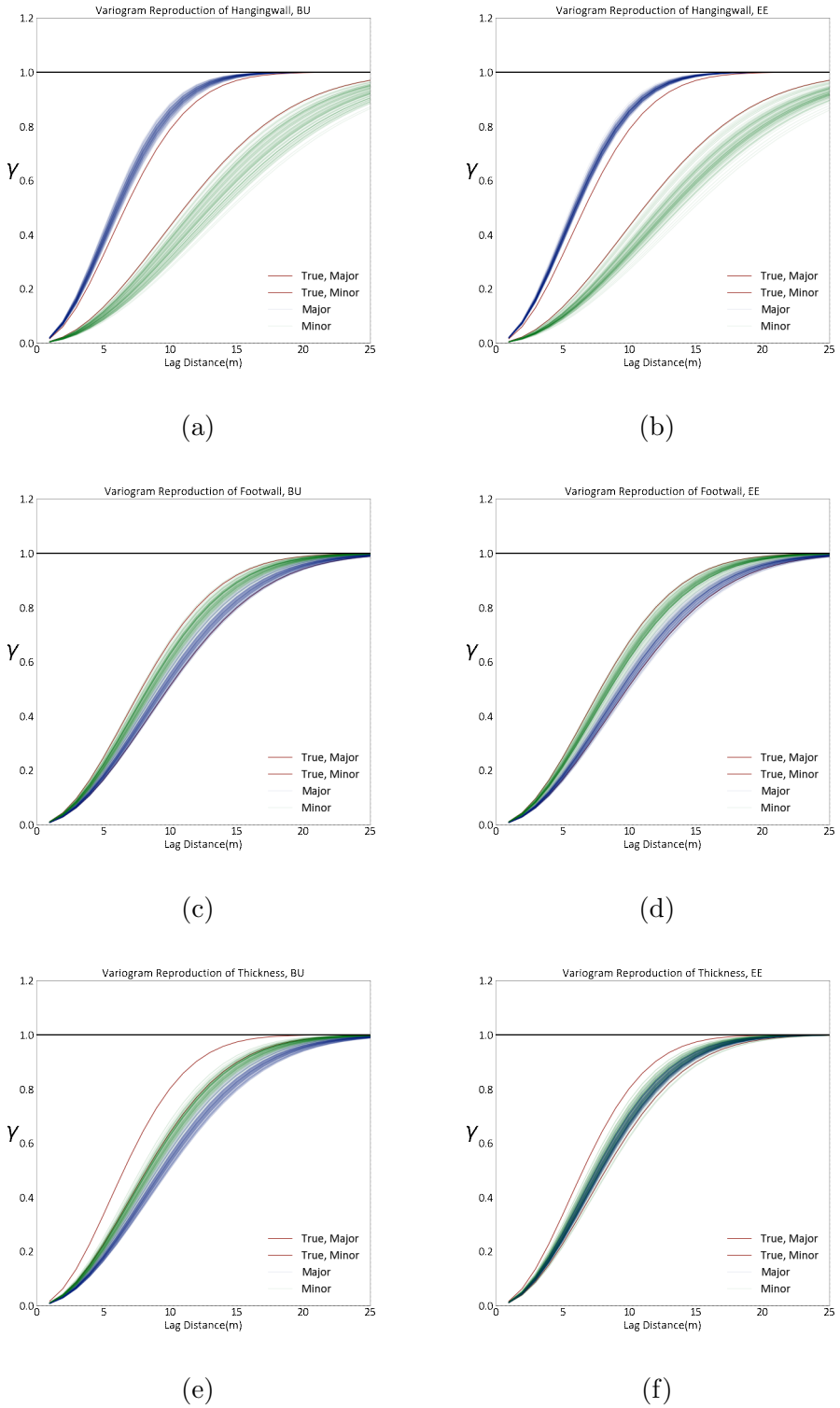
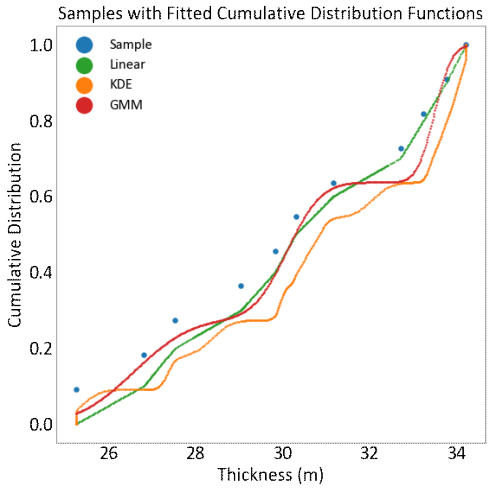



Figure 3.10: Variogram Reproduction of Single-Layered Deposit Example 1

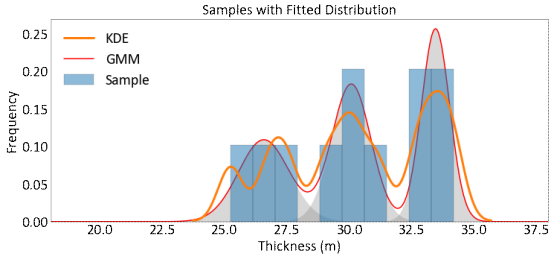
In plot (a), (c), (e), the variogram of the imputed realizations by Bayesian updating of the elevation of the hangingwall, the footwall, and the thickness are plotted, respectively. In plot (b), (d), (f), the results by error ellipses are plotted.

calculation and modeling are different.

Three methods, linear interpolation on a non-parametric CDF (Linear), kernel density estimation (KDE), and Gaussian mixture model (GMM) form different empirical distributions and allow transformation from original units to normal score units and vice versa. These three methods allow a comparison. Figure 3.11a shows fitted cumulative distribution functions by the three methods and sample values of the thickness. The sample values are in blue, and three methods use these samples values to construct non-parametric distributions. It seems that the empirical distributions fitted by Linear in green and GMM in red provide better results than the one fitted by KDE in orange. Figure 3.11b shows a histogram of the thickness and the probability density function fitted by KDE and GMM. Using the method of KDE and GMM enables imputation to provide extreme values that are beyond the range of the original distribution of geometric variables. A comparison of the imputation results will be show in Section 3.2.3.




(a) CDF of Thickness Fitted with Different Methods




(b) Histogram of Thickness Fitted with KDE and GMM

Figure 3.11: Fitted Thickness Distributions of Single-Layered Deposit Example 2

In (a), the empirical cumulative distribution of the sample of the 11 thickness values is shown with three different fitting methods. In (b), the histogram of the thickness distribution is shown with the fitted probability density function.

The variograms of the variables are calculated and modeled based on the sample in Figure 3.5. The true variograms of the three geometric variables in Section 3.1.2

are shown on the left of Figure 3.12 and the calculated and model variograms using the samples are shown on the right of Figure 3.12. The variograms of the elevation of the hangingwall and the variograms of the elevation of the footwall are less stable but can still be modeled, however, because there are fewer samples for the thickness, the variogram of the thickness is very different from the true variogram of the thickness. It can not be used for kriging. In this example, the variogram of the thickness is set to be the variogram of the true thickness in Figure 3.12e. If the true variogram of the thickness is not available, it is possible to set the variogram of the thickness to be the variogram of a more stable surface because geometric variables tend to be very continuous. In this example, the variogram of the thickness can be set to be the variogram of the elevation of the footwall shown in Figure 3.12d.

Then, the same method of forming and merging distributions is applied except that the three non-parametric methods are used to conduct transformation on the three geometric variables. After imputing 100 realizations, every location will have one value for all the geometric data and the imputed values are validated against the true values. The results using GMM are shown in this section.

The three summary statistics which compare the E-type values and true values are summarized in Table 3.2. The calculation of the three summary statistics can be found in Section 3.1.2. The overall performance based on the non-parametric methods is not as accurate as the performance based on the correct distribution functions. The error rate of the elevation changes from 1 to 1.3. The error rate of the thickness changes from 3 to 4. The correlation decreases from 0.9 to 0.8. The RMSE increases from 1 to 1.5. The results of the elevation are different from the results of the thickness.

The best result is bolded in the table. From the results, the method of Bayesian updating outperforms very slightly the method of error ellipses. The slight advantage of Bayesian updating in this example is smaller than the first example because prior distribution is uncertain.

Accuracy and bias of the results are also compared in Figure 3.13a, 3.13c, and 3.13e. In these scatter plots, e-type estimates are plotted on the x-axis against the true values on the y-axis. The overall regression line shows a moderate to strong correlation, which is consistent with the calculated correlation. The regression line illustrates a slight conditional bias. This conditional bias is larger than the bias in

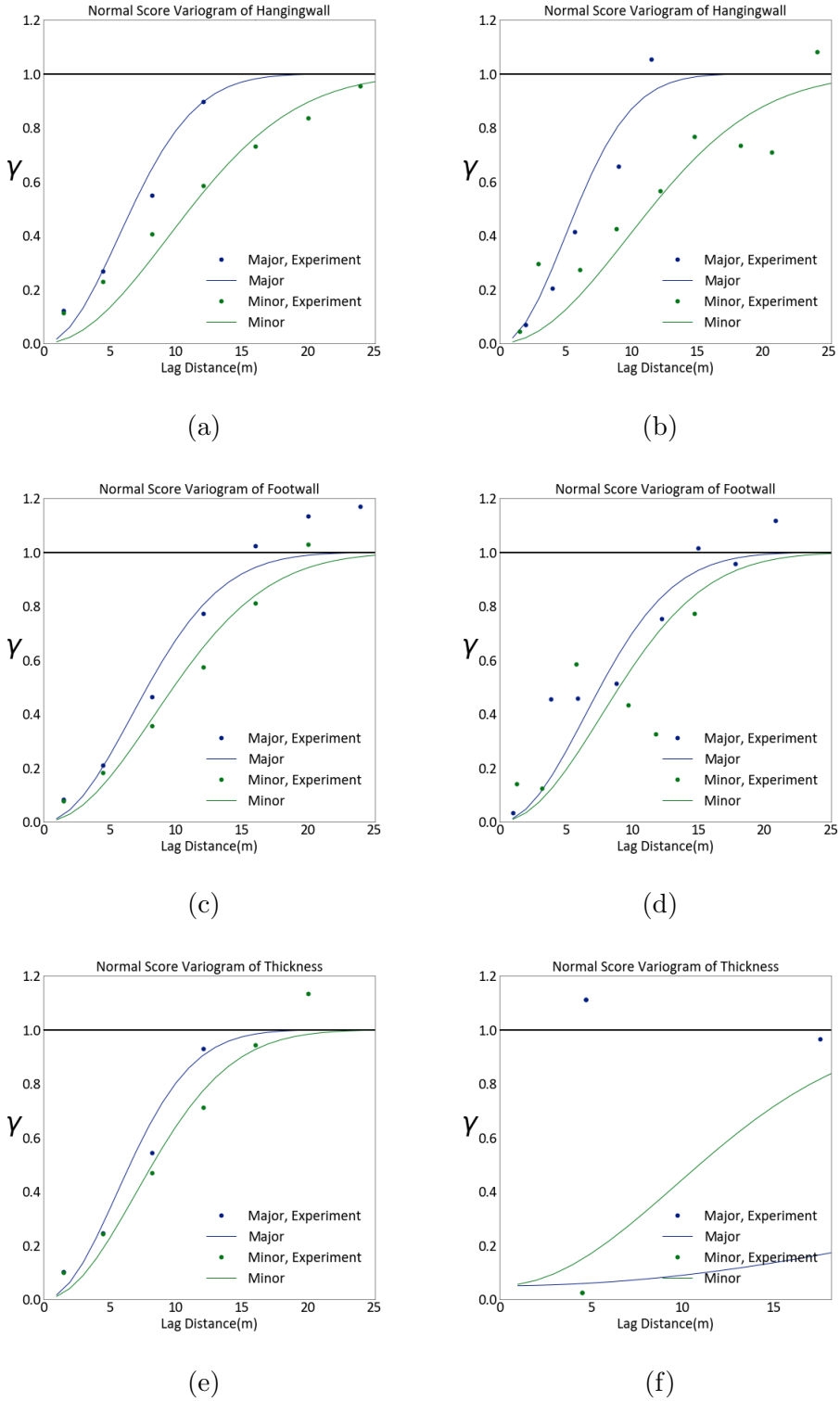



Figure 3.12: Modeled Variograms for Single-Layered Deposit Example 2

In plot (a), (c), (e), the variogram of the elevation of the hangingwall, the footwall, and the thickness from the true data are plotted, respectively. In plot (b), (d), (f), the modeled variograms from the sample are plotted. The variogram of the thickness can not be used.

Table 3.2: Summary Statistics for Single-Layered Deposit Example 2

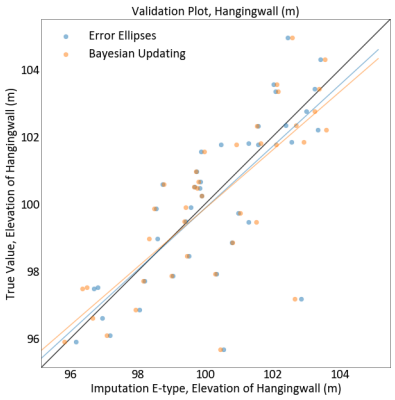
| Variable | Merging Method | Error Rate | Correlation | RMSE |
|---|---|---|---|---|
| $W_{HW}$ | EE | **1.180** | 0.748 | 1.632 |
| $W_{HW}$ | BU | 1.189 | **0.761** | **1.615** |
| $W_{FW}$ | EE | 1.505 | 0.730 | **1.395** |
| $W_{FW}$ | BU | **1.503** | **0.733** | 1.397 |
| TH | EE | **3.925** | **0.905** | 1.518 |
| TH | BU | 3.928 | **0.905** | **1.510** |

the previous example. The method tends to over-estimate the low values and under-estimate the high values again. The two merging methods are also plotted and again similar results are produced.

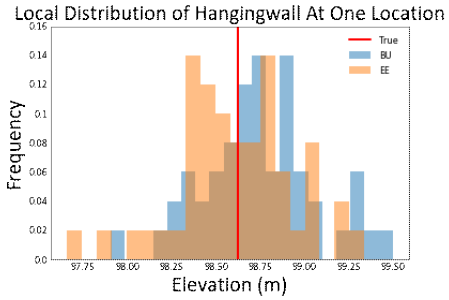
Histograms of local distributions of the elevation of the hangingwall, the elevation of the footwall, and the thickness at one random location are plotted in Figure 3.13b, 3.13d, and 3.13f, respectively. The local distributions from 100 realizations are plotted in blue and orange, and the true values are plotted in a red line. The local estimates can be slightly different from the true values. Again, the two merging methods show similar results.

Reproduction of the cumulative distribution of the imputed one hundred realizations is shown in Figure 3.14. The imputed distributions are plotted in gray. The true distributions are plotted in red and the sample distributions plotted in blue are shown for comparison. The imputed distributions usually overlap with the true distributions as well as the sample distributions.

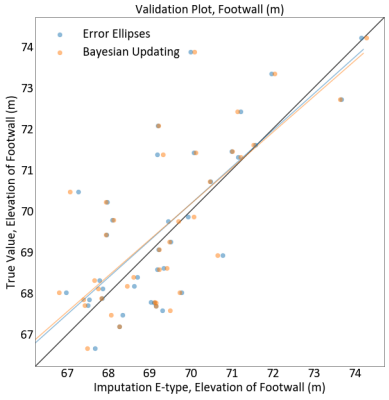
The method reproduces the cumulative distribution approximately well overall, but it reduces the variance for all three geometric variables because the transformation is based on the sample data. This result is not clear in the reproduction of the elevation of the hangingwall, but the reproduction of the elevation of the footwall shows a reduction in variance and some deviations. Furthermore, this result is very clear in the reproduction of the thickness distribution. In Figure 3.14e and Figure 3.14f, the reproduction of the thickness is mainly based on the samples. As a result, the reproduction is not as accurate as the previous case. After comparing the results



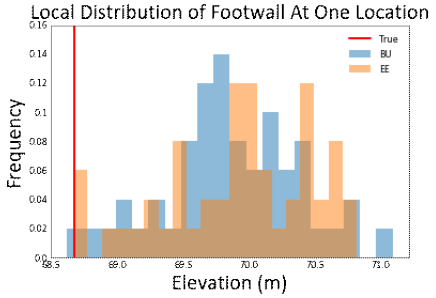

(a)



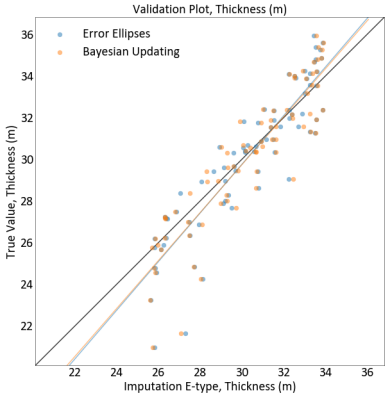

(b)



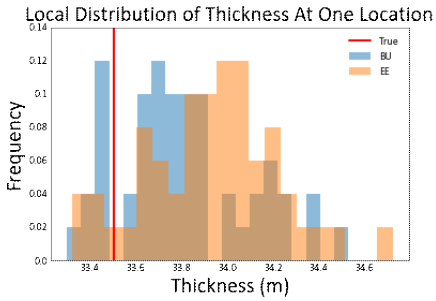

(c)




(d)




(e)




(f)

Figure 3.13: Validation Plot for Single-Layered Deposit Example 2

In scatter plots (a), (c), (e), the imputed values of the elevation of the hangingwall, the elevation of the footwall, and the thickness are plotted against the true values, respectively. In histograms (b), (d), (f), one local distribution of the three geometric variables is plotted.

by Bayesian updating on the left of Figure 3.14 and the results by error ellipses on the right of Figure 3.14, it seems that the two merging methods perform similarly.

Reproduction of the variograms of the imputed one hundred realizations are displayed against the true variograms in Figure 3.15. The true variograms are plotted in red, and the variograms from the realizations are plotted in blue and green. The method can reproduce approximately reproduce the spatial relationship. However, the reproduction is not as good as the reproduction in the previous example. The anisotropy does not appear to be reproduced fully and the reproduction of the variogram of the thickness appears poor. After comparing Figure 3.15a with Figure 3.15b, Figure 3.15c with Figure 3.15d, and Figure 3.15e with Figure 3.15f, it seems that the two merging methods perform nearly the same for the elevation. Error ellipses performs better for the thickness.

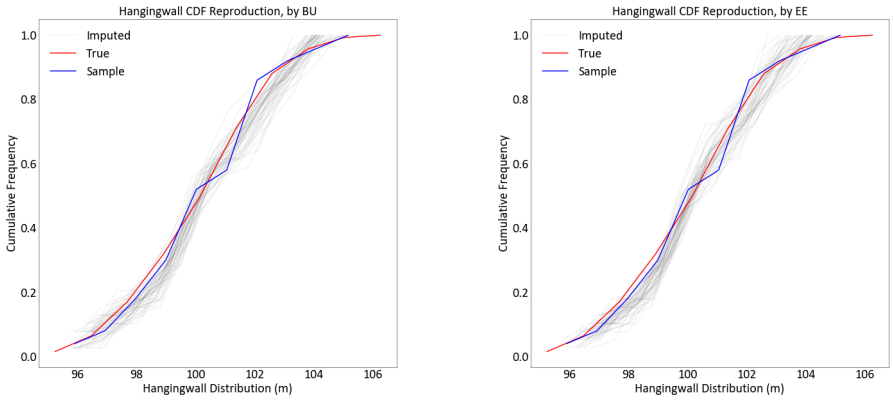
The performance of the method based on GMM provides reasonable results. Of course, it is better to have the full correct distribution. These two examples demonstrate the method of multiple imputation, which includes the primary spatial elevation values and secondary collocated thickness values together when constructing local conditional distributions. The correlation, the reproduction of the global distribution, and the reproduction of spatial continuity appears acceptable.

## 3.2 Sensitivity Analysis of Single-Layered Deposit

Various decisions and parameters during the process of multiple imputation can influence the performance of the method. These parameters include the angle tolerance, the sequence of imputation, the transformation methods, as well as the merging method. Sensitivity analysis is conducted in order to find the best parameters to choose when conducting multiple imputation. The correlation and the RMSE values between the true values and the imputed results calculated in different cases are used to justify different decisions and parameters.

### 3.2.1 Angle Tolerance

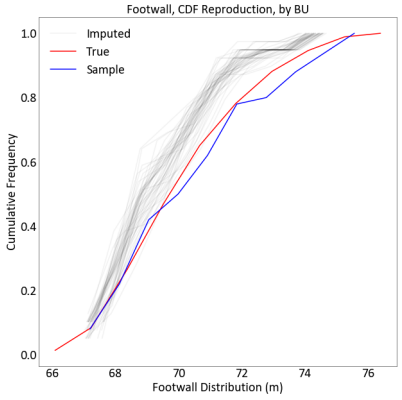
During multiple imputation, there are many quantile-to-quantile transformations. As a result, having a representative distribution, especially for thickness, is very impor-



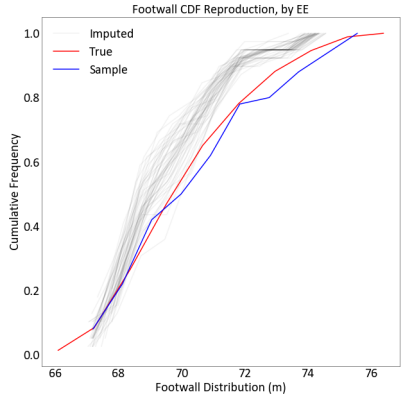

(a)



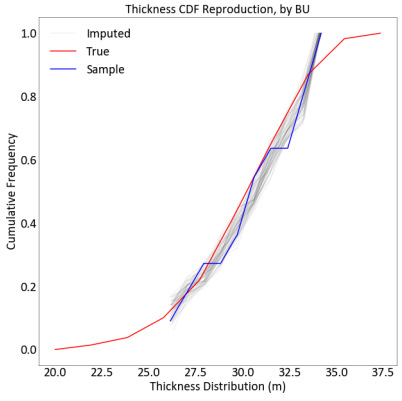

(b)



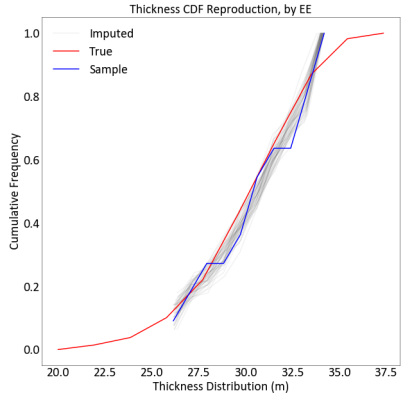

(c)




(d)




(e)




(f)

Figure 3.14: Distribution Reproduction of Single-Layered Deposit Example 2

In plot (a), (c), (e), the distribution of the imputed realizations by Bayesian updating of the elevation of the hangingwall, the footwall, and the thickness are plotted, respectively. In plot (b), (d), (f), the results by error ellipses are plotted.

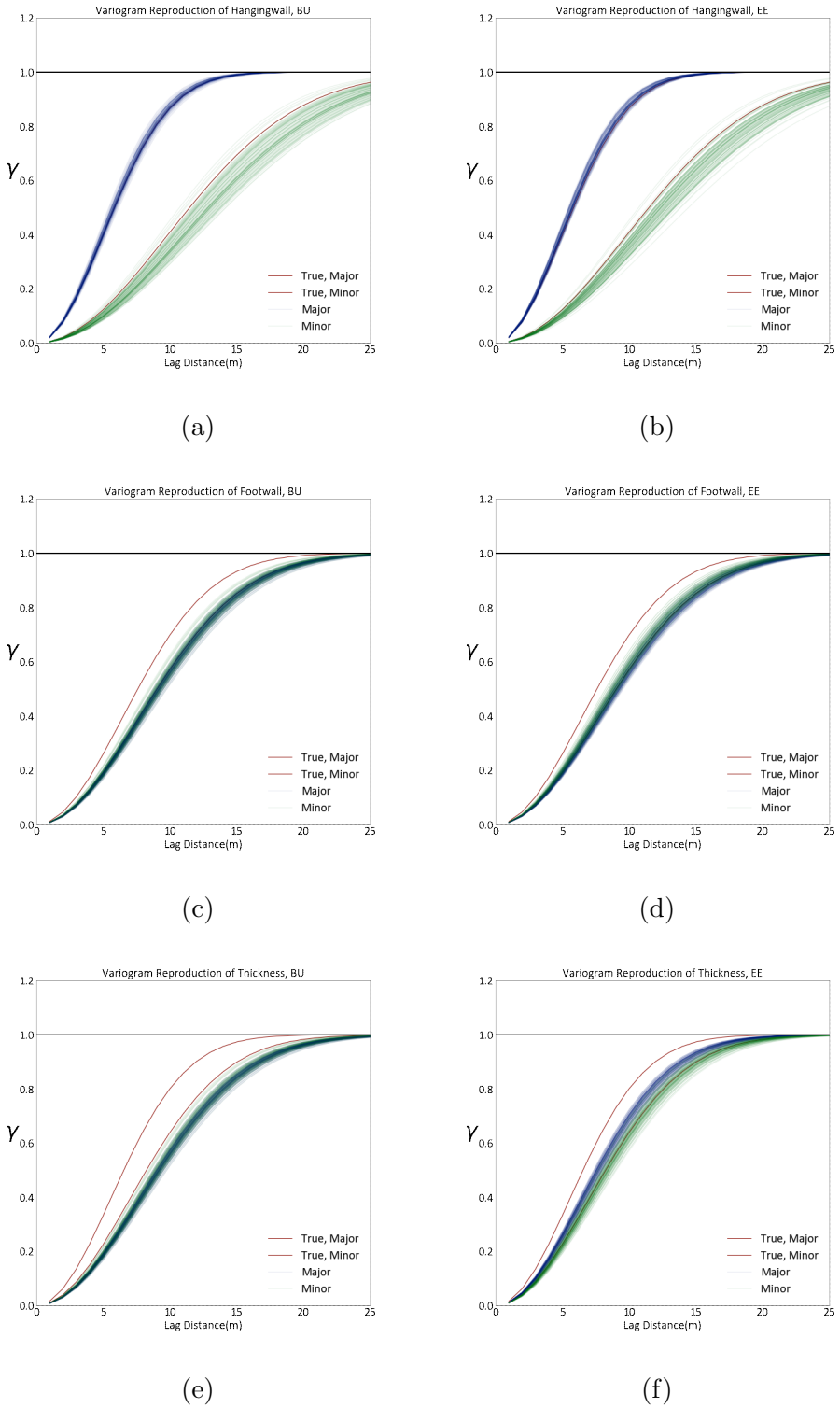
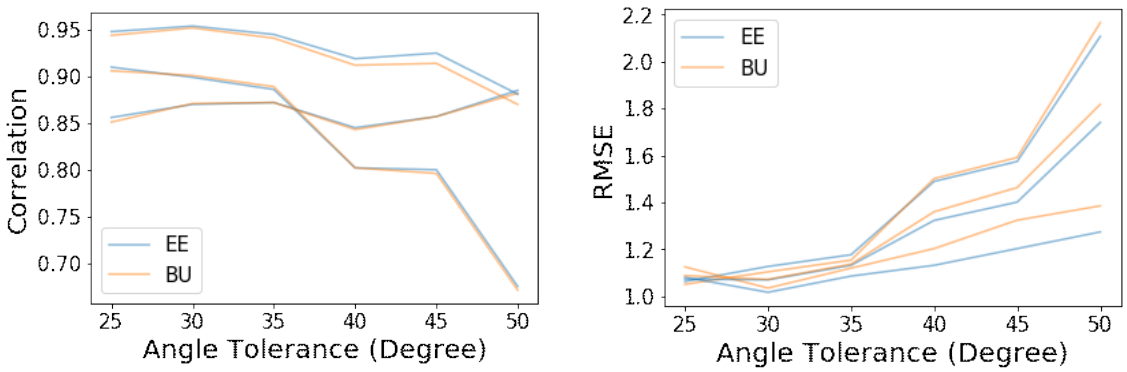



Figure 3.15: Variogram Reproduction of Single-Layered Deposit Example 2

In plot (a), (c), (e), the variogram of the imputed realizations by Bayesian updating of the elevation of the hangingwall, the footwall, and the thickness are plotted, respectively. In plot (b), (d), (f), the results by error ellipses are plotted.

tant. The global distribution of thickness is built by subtracting the elevation of the footwall from the elevation of the hangingwall and depends on the angle tolerance used. In Figure 3.16, this relationship is displayed by plotting the cumulative distribution of the drill holes angles and the corresponding thickness samples. From Figure 3.4 and Figure 3.5 in Section 3.1.2, if the angle tolerance is 30 degrees, there are 11 thickness values. From Figure 3.16, if the angle tolerance is 25 degrees, there are 7 thickness values. If the angle tolerance is 35 degrees, there are 13 thickness values. If the angle tolerance is 40 degrees, there are 19 thickness values. The higher the angle tolerance is, the more thickness values are included in the calculated distribution of thickness.

It is necessary to have enough data to construct the distribution of thickness. However, a large angle tolerance and more thickness values may decrease the performance because the calculated thickness by Equation 2.2 is not exactly the true thickness. If the angle tolerance is too big, the calculated thickness values start to differ from the true thickness distribution. Different angle tolerances are tried, and the results of summary statistics are shown in Figure 3.17. Different angle tolerances from 25 degrees to 50 degrees are plotted on the x-axis, an the corresponding correlation and RMSE are plotted on the y-Axis.




(a) (b)

Figure 3.17: Correlation and RMSE of Different Angle Tolerance

In (a), the correlation of the imputed values with the true values decreases with the increase of the angle tolerance. In (b), the RMSE of the imputed values with the true values increases with the increase of the angle tolerance.

Based on Figure 3.17, having an angle tolerance of 30 degrees appears reasonable.

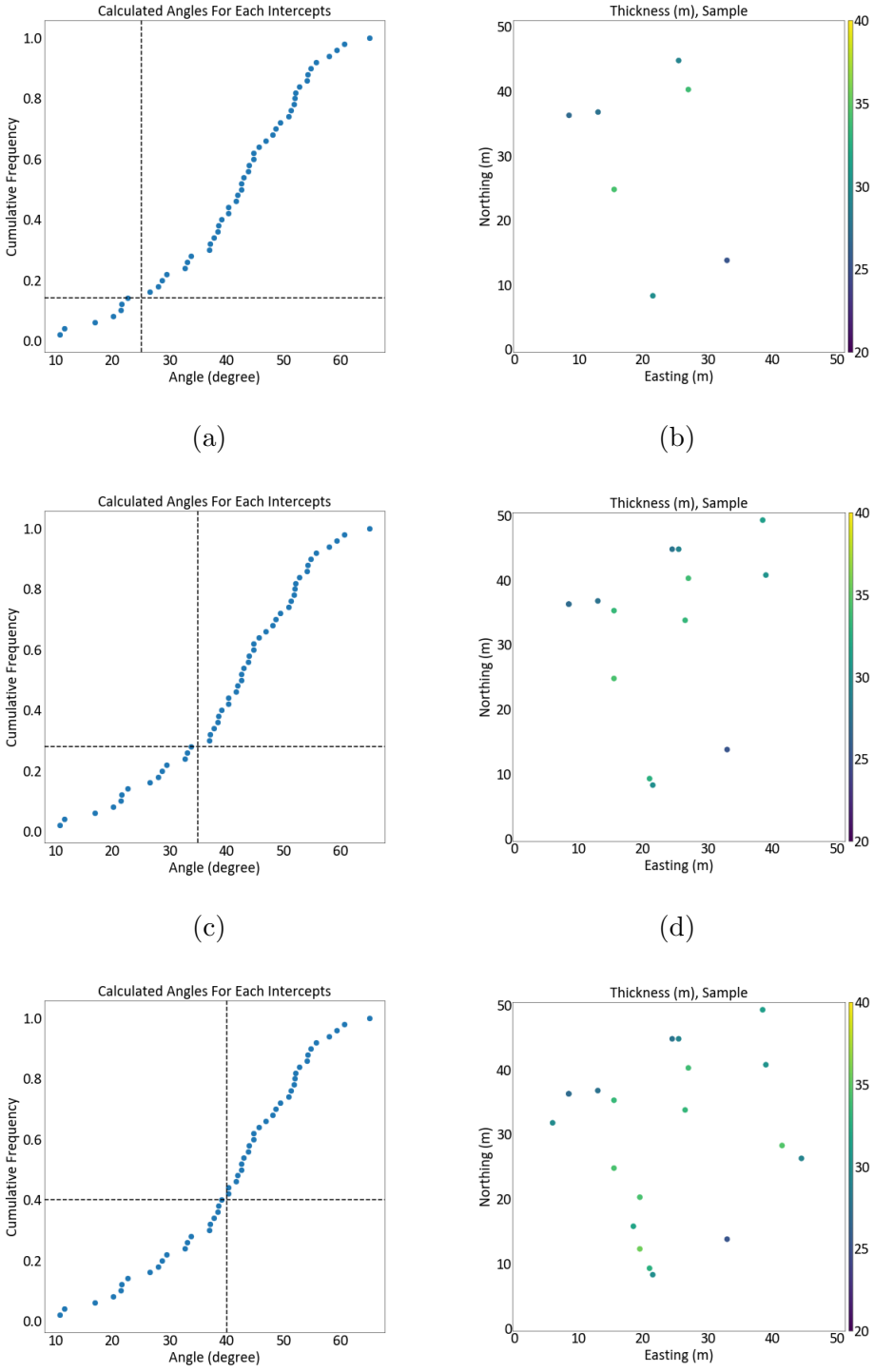


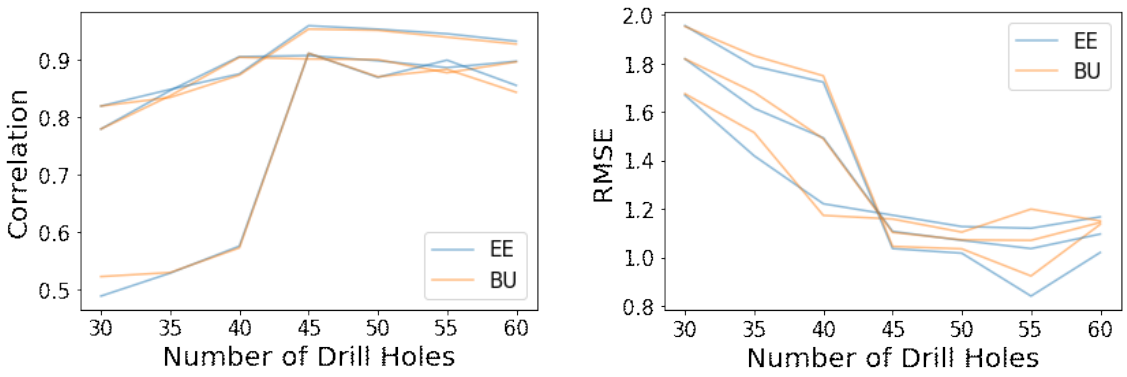

Figure 3.16: Angle tolerance and sample size

In (a), (c), (e), different angle tolerances are shown. In (b), (d), (f), the corresponding thickness values based on the angle tolerances are shown. The larger the angle tolerance, the bigger the sample size.

When the angle tolerance is larger than 35 degrees, the performance of the method decreases (the RMSE increases) because many incorrect thickness values are kept in the distribution. It is recommend to use trial and error to find the best angle tolerance for different deposits with different thickness values.

On the other hand, if too few thickness values are kept with a smaller angle tolerance, the distribution of thickness will be unstable. It is important to have enough sample values to form distributions, calculating and modeling variograms, and form kriging equations. From Figure 3.18, once enough drill holes are available, the accuracy will not change significantly. With a angle tolerance of 30 degrees, the minimum drill hole number that can produce reasonable result is 45. After this number of drill holes, the performances look the same.

By considering the angle tolerance (30 degrees) and the number of drill holes together, 10 thickness values start to produce reasonable imputation results. This number comes from the number of drill holes timing the percentage of perpendicular drill holes after applying the angle of tolerance.




(a) (b)

Figure 3.18: Correlation and RMSE of different drill holes number

In (a), the correlation of the imputed values with the true values increases with the increase of the drill hole number. In (b), the RMSE of the imputed values with the true values decreases with the increase of the drill hole number.

It is important to mention that these numbers are specific for these examples. The angle tolerance and the number of drill holes must be considered together in order to achieve the number of samples required for forming empirical distributions, variogram calculation and modeling, as well as forming the kriging equations.

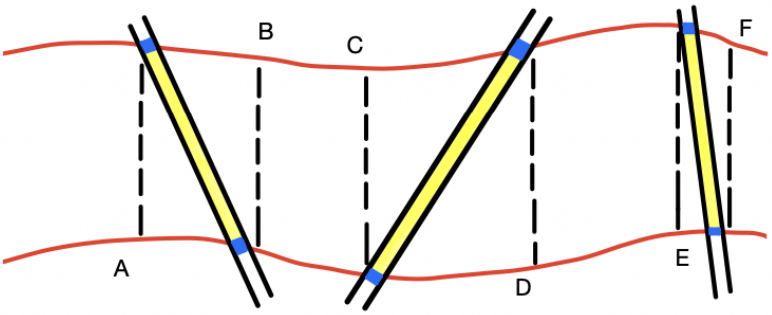



Figure 3.19: Schematic Illustration for Demonstration of Sequence of Imputation

A schematic illustration of three drill holes is shown. Six locations A, B, C, D, E, and F are going to be imputed. Several sequences are applied. (not to scale)

### 3.2.2 Imputation Sequence

Imputation starts from the most constrained location to the least constrained location. There are many ways to measure the degree of constraint. The kriging variance is one way. A smaller kriging variance means a higher constraint. Thus, the imputation could start from the location with the smallest kriging variance to the location with the highest kriging variance. Another practical measure is the angle of the drill hole (Bai and Deutsch, 2021a). Locations having vertical drill holes are more constrained than locations having inclined drill holes. A schematic illustration shown in Figure 3.19 is used to demonstrate various sequences of imputation considered.

Independent global kriging of the elevation of the hangingwall and the elevation of the footwall are used to determine the kriging variance of each location under imputation. The hypothetical calculated kriging variances are listed in Table 3.3.

There are three imputation sequences considered. In the first sequence (Kriging Variance), the calculated kriging variance values at the hangingwall and at the footwall are combined together into one sequence. This sequence is shown in the following path. The imputation starts at the location with the smallest kriging variance to the location with the largest kriging variance.

$$A\,(0.1) \Rightarrow B\,(0.2) \Rightarrow C\,(0.3) \Rightarrow E\,(0.4) \Rightarrow F\,(0.5) \Rightarrow D\,(0.6) \tag{3.4}$$

In the second sequence (Layer), the calculated kriging variance values at the hang-

Table 3.3: Kriging Variance of the Schematic Example

| Location | Kriging Variance |
|---|---|
| A | 0.1 |
| B | 0.2 |
| C | 0.3 |
| D | 0.6 |
| E | 0.4 |
| F | 0.5 |

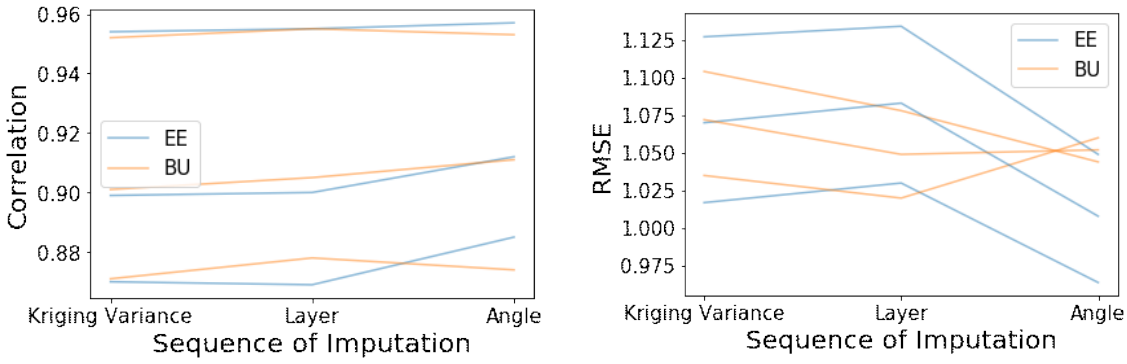
ingwall and at the footwall are compared separately at first and then combined. A more stable layer is decided at the first by comparing the variogram range. The imputation starts at the layer with the smallest variance and goes to the other layer after all locations at the first layer are imputed. It is assumed that the hangingwall is more stable. The sequence of the schematic example is shown in the following path.

$$\begin{aligned} B\,(0.2) \Rightarrow C\,(0.3) \Rightarrow F\,(0.5) \Rightarrow \\ A\,(0.1) \Rightarrow E\,(0.4) \Rightarrow D\,(0.6) \end{aligned} \tag{3.5}$$

The final sequence considered (Angle) is based on the drill hole angles. In this method, the angles of all the drill holes are calculated first. The imputation starts from the drill hole with the smallest drill hole angle. Within one drill hole, the sequence starts from the location with the smallest kriging variance. As a result, the imputation in the schematic example starts from the drill hole with E and F and to the drill hole with C and D. The sequence is shown in the following path.

$$\begin{aligned} E\,(0.4) \Rightarrow F\,(0.5) \Rightarrow \\ A\,(0.1) \Rightarrow B\,(0.2) \Rightarrow \\ C\,(0.3) \Rightarrow D\,(0.6) \end{aligned} \tag{3.6}$$

It is shown that this method provides the best results. The RMSE decreases significantly while the method is used. The results shown in Figure 3.20 reflect that there is no major difference between these two sequences.




(a) (b)

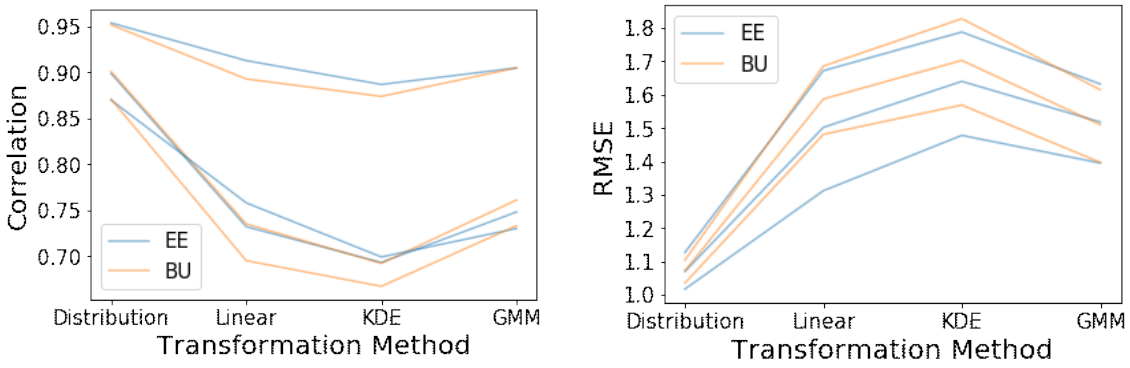
Figure 3.20: Correlation and RMSE of Different Sequence

In (a), the correlation of the imputed values with the true values increases with different sequence is shown. In (b), the RMSE is shown. The sequence based on the angle is the best.

The performance of the two merging methods: Bayesian updating and error ellipses can usually be improved together. However, in Figure 3.20a, when the third imputation sequence (Angle) is used, the performance of Bayesian updating decreases and the performance of error ellipses increases. As a result, the two merging method can behavior differently.

### 3.2.3 Transformation Method

The true distributions of the geometric data variables are not available to conduct the transformation; the transformation functions must be derived from the samples. Three non-parametric transformation methods have been tried. They are linear transformation, KDE, and GMM. From Figure 3.21, the transformation with the parametric function performs the best apparently and among the three non-parametric methods, the method of GMM produces the best results.



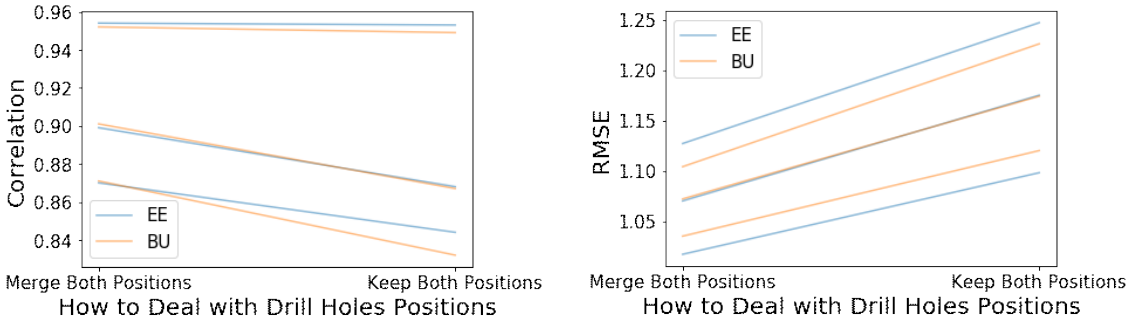

(a) (b)

Figure 3.21: Correlation and RMSE of Different Transformation Methods

In (a), the correlation of the imputed values with the true values with different transformation methods are shown. In (b), the RMSE are shown. The performance of GMM is the best in the three non-parametric method.

### 3.2.4 How to Deal with Drill Holes Positions

While the drill holes with an angle smaller than the angle tolerance are considered as perpendicular to the plane of continuity, the $U$ and $V$ coordinates of the pierce points at the hangingwall and the footwall are not exactly the same. Two methods to deal with this problem are tried. The first one is to merge these two pierce points into one point having the same $U$ and $V$ coordinate by using the mean of the $U$ and $V$ coordinates. The second one is to assume two drill holes having different $U$ and $V$ coordinate with the same thickness values. It is better to combine the two pierce points into one point from Figure 3.22 because the second method can increase the redundancy in the sample.




(a) (b)

Figure 3.22: Correlation and RMSE of Different Drill Holes Treatments

In (a), the correlation of the imputed values with the true values with different drill hole treatment methods are shown. In (b), the RMSE are shown. The method of merging two points is better.

### 3.2.5 Merging method

There are two merging methods: Bayesian updating and error ellipses. In the two previous examples in Section 3.1.2 and Section 3.1.2, Bayesian updating performs slightly better than error ellipses. From Figures 3.17, 3.18, 3.20, 3.21, 3.22 in Section 3.2, the method of error ellipses performs slightly better than Bayesian updating by comparing the correlation and RMSE. The method of error ellipses appears more stable than Bayesian updating, especially when the distributions of the variables are not known.

## 3.3 Discussion on Single-Layered Deposit

The overall performance of the method appears reasonable. However, there are some limitations related to the method. The problems come from transformation as well as the addition or subtraction of the thickness.

The first limitation is that there is no guarantee that when transforming the secondary collocated distribution derived from the thickness to the opposite elevation, the transformed values will be in the range of the original elevation distribution. In some extreme cases, imputed values must be truncated at the tails of the distributions to enable transformation. These cases do not happen if a Gaussian distribution is used to form the true dataset. However, if a distribution with a finite upper and lower limit, like uniform distribution or triangular distribution, is used to form the

true dataset, truncation at the limit could happen. For example, Figure 3.23 shows a validation plot using a true dataset formed by a triangular distribution. It is shown that there are many imputed values at the minimum values of the elevation of the footwall. During the process of imputation, many elevation values that are lower than the minimum value of the distribution are generated and they are truncated around the minimum value. Although this is an extreme cases, the method of multiple imputation with truncation tends to have a lower variability than the true values.

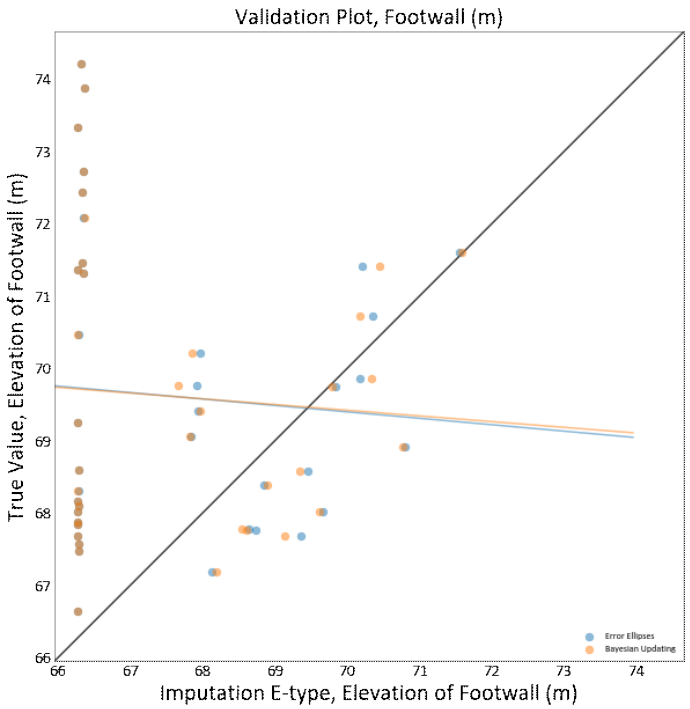



Figure 3.23: Scatter Plot of an Example with Many Truncation

There are many imputed values around the minimum values of the elevation of the footwall because of truncation.

Because of this problem, it is interesting to see that in some cases, geometric variables with higher variance values can be imputed with a high accuracy because it is less likely that the imputed values are truncated during the transformation process. About 1 percent of the imputed values are truncated. Furthermore, using a distribution of a fitting method like GMM without upper limit and lower limit can reduce the probability of truncation.

The second limitation is that when drawing the imputed value from the merged distribution, the drawn elevation value is always in the range of the distribution of the

elevation. However, again there is no guarantee that the calculated thickness value is in the range of the distribution of thickness. Iteration is used to try to find a valid pair of elevation and thickness value based on the merged distribution at the beginning. However, sometimes it is impossible to find a pair that satisfies both distributions. As a result, a random pair is drawn and set as the imputed results. This can decrease the performance of the method. In the examples shown above about 6 percent of the imputed values come from this iterative method and about 1 percent of the imputed values come from random drawing.

The third limitation is that the distribution of the thickness from the calculated thickness values does not always represent the true distribution of the thickness. The distribution from the calculated thickness values tends to have less variability for the case considered. When the parametric distributions are available, this effect is not very obvious. However, when the parametric distributions are not available, the calculated thickness values reduce the variability of the thickness greatly. In Figure 3.24, the reproduction of the thickness is only based on the samples. Furthermore, if many thickness values from inclined drill holes are used in the calculation, the calculated thickness distribution may not represent the true thickness values correctly.

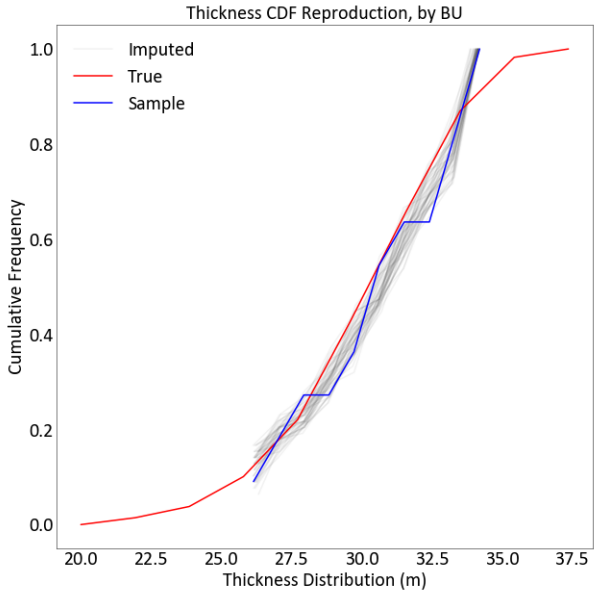



Figure 3.24: CDF Reproduction of Thickness with A Lower Variability

The reproduction of thickness values in black are derived from the sample in blue. They have lower variability than the true distribution in red.

Another characteristic of the method is that the variance of the primary distribution and the secondary distribution can be very different. There are always more elevation values than thickness values, so if the same variograms and the same searching radius are used for the elevation and the thickness, the local conditional variance of the thickness is always larger than the local conditional variance of the elevation. This variability stays when transforming the thickness values to the elevation values. So, when these two elevations distributions are merged, the primary elevation distribution derived from the spatial sources tends to dominate the final merged distribution and the secondary elevation distribution has less impact on the final merged distribution. This can further reduce the variability of the imputed values. This phenomenon decreases through the process of imputation, as more and more imputed thickness values are available.

# Chapter 4

# Multi-Layered Deposits

*The method of multiple imputation used for multi-layered tabular deposits extends from the basic method used for single-layered deposits. The difference is the sequence of imputation and the details of how thickness is transformed into elevation. A synthetic example demonstrates the application of the method. The imputed results are compared with the true values in order to assess the performance of the method. Then, a sensitivity analysis demonstrates that using the elevation from the closest opposite surface is the best way to transform thickness into elevation and the best sequence is to process imputation according to drill hole angles. This chapter ends with some challenges related to multi-layered deposits, such as the number of imputation locations and drill hole configurations that can affect the sequence of multiple imputation. The method are changed to fit this kind of configurations to run imputation smoothly.*

## 4.1 Difference Between Single-Layered and Multi-Layered Deposits

The definition of tabular vein deposits includes both single-layered and multi-layered deposits with simple structures (Carvalho, 2018). The method applied to single-layered deposits in Chapter 2 and Chapter 3 can be extended to multi-layered deposits. The basic methodology of forming a primary spatial distribution of elevation and a secondary collocated distribution of thickness remains the same. The difference is that there are different thickness values that can be used when adding to or subtracting from the known surface elevation values. This difference can change the sequence of imputation. Sensitivity analysis also shows that the merging method can behave differently.

## 4.2 Demonstration of Multi-Layered Deposits

In this section, a synthetic dataset is used to demonstrate the method. The workflow is shown and the results are compared with the true data by comparing summary statistics, reproduction of distributions and variograms. Because there is no conventional name for different surfaces in multi-layered deposits like hangingwall and footwall in single-layered deposits, the surfaces are named as Wall 1, Wall 2, Wall 3, ... from the top to the bottom. The thickness between Wall 1 and Wall 2 is named Thickness 12. For example, if this naming convention is applied to a single-layered deposit, the hangingwall is named as Wall 1, the footwall is named as Wall 2, and the thickness is named as Thickness 12.

### 4.2.1 True Data for Demonstration of Multi-Layered Deposits

A full geometric dataset is used to draw samples and evaluate the performance of the method. The dataset is generated by unconditional simulation of elevation of Wall 1, Wall 2, and Wall 3. The simulated normal score values are these three elevations are transformed into their original units separately. The elevation of Wall 1 follows a normal distribution with a mean of 100 and a variance of 5, the elevation of Wall 2 follows a normal distribution with a mean of 75 and a variance of 5, and the elevation of Wall 3 follows a normal distribution with a mean of 40 and a variance of 5.

Because the elevations are generated independently, there exists a possibility that the three surfaces would cross. It is checked that there is no crossing of these surfaces in the true data by assuring that the elevation of Wall 1 is larger than the elevation of Wall 2, and the elevation is larger than the elevation of Wall 3 at all locations.

After the simulation of the elevation values, the values of Thickness 12, Thickness 23, and Thickness 13 of the true dataset can be calculated from the simulated surfaces. The following equations show the calculation of thickness and the naming convention.

$$\text{Thickness } 12\,(u,v) = \text{Wall } 1\,(u,v) - \text{Wall } 2\,(u,v) \tag{4.1}$$

$$\text{Thickness } 23\,(u,v) = \text{Wall } 2\,(u,v) - \text{Wall } 3\,(u,v) \tag{4.2}$$

$$\text{Thickness } 13\,(u,v) = \text{Wall } 1\,(u,v) - \text{Wall } 3\,(u,v) \tag{4.3}$$

Thickness 13 can also be calculated by adding Thickness 12 and Thickness 23.

$$\text{Thickness } 13\,(u,v) = \text{Thickness } 12\,(u,v) + \text{Thickness } 23\,(u,v) \tag{4.4}$$

All thickness values are greater than zero. Thickness 12 follows a normal distribution with a mean of 25 $(100 - 75)$ and a variance of 7.07 $(\sqrt{5^2 + 5^2})$, Thickness 23 follows a normal distribution with a mean of 35 $(75 - 40)$ and a variance of 7.07 $(\sqrt{5^2 + 5^2})$, and Thickness 13 follows a normal distribution with a mean of 60 $(100-40)$ and a variance of 7.07 $(\sqrt{5^2 + 5^2})$. The distributions of the elevation and the thickness are used to build the cumulative distribution function (CDF) and inverse cumulative distribution function to conduct data transformation.

Figure 4.1 shows 2-D location maps of the simulated values in the true dataset. The elevations are shown on the left, and the thicknesses are shown on the right.

After the generation of the true dataset, it is possible to draw samples to conduct multiple imputation. A set of synthetic drill holes are drawn randomly to mimic a drill hole campaign by selecting 30 pairs of $U$ and $V$ coordinates at Wall 1 and Wall 3. It is assumed that all the drill holes are straight and these straight lines connect pairs of points at Wall 1 and Wall 3 to form drill holes samples. The coordinates of the pierce points between the drill holes and Wall 2 are the gridded coordinates that have the smallest distance between the drill holes and Wall 2.

### 4.2.2 Multi-Layered Deposits Example

The same workflow of multiple imputation is applied. The required prerequisites are the same, including angle and thickness calculation, variogram calculation and modeling, and normal score transform. The sample used for imputation can be derived by calculating angles and thickness values based on the angle tolerance. The angles of all the drill holes are calculated and the best angle tolerance is chosen to be 20 degrees using trial and error. The way of choosing the best angle tolerance is the same as documented in Section 3.2.1. The same workflow is applied with different angle tolerances, and the best angle tolerance is the one that has the highest correlation

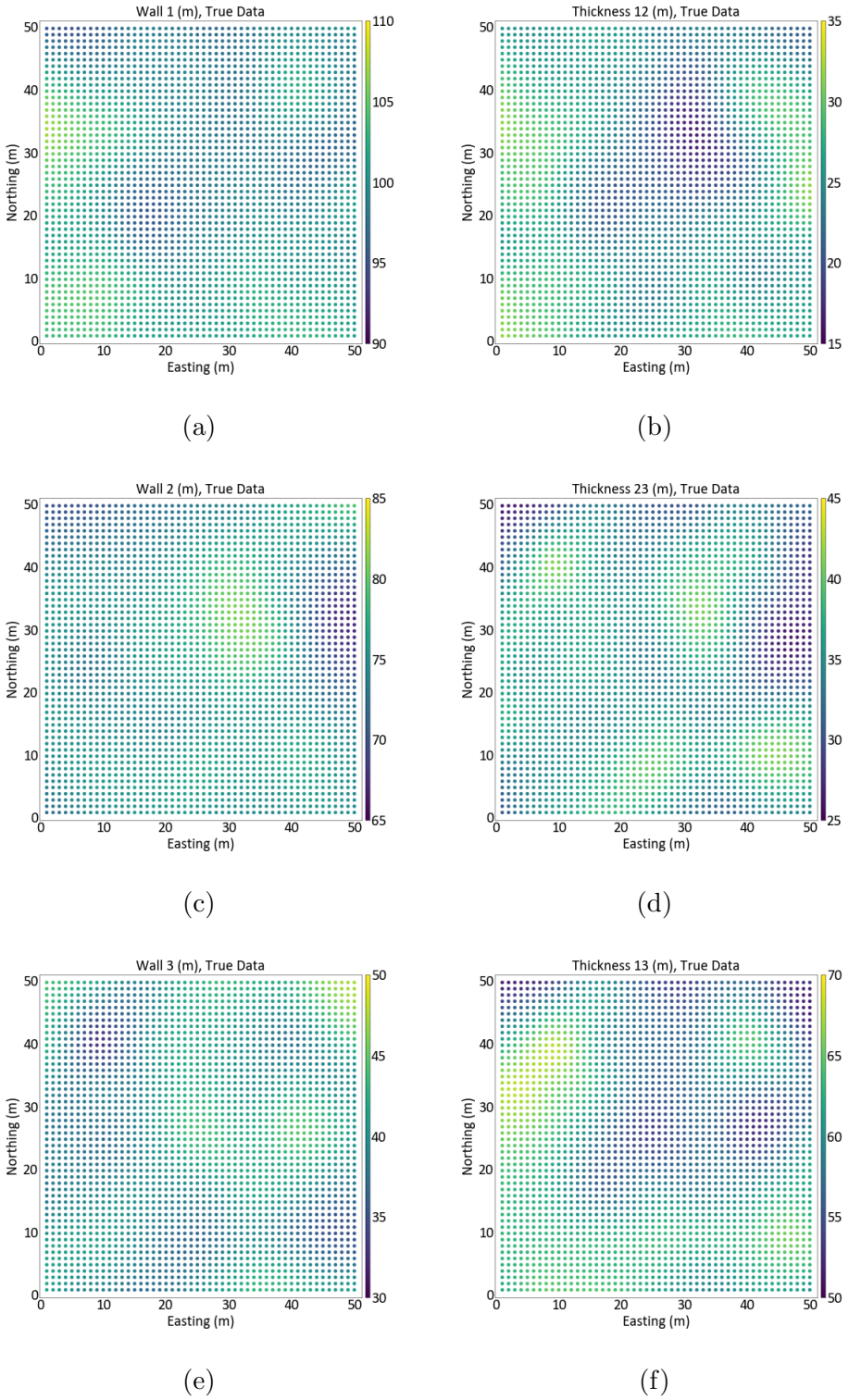



Figure 4.1: Location Map of the Simulated True Dataset for Multi-Layered Deposit

In (a), (c), (e), the location maps of the true data of elevation of Wall 1, Wall 2, and Wall 3 are shown. In (b), (d), (f), the location maps of the true data of Thickness 12, Thickness 23, and Thickness 13 are shown.

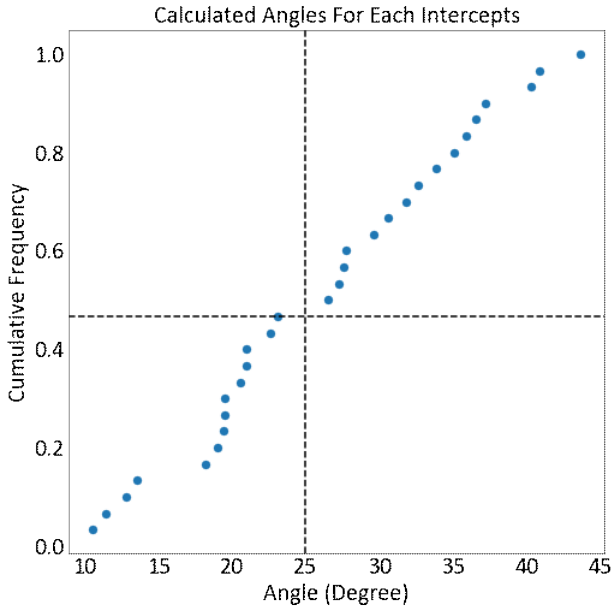



Figure 4.2: Cumulative Distribution of the Angles of the Drill Holes for Multi-Layered Deposit Example

The cumulative distribution function of the angles of the drill holes are plotted and an angle tolerance of 25 degrees is used.

and lowest RMSE. Any thickness value provided by drill holes that have angles that are larger than 25 degrees will not be used to form the distribution of the thickness. Figure 4.2 shows the calculated angles of the drill holes, and the dashed line indicates the angle tolerance of 25 degrees. Based on the angle tolerance, the thickness values chosen form the distribution of the thickness. Other drill holes are thus considered as inclined drill holes and will be imputed.

After the calculation of the angle tolerance and thickness, the geometric data samples used in multiple imputation can be formed. Figure 4.3 illustrates 2-D location maps of the geometric variables of the samples. There are 30 drill holes with 30 samples of elevation on each surface. However, there are only 14 thickness samples based on the drill hole angles and the angle tolerance. The number of the samples of the thickness values is always smaller than the number of the elevation values because inclined drill holes provide only elevation samples.

In this example, it is assumed that all the distributions and the variograms of the six geometric variables are available. The parametric cumulative distribution functions

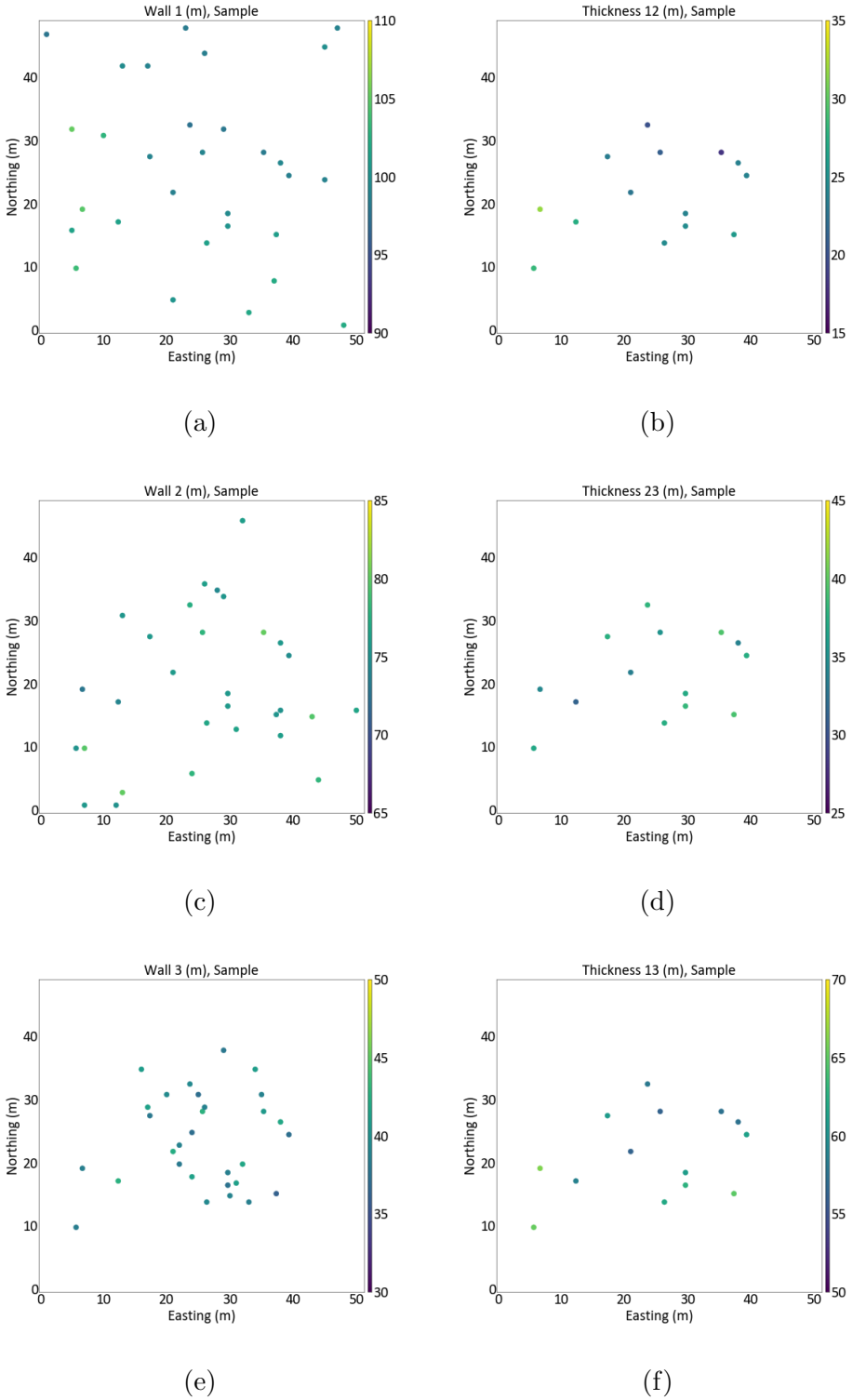



Figure 4.3: Location Maps of the Sample of Multi-Layered Deposit Example

In (a), (c), (e), the location map of the samples of elevation of Wall 1, elevation of Wall 2, and elevation of Wall 3 are shown. In (b), (d), (f), the location map of the samples of Thickness 12, Thickness 23, and Thickness 13 are shown.

and inverse cumulative distribution functions can be used to conduct transformations. The variograms are calculated and modeled from the true data set, which are shown in Figure 4.4. If the experimental variograms are calculated and modeled based on closely-gridded data, the experimental variograms will be almost identical to the true variograms (Bai and Deutsch, 2020a).

After these prerequisites, the process of imputation follows the basic methodology demonstrated in Chapters 2 and 3. When the imputation proceed to a location being imputed, local elevation and thickness distribution comes from surrounding samples by kriging. However, there are two collocated elevations can be used to transform the thickness distribution to the elevation distribution. For example, if the elevation of Wall 1 is being imputed, it is possible to get the secondary distribution by the following two ways.

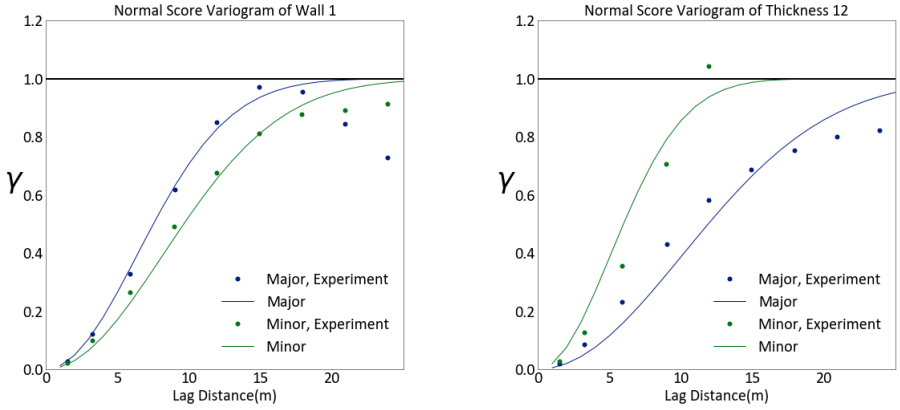
$$\text{Wall } 1 = \text{Wall } 2 + \text{Thickness } 12 \tag{4.5}$$

$$\text{Wall } 1 = \text{Wall } 3 + \text{Thickness } 13 \tag{4.6}$$

The two merging methods allow a merging of the spatial and collocated information afterwards and a realization is drawn from this merged distribution. The sequence of imputation starts from the most constrained location to the least constrained location. The discussion about the elevation values used and the sequence of imputation are in Section 4.3.

After imputation, every location will have a set of all six geometric variables containing elevation of Wall 1, Wall 2, Wall 3, Thickness 12, Thickness 23, and Thickness 13. The results of the imputed values are validated against the true values. Multiple imputation transfer the heterotopic geometric dataset to many homotopic geometric datasets.

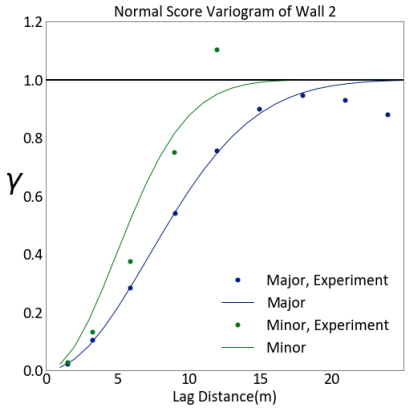
The realizations are compared with the true data. E-type values are considered with three summary statistics, which are error rate, Pearson correlation coefficient ($\rho$), and root mean square error (RMSE). These summary statistics are defined in Section 3.1.1. The results of the performance of multiple imputation for multi-layered deposits are shown in the following Table 4.1. The error rate and the RMSE of the



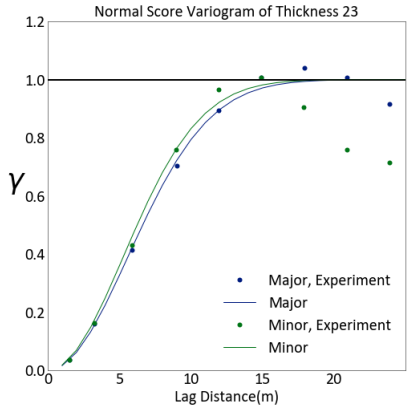

(a)



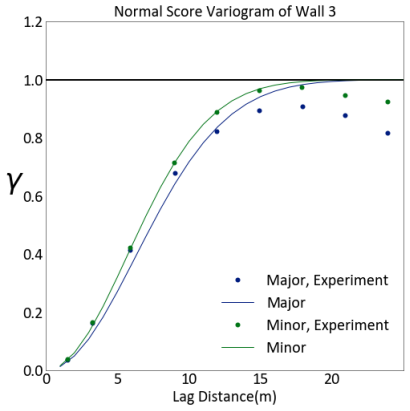

(b)



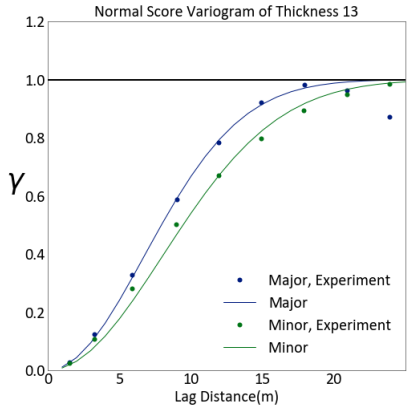

(c)




(d)




(e)




(f)

Figure 4.4: Modeled Variograms for Multi-Layered Deposit Example

In (a), (c), (e), the calculated and modeled variogram of the elevations are shown. In (b), (d), (f), the calculated and modeled variograms of the thicknesses are shown.

Table 4.1: Summary Statistics for Multi-Layered Deposit Example

| Variable | Merging Method | Error Rate | Correlation | RMSE |
|---|---|---|---|---|
| Wall 1 | EE | **0.654** | **0.927** | **0.848** |
| Wall 1 | BU | 0.711 | 0.916 | 0.858 |
| Wall 2 | EE | **1.066** | **0.887** | **0.979** |
| Wall 2 | BU | 1.181 | 0.855 | 1.101 |
| Wall 3 | EE | **3.011** | **0.794** | **1.374** |
| Wall 3 | BU | 3.222 | 0.727 | 1.500 |
| TH 12 | EE | **3.474** | **0.933** | **1.149** |
| TH 12 | BU | 3.679 | 0.927 | 1.214 |
| TH 23 | EE | **4.401** | **0.889** | **1.438** |
| TH 23 | BU | 4.491 | 0.867 | 1.517 |
| TH 13 | EE | **2.001** | **0.941** | **1.925** |
| TH 13 | BU | 2.028 | 0.924 | 1.963 |

elevation are lower than the error rate and the RMSE of the thickness, but the correlation of the elevation are lower. The correlation of the result is around 0.85, which is lower than the result of single-layered deposit (around 0.9). The RMSE is around 1.5, which is higher than the result of single-layered deposit (around 1.0). The overall performance of the method applied to multi-layered deposit is weaker. The better results of the two merging methods are bolded in the table. Here, the method of error ellipses outperforms the method of Bayesian updating consistently and slightly.

Accuracy and bias are also compared in Figure 4.5. In these scatter plots, e-type estimates are plotted on the x-axis against the true values on the y-axis. The overall performance has a correlation of 0.85, showing a strong correlation. The regression line illustrates some conditional biases. Considering the e-type tends to reduce the variance of the geometric variables; in other words, it tends to overestimate low values and underestimate high values. This phenomenon is most obvious for Thickness 13, considering the realizations mitigates this.

The two merging methods produce similar results, as they usually appear as close pairs in the scatter plots. But the difference is bigger than observed in single-layered deposits. From the summary statistics, the method of error ellipses performs better.

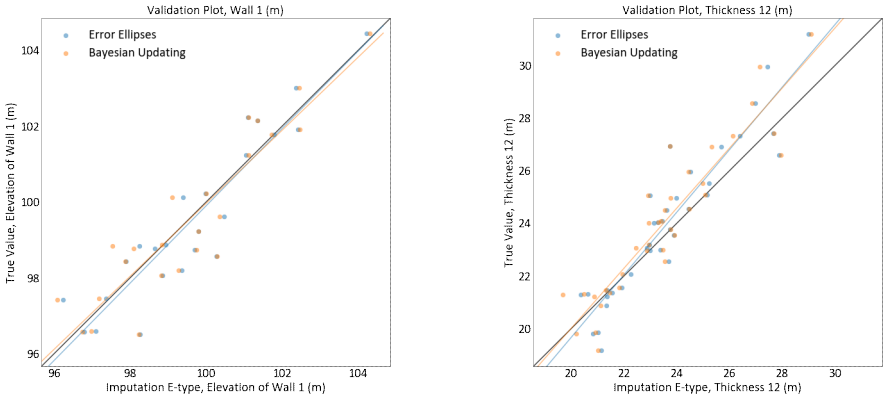
Reproduction of the CDF of the imputed realization is shown in Figure 4.6. As the method of error ellipses performance better, the results of it are shown. The distribution of realizations are plotted in gray. The true distribution in red and the sample distribution in blue are plotted for comparison. The method reproduces the distribution approximately, but it reduces the variances for all geometric variables slightly as the imputed distributions only overlap with the true distributions in the middle. Some reproductions reduce the variance greatly, as shown in Figure 4.6d and in Figure 4.6e.

Reproduction of variograms of the imputed realization are displayed against the true variograms in Figure 4.7. The true variograms are plotted in red and the imputed variograms are plotted in green and blue. The method reproduces the variograms overall, but it seems that it does not reproduce the anisotropy very well again, especially in Figure 4.7b.

The overall performance of the method of multiple imputation applied to multi-layered deposits appears reasonable. The correlation is acceptable (around 0.85) and the RMSE is around 1.5, and the reproduction of the distribution follows the sampled data. But the variance is lower and the variogram reproduction is not very good.

## 4.3 Sensitivity Analysis of Multi-Layered Deposits

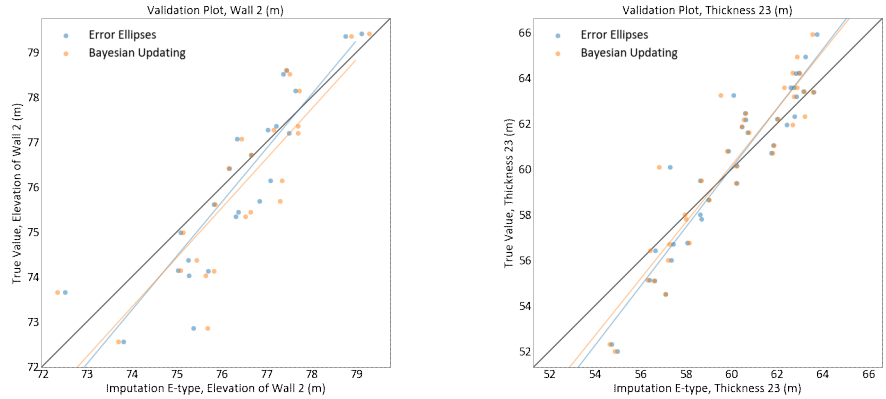
There are many decisions and parameters that can influence the performance of the method of multiple imputation. In Chapter 3, some of them, like angle tolerance and transformation method, are discussed. Sensitivity analysis shows that these parameters have the same influence on multiple imputation applied to multi-layered deposits as the method applied to single-layered deposits discussed in Chapter 3. However, the merging method of error ellipses performs consistently better than Bayesian updating for multi-layered deposits probably because the scenarios are more complicated, and the method of error ellipses is more stable. In this section, a sensitivity analysis is used to find the best sequence to conduct imputation and the best opposite elevation




(a)



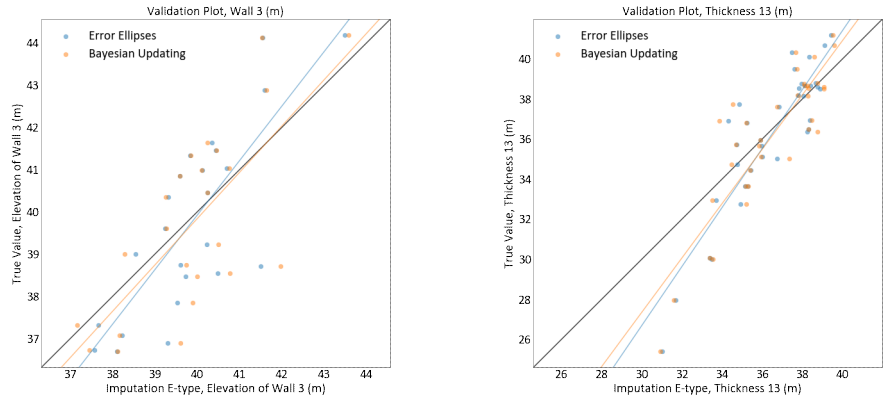

(b)




(c)




(d)



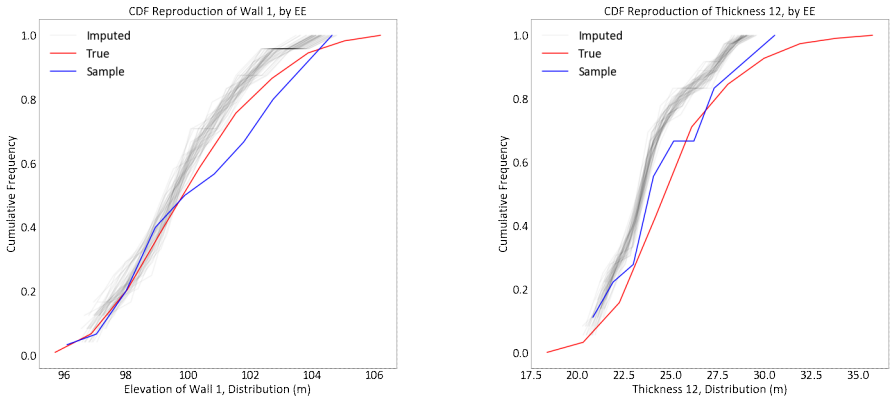

(e)




(f)

Figure 4.5: Validation Plot for Multi-Layered Deposit Example

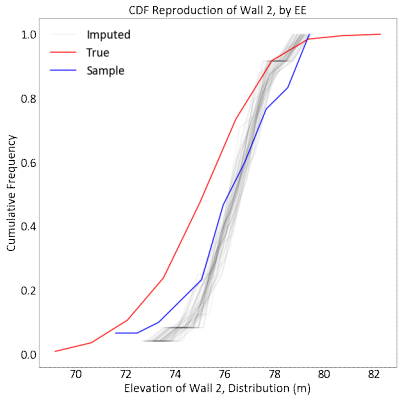
In (a), (c), (e), the imputed values of the elevations are plotted against the true values. In (b), (d), (f), the imputed values of the thicknesses are plotted against the true values.



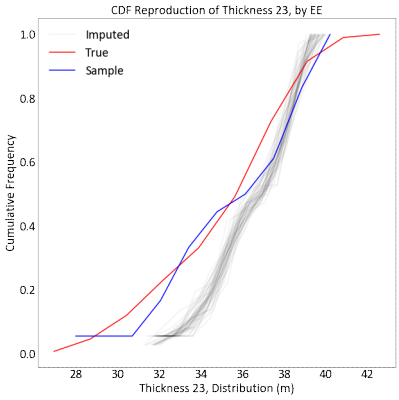

(a)



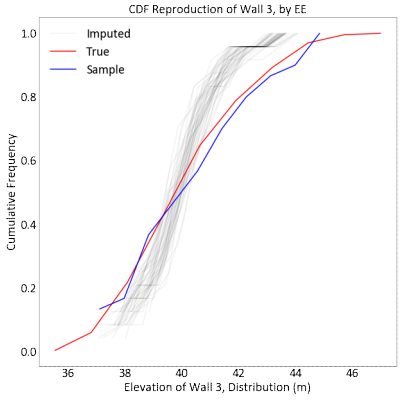

(b)



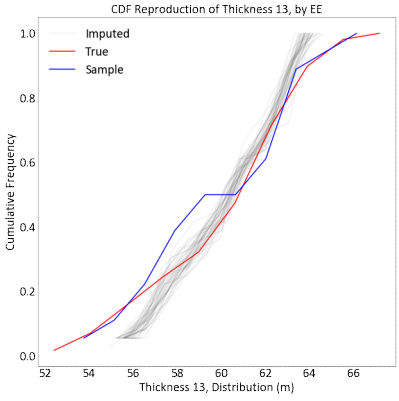

(c)




(d)




(e)




(f)

Figure 4.6: Distribution Reproduction of Multi-Layered Deposit Example

In (a), (c), (e), the reproduction of distribution of the imputed realizations of the elevations are plotted. In plot (b), (d), (f), the reproduction of distribution of the imputed realizations of the thicknesses are plotted.

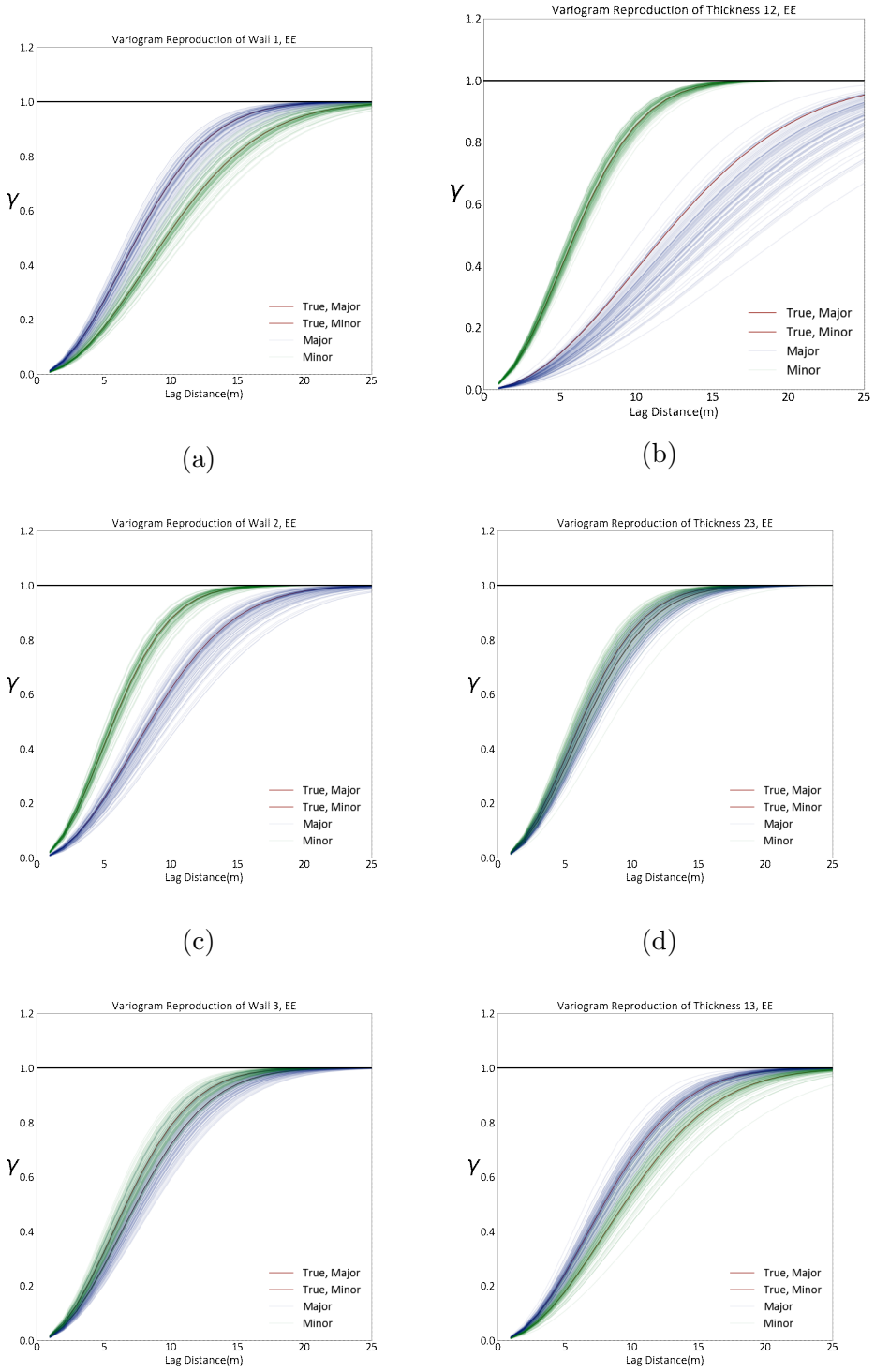



Figure 4.7: Variogram Reproduction of Multi-Layered Deposit

In (a), (c), (e), the variogram reproduction of elevation are shown. In (b), (d), (f), the variograms reproduction of thicknesses are shown.

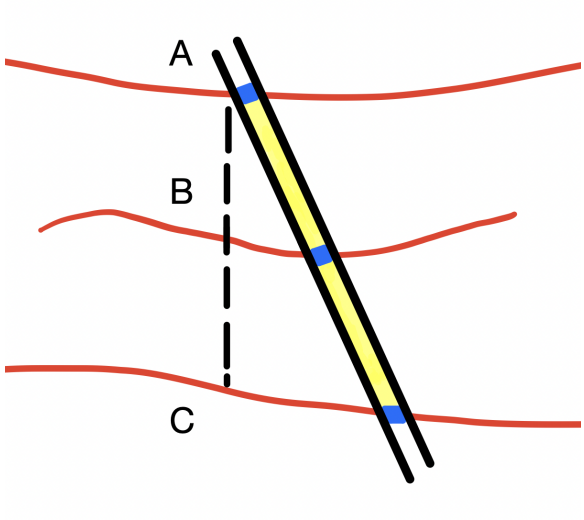



Figure 4.8: Schematic Illustration of Difference Elevations

A schematic illustration shows three elevations. If C is being imputed, existed pierce point A or previously imputed B can be both used as an elevation value. (no to scale)

to use to transform thickness distributions into elevation distributions.

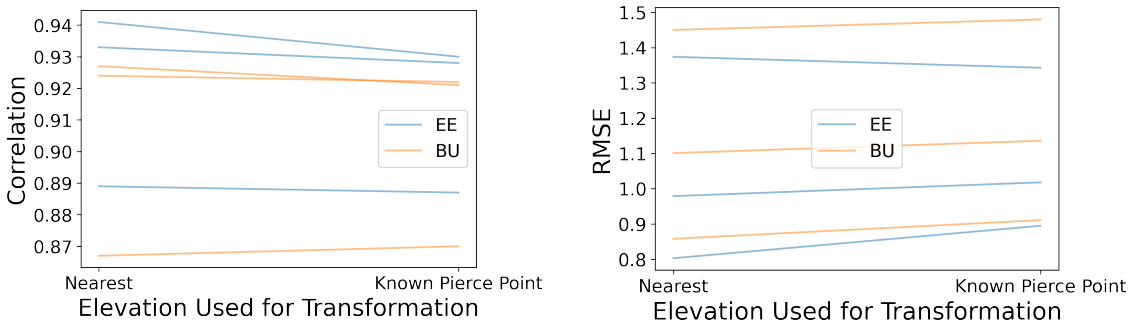
### 4.3.1 Elevation Value Used

The first parameter is which elevation value should be used when transforming thickness into elevation. During imputation, to add to or subtract from the simulated conditional distribution of thickness in order to merge this thickness distribution and elevation distribution.

During the transformation of multi-layered deposits, many elevation values can be used. These choices are demonstrated by a schematic illustration. It is assumed that location C in being imputed in Figure 4.8. It is possible to use the elevation value at pierce point A, or to use the previously imputed elevation value at location B. The first choice (Known Pierce Point) uses the observed pierce point of the drill hole, which is the most certain elevation value can be used of this location. In this example, A is the observed pierce point. The second choice (Nearest) uses the nearest opposite surface which has the smallest difference of elevation. Here the elevation of B is closer to the elevation of C than the elevation of A. This opposite elevation of B can be derived derived from previously imputed values.

In Figure 4.9, the result of the sensitivity analysis are shown by comparing corre-

lation and RMSE. Using the nearest elevation is a better choice because it has higher correlation and lower RMSE.



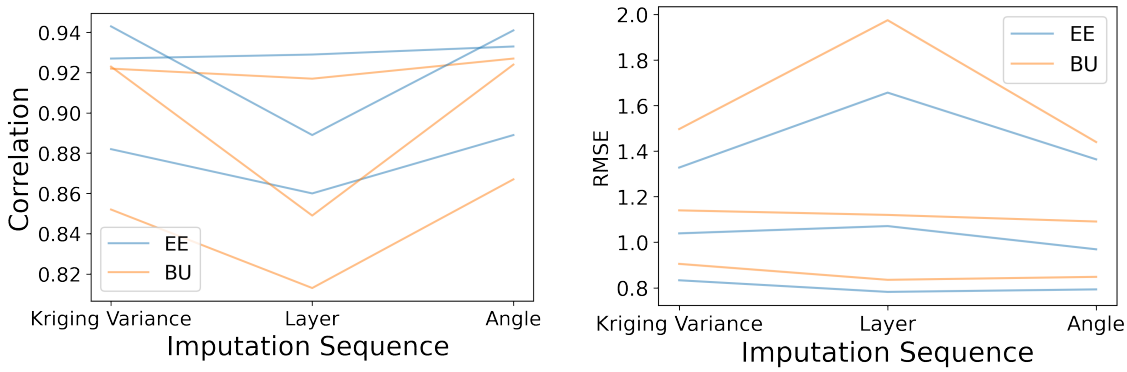

(a) (b)

Figure 4.9: Correlation and RMSE of Different Elevation Used

In (a), the correlation of the imputed values with the true values with different elevation used during transformation are shown. In (b), the RMSE are shown. The method of using nearest elevation value is better.

### 4.3.2 Imputation Sequence

Another parameter analyzed is the sequence of imputation. A schematic example in Section 3.2.2 demonstrates the three imputation sequences considered. The first sequence (Kriging Variance) is to use global kriging to decide the kriging variance and start from the lowest kriging variance to the highest variance. The second sequence (Layer) is to start from the most stable elevation and use global kriging to decide the sequence at this elevation and go to the next elevation afterwards. The third sequence (Angle) is to start from the least inclined drill hole to the most inclined drill hole (Bai and Deutsch, 2021b). These three sequences are compared for multi-layered deposits. The results are shown in Figure 4.10 and show that the best sequence is to use the angle of the drill holes because it produces the highest correlation and lowest RMSE. The optimum sequence is the same as the one for single-layered deposits. Furthermore, for one inclined drill hole, the sequence of its different imputation locations has no significant influence on the performance, but it is still better to use the kriging variance to proceed the imputation for one drill hole.




(a) (b)

Figure 4.10: Correlation and RMSE of Different Imputation Sequence

In (a), the correlation of the imputed values with the true values with different imputation sequence are shown. In (b), the RMSE are shown. The method of using angles to determine the sequence is the best.

As a conclusion, it is better to use the nearest known elevation value to transform thickness distribution into elevation distribution and the best imputation sequence is to start from the most vertical drill hole to the most inclined drill hole. Influences of other parameters remain the same as in Section 3.2, and the method of error ellipses performs better.

## 4.4 Discussion on Multi-Layered Deposits

Multiple imputation applied to multi-layered deposits is similar to single-layered deposits. Furthermore, the performance of the method and the results from the sensitivity analysis remain quite similar. One difference is that the method of error ellipses is consistently better than the method of Bayesian updating. However, there are some extra limitations about multiple imputation that is applied to multi-layered deposits because of the increasing number of surfaces.

The first limitation is that the number of the locations that are required to be imputed increases quadratically with the number of the layers of the deposits. For example, there are 2 locations that are required to be imputed for an inclined drill hole at a single-layered deposit. However, there are 6 locations that are required to be imputed for an inclined drill hole at a double-layered deposit. The relationship between the number of the positions that are required to be imputed and the number

of the layers of the deposits follows the following equation:

$$N = n^2 + n \tag{4.7}$$

where $N$ is the number of the locations that are required to be imputed, and $n$ is the number of the layers. As a result, it takes longer to impute geometric data from multi-layered deposits.

A second problem is that as the number of the layers and the number of the locations that are required to be imputed increase, the imputation process may need to change. The situation is shown in a schematic illustration of configuration of drill holes in Figure 4.11. If there are 2 inclined drill holes, there should be 4 locations that are necessary to be imputed. However, there exists the possibility that the location at the footwall that is required to be imputed coincides with the known footwall location provided by other inclined drill hole. As a result, although there are 2 inclined drill holes, there is no need to conduct imputation in this situation. For example, in theory, it is necessary to impute inclined drill hole 1 as well as inclined drill hole 2, however, the geometric variables do not need to be imputed because location A and location B are already highly constrained. Another more common example is that for inclined drill hole 3 and inclined drill hole 4, it is necessary to imputed at location C and location E and it is not necessary to impute at location D because the inclined drill holes here do provide both the elevation and the thickness.

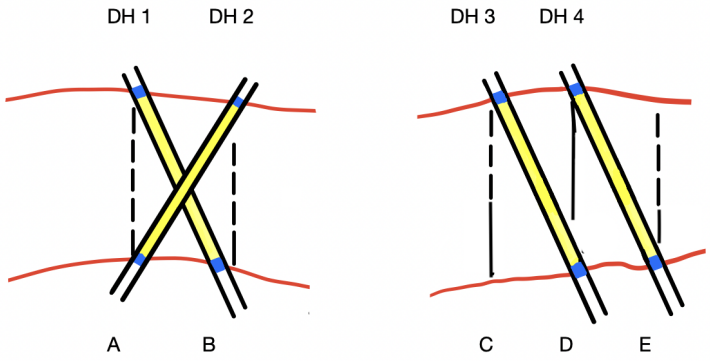
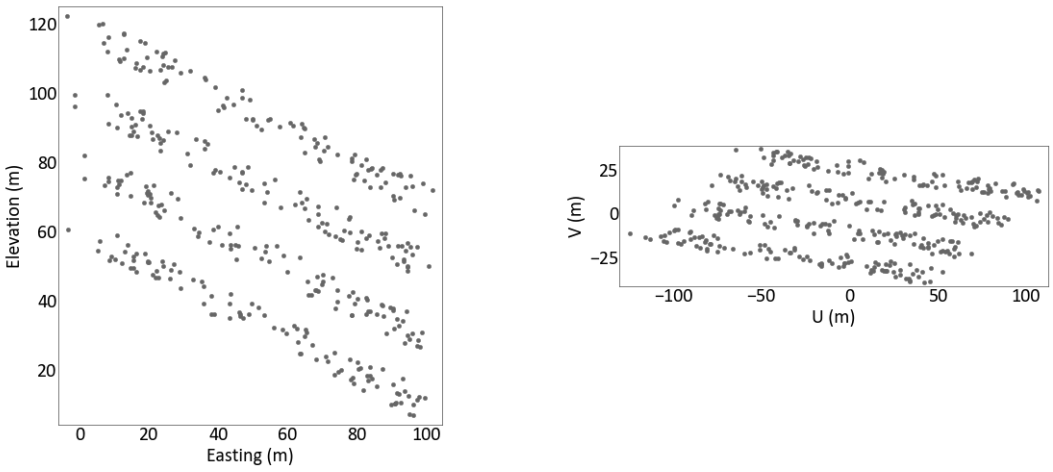



Figure 4.11: Configuration of Situation Changing the Imputation Process

Two situations that can change the process of imputation are plotted. Location A, B, and D are already highly constrained. There is no need to impute at these locations. (no to scale)

If the imputation still proceeds at the inclined drill holes here, the covariance matrices can take both the imputed value and the true value at the same time. This leads to singular covaraince matrices that can not be used to form kriging systems. This problem occurs more often when multiple imputation is applied to multi-layered deposits. This problem can be fully solved by modifying the method to skip this kind of highly constrained locations in order to proceed multiple imputation smoothly. Nevertheless, it is important to avoid having singular matrices during imputation.

A third limitation is that TLS can be less efficient when there are many surfaces (Bai and Deutsch, 2021a). As shown in Figure 2.1, TLS performs quite well when there are only two surfaces. In Figure 4.12, TLS is applied to inclined surfaces in order to build a local coordinates system in order to derive the thickness distribution. However, even after applying TLS, the thickness values between different surfaces are still difficult to calculate sometimes.




(a) (b)

Figure 4.12: TLS for Multi-Layered Deposits

The coordinates of the pierce points of the drill holes with inclined surfaces are shown in (a). After TLS, the coordinates are shown in (b). It is still hard to calculate the thickness here.

This problem can be solved by fitting each plane individually by TLS. Figure 4.13 illustrates the result of fitting by the new method, which allows a better calculation of thickness values.

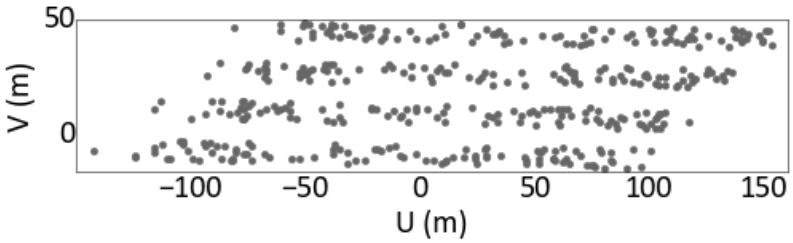



Figure 4.13: TLS Fitted to Each Plane for Multi-Layered Deposits

After the new fitting by TLS, the coordinates are shown. It is easier to calculate the thickness.

One final discussion is about the two different types of continuity. Different geometries of the tabular deposits can influence the performance of multiple imputation. There are two types of continuities discussed for simple geometries of tabular vein deposits (Carvalho, 2018). In Figure 4.14, these two types of continuities are shown. In A, the elevation of the footwall is more stable than the thickness. In B, the thickness is more stable than the elevation of the footwall.

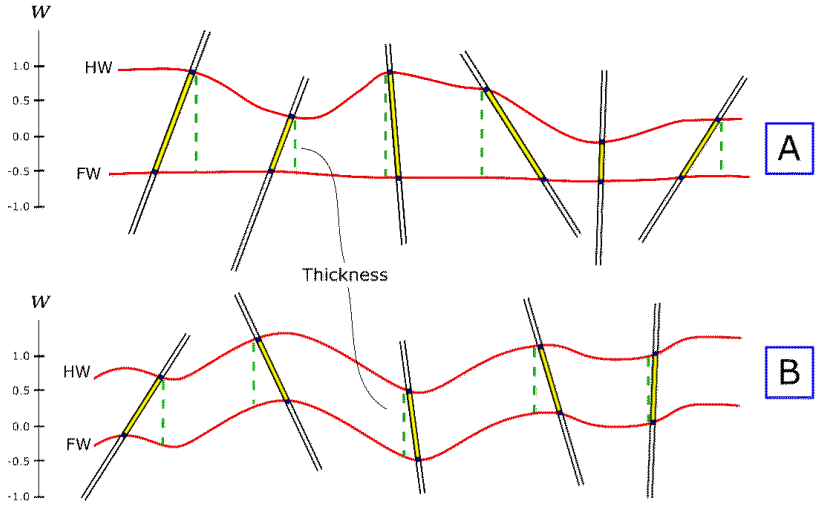



Figure 4.14: Two Continuities

In A, the elevation of the footwall is more stable than the thickness. In B, the thickness is more stable than the elevation (Carvalho, 2018). (no to scale)

It is found that multiple imputation performs better in case B when the thickness is more stable. It is probably because thickness distribution is calculated from the

samples, and if it is possible to have a representative and stable thickness distribution, the higher uncertainty of thickness can be reduced.

It is important to consider these limitations during multiple imputation for multi-layered deposits. It is not necessary to impute at locations that are already highly constrained even there are inclined drill holes there.

# Chapter 5

# Imputation Along Long Deviated Drill Holes

*The basic method of multiple imputation starts from imputation at pierce points that can provide geometric variables directly. Yet the internal data along drill holes provides additional information, especially when drill holes are highly deviated. There are constraints along the drill holes that the surfaces of the deposits cannot cross the drill hole intersection in the vein. It is necessary to impute along drill holes to integrate this internal data. Assumptions about the mean and the variance of the distribution of thickness as well as the opposite elevation are made. A 2D schematic example demonstrates the method and a 3D synthetic example is presented. Sensitivity analysis shows that imputation along drill holes has a lower but acceptable accuracy and it is better to impute at pierce points at first.*

## 5.1 Assumptions for Imputation Along Drill Holes

Geometric data including elevation and thickness are provided by the pierce points between drill holes and surfaces defining the deposits. The method of multiple imputation in Chapters 2, 3, and 4 quantifies geometric uncertainty near these pierce points. Highly deviated drill holes show increased geometric uncertainty along long deviated drill holes. In Figure 5.1, a schematic drawing shows a configuration of drill holes. It is possible that the drill hole stays inside the vein between the hangingwall and the footwall intersections (in red) or outside the layers (in orange). Furthermore, for long drill holes, there are few drill holes that can provide geometric information about thickness. Although the data along these drill holes can not provide geometric data like pierce points, this information is important and can be used for modeling as well.

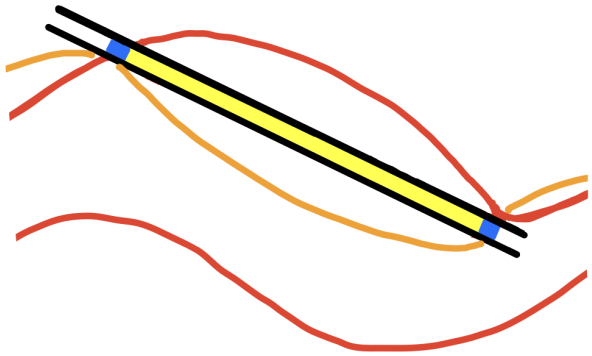


Figure 5.1: Configuration of Drill Holes for Long Deviated Drill Holes

A schematic drawing shows that there can be many possible configurations along drill holes with the same pierce points. (not to scale)

During the workflow of multiple imputation, the secondary thickness distribution is transformed into primary elevation distribution using an opposite elevation value. However, the geometric data provided along the drill holes have neither a direct thickness distribution nor an opposite elevation value for transformation. As a result, some assumptions about the elevation used during transformation and the distribution of thickness should be made in order to extend the basic method for geometric data imputation along long drill holes.

Several coordinate points on the projection of the drill hole onto the plane of reference are added to the geometric dataset, and geometric data of these points will be imputed in the workflow. The assumptions of the relative elevation and the relative thickness are based on the coordinates of these coordinate points along drill holes that depends on the trajectory along the length of the drill holes.

The first assumption is about the opposite elevation. It is necessary to calculate the opposite elevation value. We assume that the drill holes are straight and the $W$ can be calculated by the following equation if the $U$ and $V$ coordinates are known.

$$\frac{U - U_{FW}}{U_{HW} - U_{FW}} = \frac{V - V_{FW}}{V_{HW} - V_{FW}} = \frac{W - W_{FW}}{W_{HW} - W_{FW}} \tag{5.1}$$

where $U_{HW}$, $U_{FW}$, $V_{HW}$, $V_{FW}$, $W_{HW}$, and $W_{HW}$ are the $U$, $V$, and $W$ coordinates of the intercepts of the hangingwall and the footwall of the drill hole. $U$ and $V$ are the coordinates along drill holes under imputation, and they must be along the projection

of the drill hole onto the plane of continuity. After the derivation of this $W$ value, the thickness distribution can be added to or subtract from it. Furthermore, these fractions $\frac{U-U_{FW}}{U_{HW}-U_{FW}}$, $\frac{V-V_{FW}}{V_{HW}-V_{FW}}$, or $\frac{W-W_{FW}}{W_{HW}-W_{FW}}$ are used in the second assumption to construct local thickness distribution.

The second assumption is about the distribution of thickness. The primary distribution of elevation in multiple imputation comes from kriging and it can be derived using the same method in Chapters 2, 3, and 4. The secondary distribution in multiple imputation comes from the collocated thickness distribution by kriging as well. However, the thickness distribution provided by kriging is the conditional distribution of the thickness from the hangingwall to the footwall. There is no direct thickness distribution along drill holes. In order to build a conditional distribution of thickness along drill holes, assumptions about the mean and the variance of the distribution are made. The assumptions are listed in the following formulae.

$$\overline{TH_{Total}} = \overline{TH_{Above}} + \overline{TH_{Below}} \tag{5.2}$$

$$\overline{TH_{Above}} = F \times \overline{TH_{Total}} \tag{5.3}$$

$$\overline{TH_{Below}} = (1 - F) \times \overline{TH_{Total}} \tag{5.4}$$

$$\sigma^2_{TH_{Total}} = F \times \sigma^2_{TH_{Above}} + (1 - F) \times \sigma^2_{TH_{Below}} \tag{5.5}$$

where $\overline{TH_{Total}}$ and $\sigma^2_{TH_{Total}}$ are the conditional mean and the conditional variance derived by kriging and back-transform, $\overline{TH_{Above}}$ and $\sigma^2_{TH_{Above}}$ are the conditional mean and the conditional variance of the thickness above the drill hole to the hangingwall, and $\overline{TH_{Below}}$ and $\sigma^2_{TH_{Below}}$ are the conditional mean and the conditional variance of the thickness below the drill hole to the footwall. $F$ is the fraction of the thickness above the drill hole over the total thickness, and it can be $\frac{U-U_{FW}}{U_{HW}-U_{FW}}$, $\frac{V-V_{FW}}{V_{HW}-V_{FW}}$, or $\frac{W-W_{FW}}{W_{HW}-W_{FW}}$.

After making these two assumptions, it is possible to use multiple imputation to quantify geometric uncertainty along drill holes. It is possible to construct a conditional distribution of thickness and an opposite elevation to transform this thickness

distribution into elevation distribution. Other steps remain the same as the basic method of multiple imputation.

## 5.2 Schematic Example of Along Drill Holes

The 2D schematic example shown at the end of Chapter 2 is used in this section to demonstrate multiple imputation along drill holes. In Table 5.1, one realization of geometric data after multiple imputation at pierce points is shown. Figure 5.2 shows a location map of the drill holes and the imputed values. The demonstration of imputation along drill holes is based on this dataset. The parameters of the data are listed in Chapter 2.

Table 5.1: Data for Schematic Example Along Drill Holes

| DH | Horizontal Coordinate | Hangingwall | Footwall | Thickness |
|---|---|---|---|---|
| 1 | 81 | 34 | 8 | 26 |
| 2 | 163 | 31 | 16.341 | 14.659 |
| 2 | 222 | 29.078 | 18 | 11.078 |

For the schematic example, it is assumed that multiple imputation proceeds to the hangingwall location at 200 along the drill hole 2. 200 is between 163 and 222. The conditional distribution of the elevation of the hangingwall at 200 can be formed by kriging using the data at 81, 163, and 222. The covariance matrix based on the distance can be derived from the variogram model. The kriging equations can be constructed and solved. The mean and the variance of the local distribution can be derived.

$$\begin{bmatrix} 1 & 1-\gamma(163-81) & 1-\gamma(222-81) \\ 1-\gamma(163-81) & 1 & 1-\gamma(222-163) \\ 1-\gamma(222-81) & 1-\gamma(222-163) & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} 1-\gamma(200-81) \\ 1-\gamma(200-163) \\ 1-\gamma(222-200) \end{bmatrix} \tag{5.6}$$

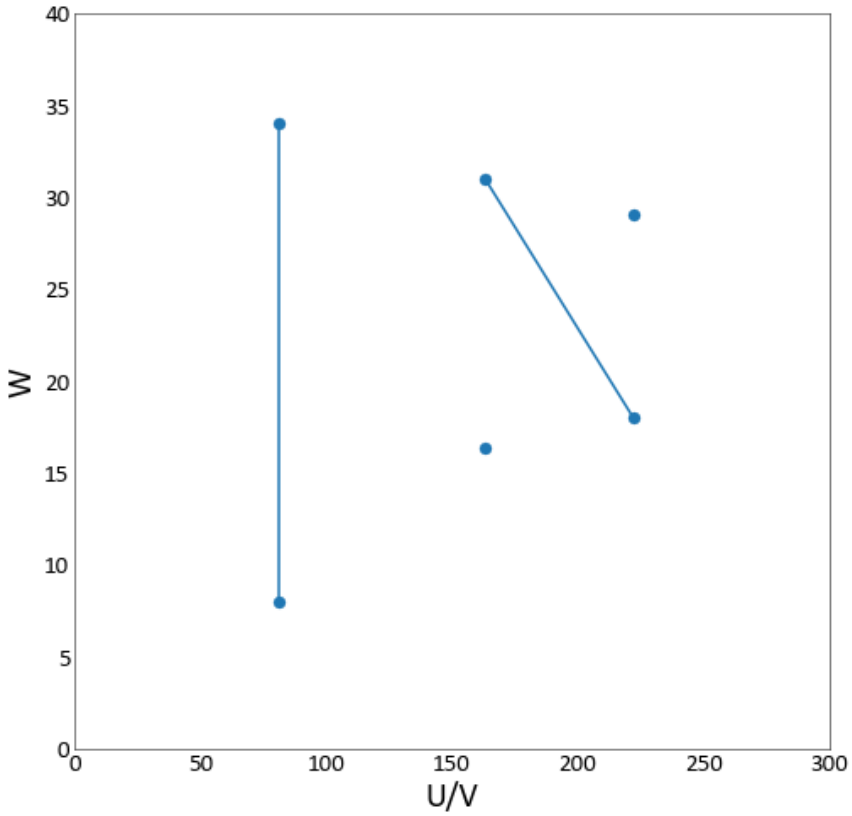



Figure 5.2: Location Plot of Schematic Example of Imputation Along Drill Hole

Data for schematic example is shown. The two individual points come from previous imputation. (no to scale)

$$\mathbf{C} = \begin{bmatrix} 1 & 0.4021 & 0.7771 \\ 0.4021 & 1 & 0.2375 \\ 0.7771 & 0.2375 & 1 \end{bmatrix} \mathbf{D} = \begin{bmatrix} 0.3423 \\ 0.8934 \\ 0.9547 \end{bmatrix} \tag{5.7}$$

$$\lambda = \mathbf{C}^{-1}\mathbf{D} = \begin{bmatrix} -0.7773 \\ 0.4742 \\ 0.6104 \end{bmatrix} \tag{5.8}$$

$$y_{\bar{W}} = -0.7773 \times 0.9153 + 0.4742 \times 0.2404 + 0.6104 \times -0.2223 = -0.0923 \tag{5.9}$$

$$\sigma^2_{y_W} = 1 - -0.7773 \times 0.3423 - 0.4742 \times 0.8934 - 0.6104 \times 0.9547 = 0.0201 \tag{5.10}$$

After the calculation, the primary mean of the elevation is $-0.0923$, and the primary variance of the elevation is 0.0201. The conditional distribution of the total thickness at 200 can be formed by kriging using the same covariance matrix. Consequently, the same kriging weights can be applied.

$$\bar{y_{Total}} = -0.7773 \times 0.9945 + 0.4742 \times -0.2889 + 0.6104 \times -0.6963 = -0.6388 \quad (5.11)$$

$$\sigma^2_{y_{Total}} = \sigma^2_{y_W} = 0.0201 \quad (5.12)$$

The conditional variance with a value of 0.0201 is the same as the variance of the elevation, but the conditional mean with a value of $-0.6388$ is different from the mean of the elevation. This distribution quantifies the distribution of the total thickness, and it is necessary to transform it into the distribution of the thickness above the drill hole. Before transforming this total thickness distribution into elevation distribution, the two stated assumptions in Section 5.1 are made. The intercept of the drill hole with the hangingwall is at $(163, 31)$, and the intercept of the drill hole with the footwall is at $(222, 18)$. As a result, the elevation along the drill hole at 200 can be calculated by the following equations.

$$\frac{200 - 163}{222 - 163} = \frac{W - 31}{18 - 31} \quad (5.13)$$

$$W = 22.847 \quad (5.14)$$

The calculated $W$ value serves as the value to which can be used to perform addition or subtraction. The conditional mean and the conditional variance of thickness above the drill hole at 200 is estimated by proportion. The calculation of the fraction is shown in the following equation.

$$\frac{31 - W}{31 - 18} = 0.627 \quad (5.15)$$

After calculating this fraction, it is possible to get the conditional distribution of the thickness above the drill hole at 200 to the hangingwall. After making this two assumptions about the opposite elevation and the thickness distribution, the

distribution of the total thickness can be transformed into the thickness above the drill hole at first. Then, the distribution of the thickness above the drill hole can be transformed into the distribution of the elevation of the hangingwall. The steps of the transformation are shown below. The difference is how to use the assumption to transform the distribution of the total thickness to the thickness above the drill hole.

1. Generate original unit values of the total thickness based on the conditional distribution $(-0.6388, 0.0201)$.

$$TH = F_{TH}^{-1}\left(G\left(y_{TH}\right)\right)$$

$$\begin{bmatrix} 10.527 & 12.263 & \cdots & 10.947 & 11.392 \end{bmatrix}$$

2. Calculate the mean and the variance of original unit values of the total thickness.

$$\overline{TH_{Total}} = 11.630$$

$$\sigma^2_{TH_{Total}} = 1.796$$

3. Apply the fraction 0.637 to the mean and the variance of original unit values of the total thickness and get the mean and the variance of original unit values of the thickness above the drill holes.

$$\overline{TH_{Above}} = 0.627 \times 11.630 = 7.294$$

$$\sigma^2_{TH_{Above}} = 0.627 \times 1.796 = 1.126$$

4. Generate original unit values of the thickness above the drill holes based on $(7.294, 1.126)$.

$$\begin{bmatrix} 8.491 & 6.370 & \cdots & 7.223 & 6.414 \end{bmatrix}$$

5. Calculate the hangingwall based on the thickness above the drill holes. The elevation at 200 is 22.847.

$$W_{HW} = 22.847 + TH_{Above}$$

$$\begin{bmatrix} 31.339 & 29.217 & \cdots & 30.070 & 29.261 \end{bmatrix}$$

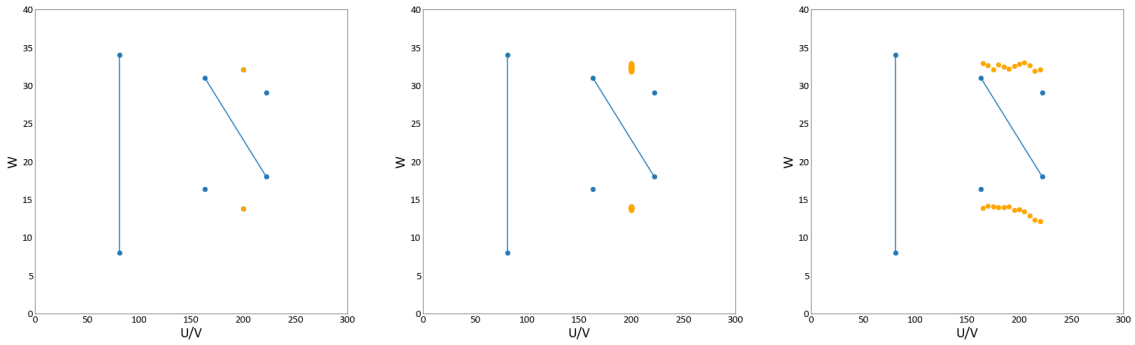
6. Normal score transform the elevation of the hangingwall.

$$y_{W_{HW}} = G^{-1}\left(F_{W_{HW}}\left(W_{HW}\right)\right)$$

$$\begin{bmatrix} 0.318 & -0.190 & \cdots & 0.018 & -0.179 \end{bmatrix}$$

7. Calculate the mean and the variance of the new distribution, which is 0.0201 and 0.0697.

After the derivation of the secondary distribution of the elevation of the hangingwall from thickness, it is possible to merge it with the primary distribution of the elevation of the hangingwall, which is calculated at the beginning. Values can be drawn from the merged distribution. Furthermore, it is possible to impute the elevation of the footwall at 200 using the same method. In Figure 5.3a, one realization of imputation along the drill hole is shown in orange.




(a) (b) (c)

Figure 5.3: Imputation results of the schematic example of along drill holes

In (a), one realization of imputation at 200 is shown. In (b), 100 realizations of imputation at 200 is shown. In (c), one realization of imputation along drill holes is shown.

In Figure 5.3b, 100 simulations of imputation are shown in orange. Furthermore, using different fractions and different elevation values, it is possible to proceed imputation along drill holes as long as the $U$, and $V$ coordinate is in the projection of the drill hole. In Figure 5.3c, one realization of imputation along the drill hole is shown in orange with an interval of 5.

## 5.3 Demonstration of Imputation Along Long Deviated Drill Holes

The dataset used in Chapter 3 is used in this section to demonstrate the method. The same workflow of imputation applied in Chapter 2, 3, and 4 can be applied to the pierce points.

For imputation along long drill holes, one more prerequisite for the method is to decide what set of points between pierce points should be imputed. After the calculation of angles of the drill holes, a set of coordinates of points within long deviated drill holes are added to the raw data and they do not provide any new geometric information. These coordinates points must be points along the projection of the drill holes onto the 2D surface of the deposit. The adding of coordinates points will be discussed in Section 5.4.1. Other prerequisites are the same. The geometric variables should be normal score transformed and the variograms should be calculated and modeled.

The same method can be applied at the pierce points. For the points along long drill holes, the similarity is that the conditional distribution of the elevation is formed by kriging, but the conditional distribution of the thickness is characterized by a proportional mean and variance and the opposite elevation is the coordinates points of the drill holes. Then, this thickness distribution is transformed into elevation distribution and merged with the primary elevation distribution. The sequence of the imputation will be discussed in Section 5.4.2.

After imputation, every pierce points will have a set of geometric variables including elevation of the hangingwall, elevation of the footwall, and thickness. Furthermore, there will be these geometric variables along long drill holes.

The realizations are compared with the true data. E-type statistics are compared with three summary statistics, error rate, correlation, and RMSE. These summary statistics are defined in Section 3.1.1. The results of the performance of imputation along long drill holes are shown in Table 5.2. The imputation of the elevation looks reasonable. The correlation of the elevation is around 0.9 and RMSE is around 1.7. However, the imputation of the thickness is not very good. The correlation is around 0.75. The best results of the two merging methods are bolded in the table. The two

methods perform similarly.

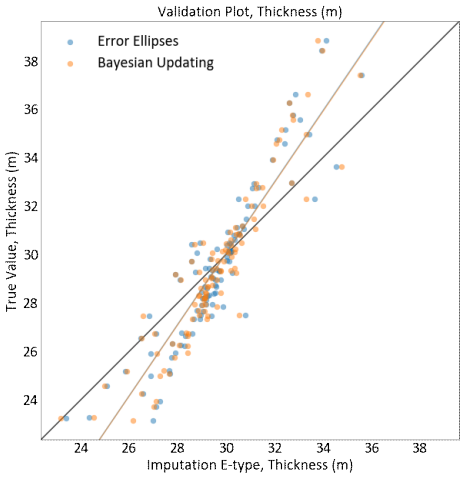
Table 5.2: Summary Statistics for Along Drill Holes

| Variable | Merging Method | Error Rate | Correlation | RMSE |
|---|---|---|---|---|
| $W_{HW}$ | EE | **1.397** | **0.928** | **1.753** |
| $W_{HW}$ | BU | 1.451 | 0.901 | 1.768 |
| $W_{FW}$ | EE | 1.940 | **0.932** | 1.707 |
| $W_{FW}$ | BU | **1.919** | 0.931 | **1.649** |
| TH | EE | 5.067 | **0.741** | 2.299 |
| TH | BU | **4.897** | 0.756 | **2.262** |

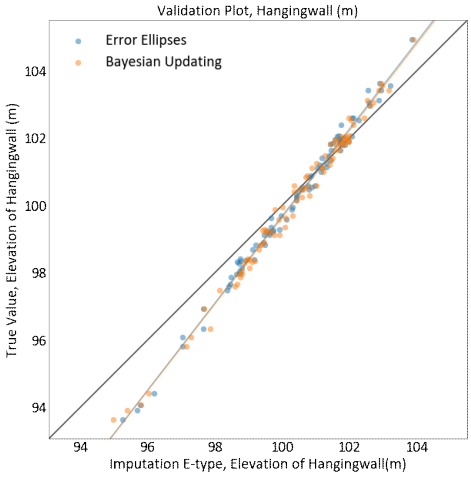
Accuracy and bias are also compared in Figure 5.4. In these scatter plots, e-types estimates are plotted on the x-axis against the true values on the y-axis. The overall performance of the imputation of elevation shows a strong correlation, however, the performance of the thickness is not as good. The regression lines illustrate some conditional bias. Again, the method tends to lower the variance. The two merging method produce similar results.

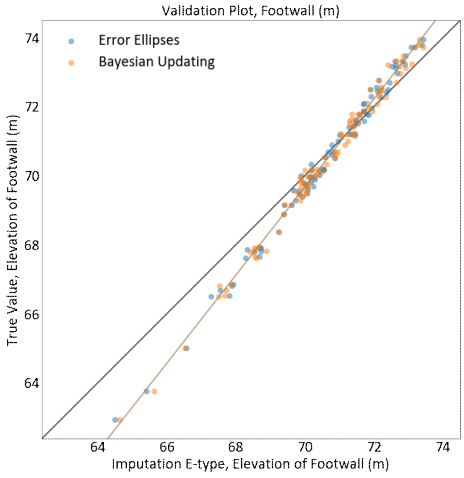
Reproduction of the CDF and variograms of the imputed realization are shown in Figure 5.5. The method reproduce the CDF and variograms quite well. The imputed results of global distribution does not reduce the variance. It is probably because the imputation along long drill holes does not have a strong constraint on elevation. The reproduction of the variogram of the thickness is slightly away from the true variogram.

The overall performance of the method seems acceptable for elevation and thickness. The reproduction of CDF and variogram are good. Furthermore, the global distribution imputed represents the true distribution. It seems that the assumptions about the distribution and the elevation used are reasonable.




(a)




(b) (c)

Figure 5.4: Validation Plot for Along Long Drill Hole

In (a), the imputed values of the thickness are plotted against the true values. In (b), (c), the imputed values of the elevation are plotted against the true values.

## 5.4 Sensitivity Analysis of Imputation Along Drill Holes

Two parameters that influence the performance of the method are evaluated in this Section. The first one is the imputation frequency and the second is the imputation sequence.

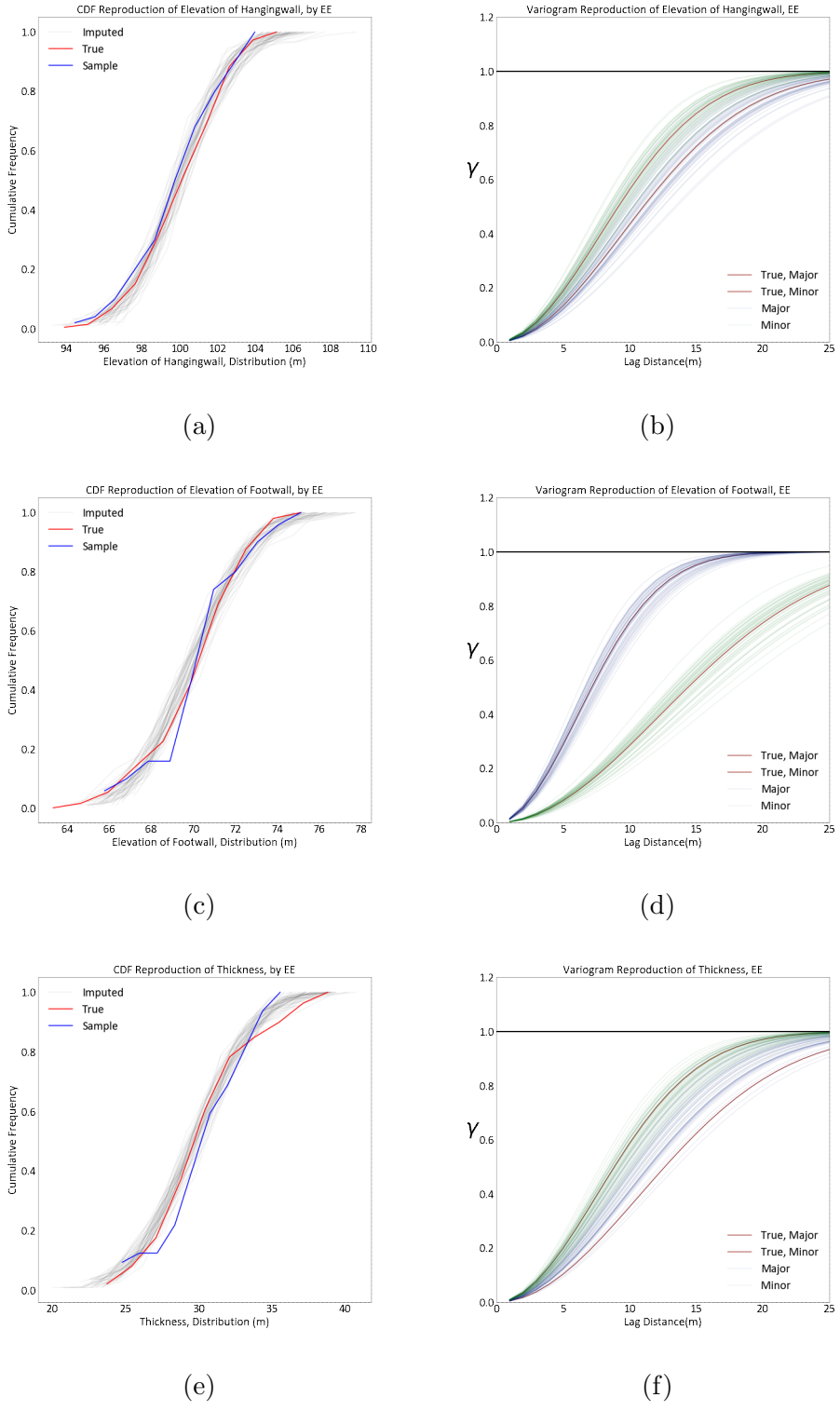



Figure 5.5: Reproduction of CDF and Variogram Along Drill Holes

In (a), (c), (e), the reproduction of the CDF are plotted. In (b), (d), (f), the reproduction of variogram are plotted.

### 5.4.1 Imputation Frequency

During the prerequisites of imputation along long drill holes, a set of coordinates must be added to the geometric dataset. The imputation frequency determines how many additional coordinates are added along long drill holes. Figure 5.6 illustrates four figures showing the imputed results of hangingwall. The imputation frequencies are 2, 4, 8, 16, respectively. In the last figure, it is possible to see the general configuration of the long drill holes imputed.

The performance of different imputation frequency is shown in Figure 5.7. It shows that there is no significant difference between different imputation frequencies. As a result, it is possible to choose any imputation frequencies. Of course, if the frequency is too small, there may be violations of the drill hole data in subsequent simulation.

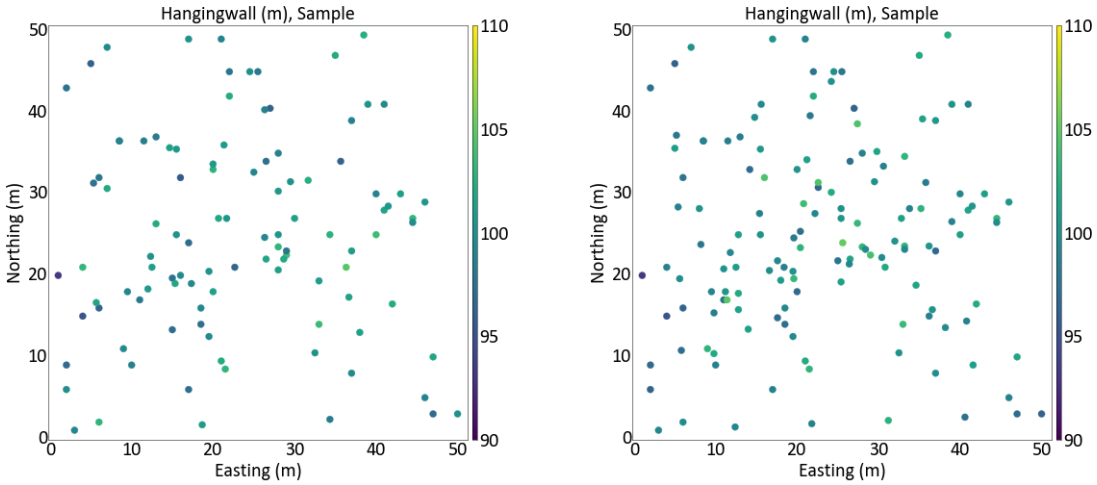
### 5.4.2 Imputation Sequence

The sequence of imputation is tested again. The imputation starts from the most vertical drill hole towards the most inclined drill holes. Here, the imputation sequence considers the area near pierce points, and the added points along the drill holes. The first sequence is to impute pierce points at first and then the points added. The second sequence is to impute both the pierce points and the points added together.

It is shown that it is better to impute at pierce points first. The sequence of the added points within a drill holes does not appear to influence the performance of the method significantly.

## 5.5 Discussion on Imputation Along Drill Holes

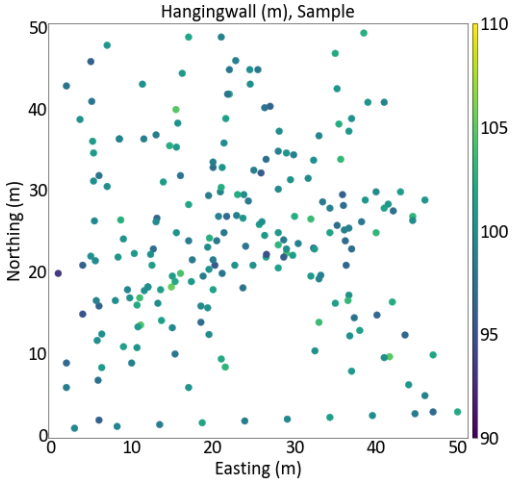
The method of multiple imputation can be use to impute a location if the coordinates of this location is along the projection of the drill hole. The overall performance of the method is acceptable. However, there are some differences between the results of the pierce points and added points. In Figure 5.9, the imputation results using the pierce points and using the added points are shown. It is shown that the results of the pierce points are better than the results of the added points. The results for elevation are quite similar, however, the results for thickness are much better for the pierce points.



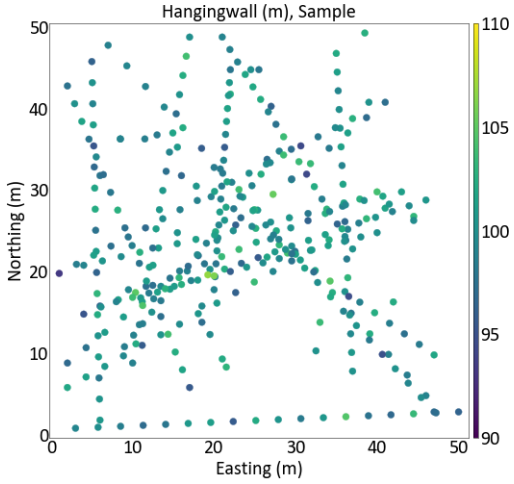

(a)




(b)




(c)

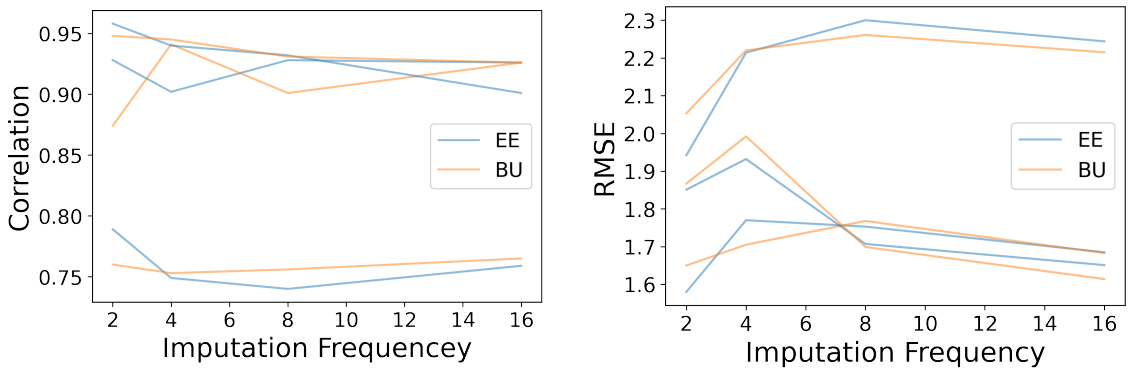



(d)

Figure 5.6: Imputed Realizations with Different Imputation Frequency

Location maps of imputed realizations with different imputation frequency are shown. It is possible to see the general configuration of the long drill holes imputed in (d).

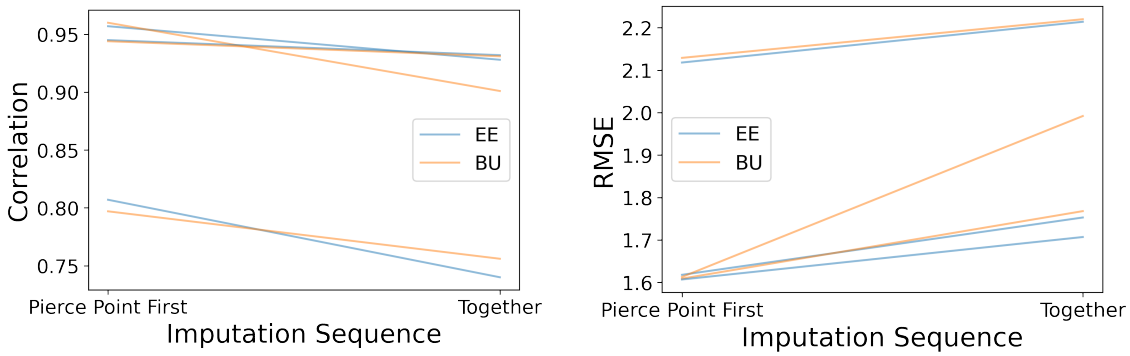
It seems that the assumptions about the elevation and the distribution of thickness can be applied. If the imputed value of the elevation are beyond the range of the original distribution, it is better to reject the imputed results and draw again.




(a) (b)

Figure 5.7: Correlation and RMSE of Different Imputation Frequency

In (a), the correlation of the imputed values with the true values with different imputation frequency are shown. In (b), the RMSE are shown. These is no significant difference between different imputation frequency.




(a) (b)

Figure 5.8: Correlation and RMSE of Different Imputation Sequence

In (a), the correlation of the imputed values with the true values with different imputation sequence are shown. In (b), the RMSE are shown. The better sequence is to impute pierce points at first.

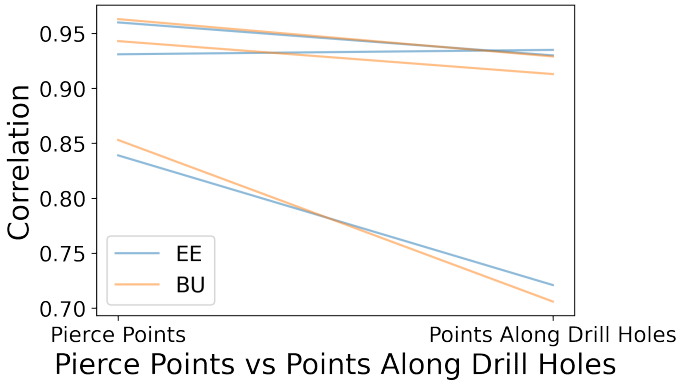



Figure 5.9: Pierce Points vs Points Along Drill Holes

The correlation of the imputed values from the pierce points and from the points along drill holes are compared. The pierce points have a higher correlation, especially for the thickness.

# Chapter 6

# Case Study

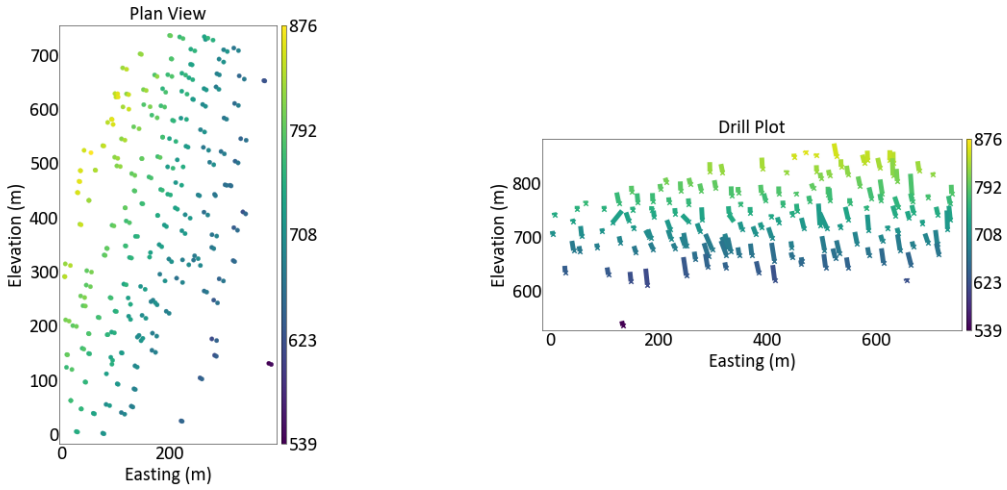
*A case study demonstrates the workflow of multiple imputation applied to tabular vein deposits to quantify geometric uncertainty to better model the geological domain. The data for the case study consist of a tabular gold vein. The workflow of geometric data imputation including the prerequisites and the results for inclined drill holes are shown. A cross validation is used to evaluate the performance of the method.*

## 6.1 Case Study 1

The data of the case study comes from a gold mine in Brazil. The information about the deposit and data description can be found in this reference (Carvalho, 2018). In Figure 6.1, the configuration of drill holes are shown. The plan view of the pierce points of the drill holes with the deposit is plotted, and the drill plot is plotted.

Multiple imputation is the method to quantify geometric uncertainty. The prerequisites of multiple imputation include TLS, angle tolerance and thickness calculation, normal score transform, and variogram calculation and modeling. The detailed calculation of TLS can be found in this reference (Carvalho, 2018). Angles of all the drill holes are calculated. In Figure 6.2, the cumulative distribution of the calculated angles as well as the angle tolerance applied are shown. The angle tolerance is set to be 34 degrees, which includes 90 percent of the drill hole angles, because most of the drill holes has small angles (perpendicular to the plane of continuity) and the geometric uncertainty is low (Carvalho, 2018).

After applying the angle tolerance, the sample of the geometric variables, including the elevation of the hangingwall, the elevation of the footwall, and the thickness, can be shown in Figure 6.3. These samples include all the pierce points of the drill holes that have smaller angles than the angle tolerance. Other drill holes that have larger




(a) (b)

Figure 6.1: Location Map of Drill Holes in Case Study 1

In (a), a location map of the plan view of the pierce points is shown. In (b), a cross section view of the drill hole is shown.

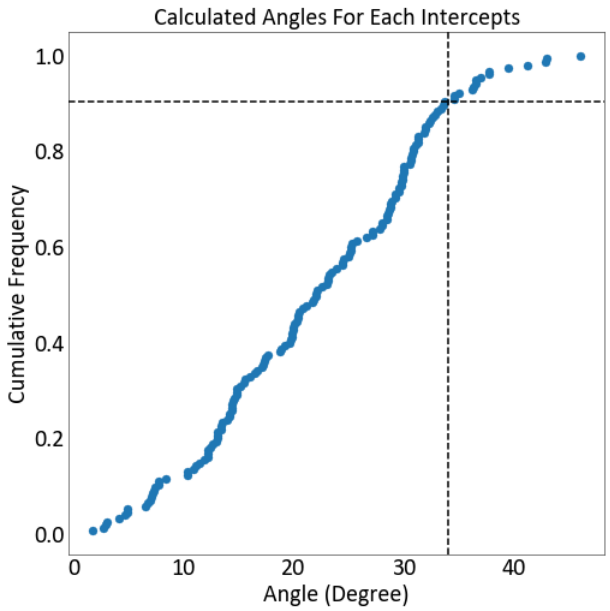
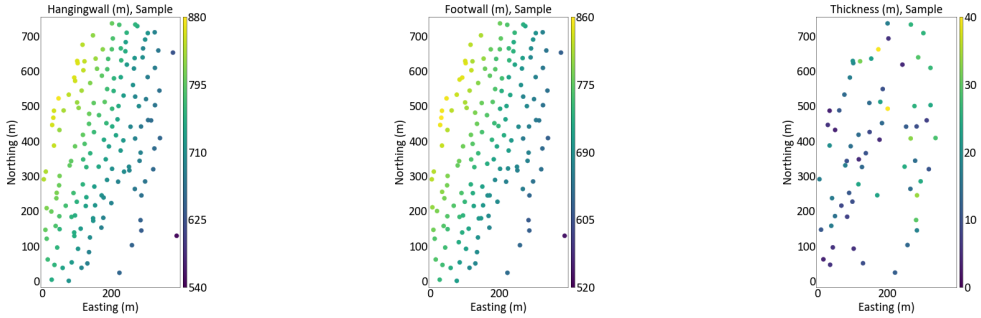



Figure 6.2: Cumulative Distribution of the Angles of Drill Holes for Case Study 1

The cumulative distribution function of the angles of the drill holes are plotted and an angle tolerance of 34 degrees is used.

angles than the angle tolerance will be imputed. There are fewer sample for the thickness than the elevation.




(a) (b) (c)

Figure 6.3: Location Map of Samples in Case Study 1

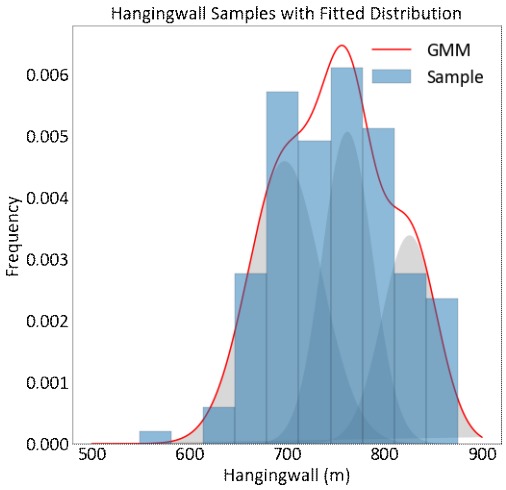
In (a), the location map of the elevation of the hangingwall is shown. In (b), the location map of the elevation of the footwall is shown. In (c), the location map of the thickness is shown.

The normal scored geometric variables are used for data imputation. Because the parametric distributions of these geometric variables are not known, they are normal score transformed by fitting these distributions by GMM. The results of the transformation are shown in Figure 6.4. The transformation also enables the calculation and modeling of variograms as well as the transformations during multiple imputation.

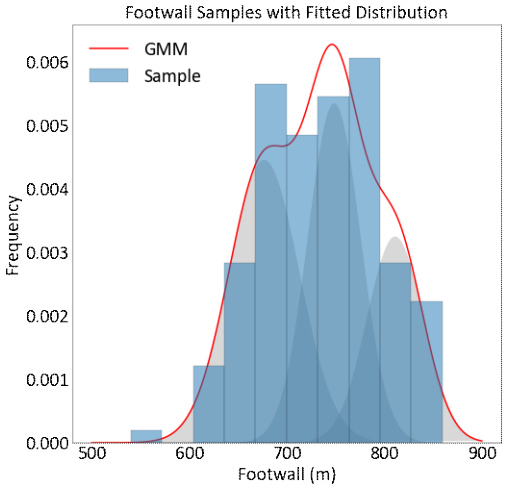
The variograms of these geometric variables can be calculated and modeled independently. The modeled variograms are shown in Figure 6.5. These variograms are fitted with a Gaussian model with a zero nugget effect, as well as a long range.

After these prerequisites, multiple imputation can be executed at both the hangingwall and the footwall. At locations that is being imputed, local distributions of elevation and thickness are formed by kriging. Then, the distribution of the thickness are transformed into the elevation. Both the distribution of elevation and the transformed distribution of thickness are merged by error ellipses and imputed values can be drawn.

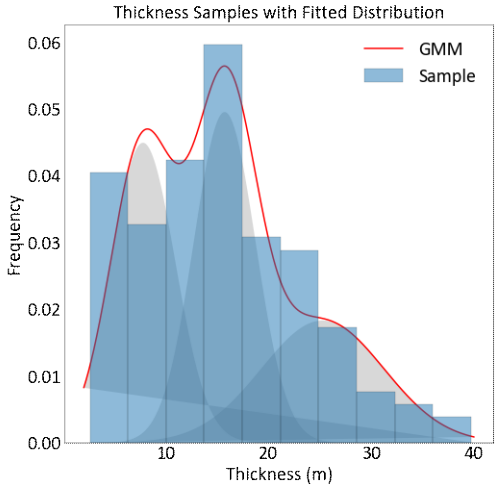
After the imputation of the geometric variables, many homotopic imputation data files will be generated. One realization is shown in 6.6 Each imputation file has different values of the geometric variables. And it is possible to use these geometric data to help the modeling of the geological domains. And other quantification of uncertainties based on these different imputation files can be processed.




(a)




(b)

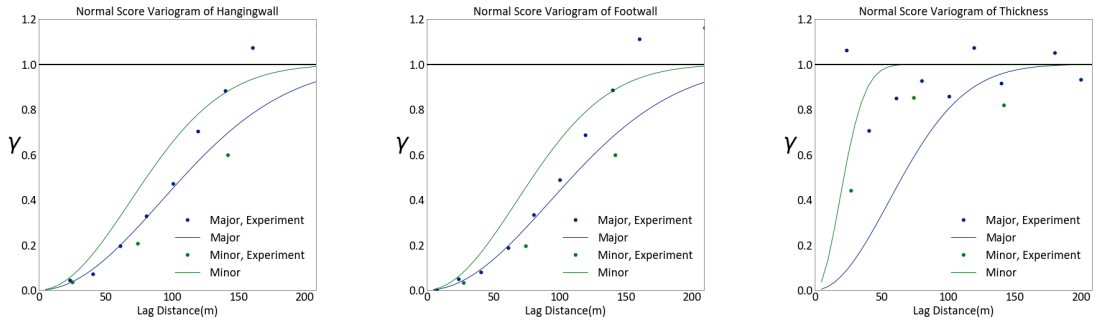



(c)

Figure 6.4: Distributions of Geometric Samples in Case Study 1

In (a), the fitted distribution of elevation of the hangingwall is shown. In (b), the fitted distribution of elevation of the footwall is shown. In (c), the fitted distribution of the thickness is shown.

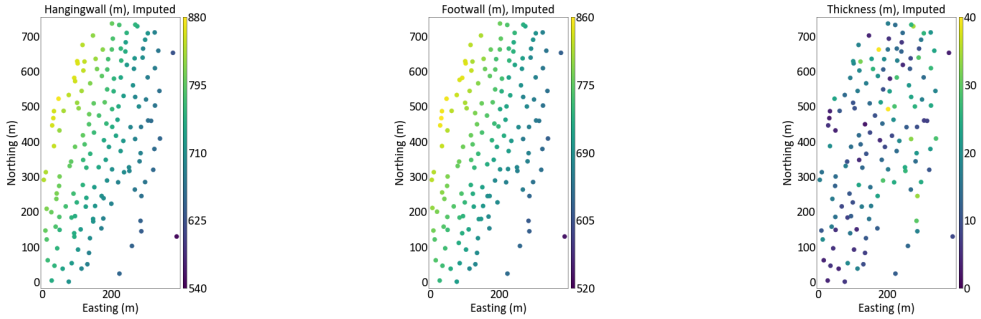
A cross-validation is used to test the results of the imputation. 70 percent of the data are used as the training data and multiple imputation is applied to impute the other 30 percent of the data, which serve as the testing data. These testing data are randomly left out. After multiple imputation, the e-type values of the imputed results are compared with the testing data. Three summary statistics: error rate, Pearson




(a) (b) (c)

Figure 6.5: Variograms of Geometric Samples in Case Study 1

In (a), the variogram of the elevation of the hangingwall is shown. In (b), the variogram of the elevation of the footwall is shown. In (c), the variogram of the thickness is shown.




(a) (b) (c)

Figure 6.6: One Realization of Case Study 1

In (a), the imputed results of the elevation of the hangingwall is shown. In (b), the imputed results of the elevation of the footwall is shown. In (c), the imputed results of the thickness is shown.

correlation coefficient ($\rho$) and root mean square error (RMSE) are summarized in Table 6.1.

From the results, the method of Bayesian updating outperforms slightly the method of error ellipses for the elevation. The error rate of the elevation can be reduced from around 3 to around 1 by Bayesian updating. However, for the thickness, the method of error ellipses performs better. As a result, it is possible better to use different merging method for different geometric variables. The overall correlation is around 0.9, which means a strong correlation. The results of the RMSE are consistent with the error rates. The better result is bolded in the table in order to compare the two merging methods.

Table 6.1: Summary Statistics for Case Study 1

| Variable | Merging Method | Error Rate | Correlation | RMSE |
|---|---|---|---|---|
| $W_{HW}$ | EE | 2.606 | 0.890 | 25.69 |
| $W_{HW}$ | BU | **1.618** | **0.967** | **14.23** |
| $W_{FW}$ | EE | 3.638 | 0.878 | 36.07 |
| $W_{FW}$ | BU | **1.122** | **0.971** | **9.727** |
| TH | EE | **6.002** | **0.937** | **2.556** |
| TH | BU | 7.883 | 0.866 | 3.512 |

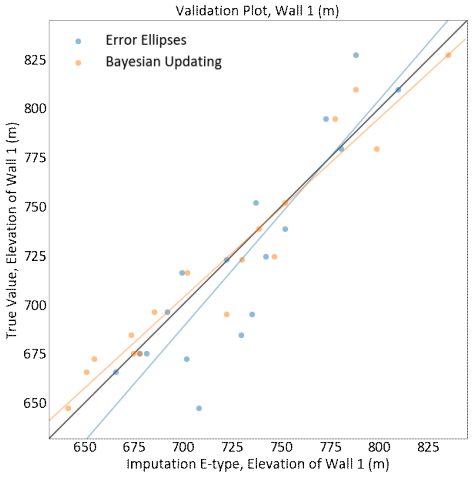
Accuracy and bias are compared in Figure 6.7. In these scatter plots, e-type estimates of the training data are plotted on the x-axis against the testing data on the y-axis. The overall performance shows a strong correlation. The regression line illustrates some conditional biases. Again, the method tends to reduce the variance of the geometric variables.

The difference of the results by the two merging methods are much more obvious than the one from the synthetic example. In the scatter plot, the close pairs of the two merging methods are not observed.

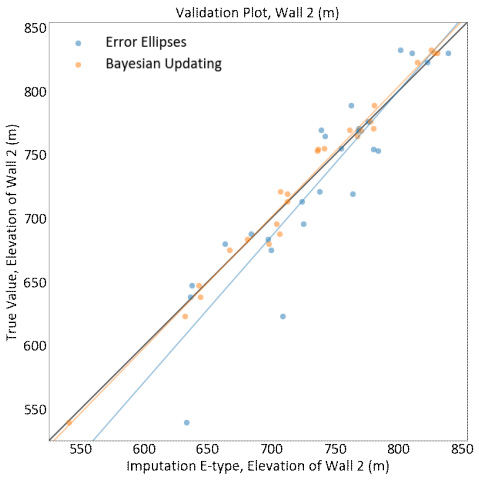
Reproduction of the CDF of the imputed realization is shown in Figure 6.8. The distribution of realizations are plotted in gray. The true distribution in red and the sample distribution in blue are plotted for comparison. The method reproduces the distribution approximately, but it reduces the variances for all geometric variables slightly.

Reproduction of the variograms of the imputed realization is shown in Figure 6.9. The method reproduces the variograms approximately for the elevation. The variogram reproduction of the thickness is more disperse. It is probably because that there are more thickness values being imputed, and the variogram of the thickness are more unstable.

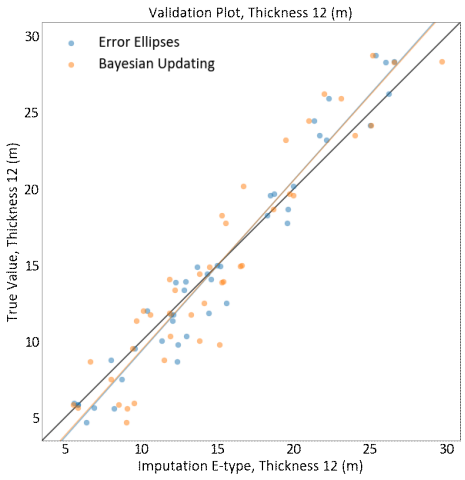
The results of the cross-validation show that the method has an strong correlation, and acceptable reproduction of the distribution and the variogram.




(a)



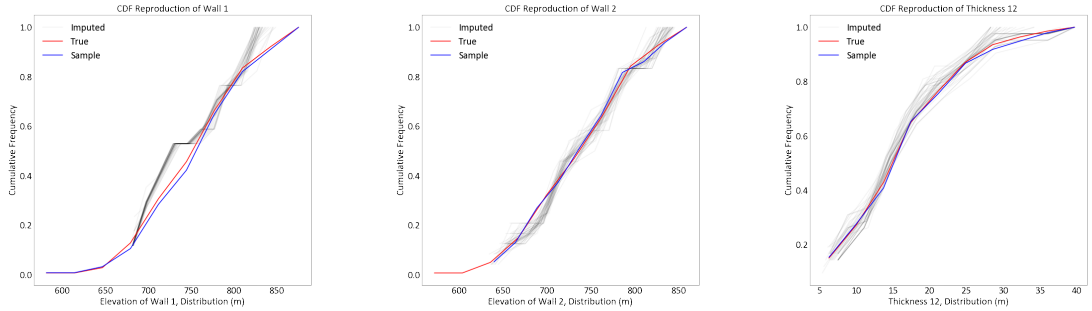

(b)




(c)

Figure 6.7: Validation Plot for Case Study 1

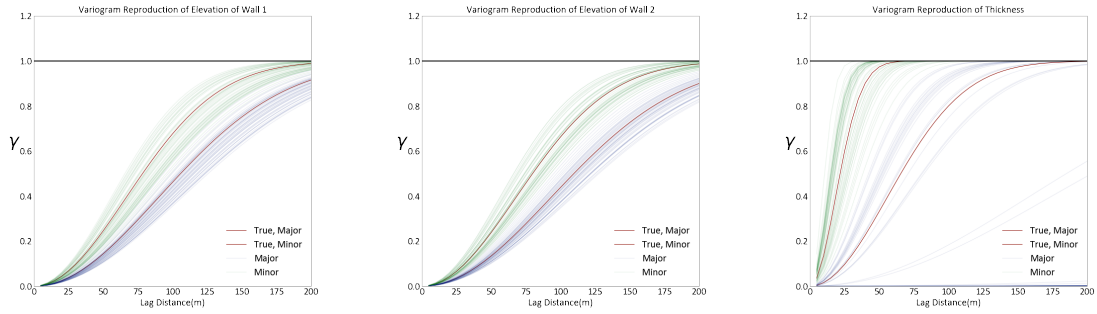
The imputed values using the training data are compared with the test data. In (a), the results of the hangingwall are shown. In (b), the results of the footwall are shown. In (c), the results of the thickness are shown.




(a) (b) (c)

Figure 6.8: Distribution Reproduction of Case Study 1

In (a), the distribution of the imputed realizations of the elevation of the hangingwall is plotted. In (b), the elevation of the footwall is plotted. In (c), the thickness is plotted.




(a) (b) (c)

Figure 6.9: Variogram Reproduction of Case Study 1

In (a), the reproduction of the variogram of the imputed realizations of the elevation of the hangingwall is plotted. In (b), the reproduction of the footwall is plotted. In (c), the reproduction of the thickness is plotted.

# Chapter 7

# Conclusion and Future Work

*The workflows of multiple imputation used to model the geological domain for single-layered, multi-layered tabular deposits and multiple imputation along drill holes are presented. Synthetic examples are used to demonstrate the steps and enable the discussion of the results and the sensitivity analysis. This chapter reviews the problem and contributions made in this thesis. Proposals for future work are presented.*

## 7.1 Review and Contributions

Geometric uncertainty is one of the uncertainties encountered during the modeling of the geological domain of tabular vein deposits. It is due to sparse drilling and it increases with inclined drill holes. Multiple imputation is a kind of sequential Gaussian simulation used to quantify this geometric uncertainty.

The prerequisites of the method includes building a local coordinates system, angle and thickness calculation, as well as normal score transform and variogram calculation and modeling. The building of a local coordinates system by Total Least Square (TLS) allows the calculation of thickness. The calculation of the angles of the drill holes decides whether a drill hole should be imputed or not. The thickness values from the vertical drill holes form the distribution of the thickness, and others will be imputed. After that, the geometric variables, including elevation (position vertical to the plane of continuity) and thickness, can be normal score transformed. Variogram calculation and modeling can be performed for these geometric variables.

Multiple imputation combines distribution of elevation and distribution of thickness in order to make realistic realization. The mean and the variance of the primary distribution of elevation is derived from kriging using nearby spatial information. The mean and the variance of the secondary distribution of thickness is derived from

kriging using nearby spatial information as well. Then, the secondary distribution of thickness is transformed into the distribution of elevation. The two distributions of elevation are merged. After forming this distribution, a realization of multiple imputation can be drawn. The workflow can proceed to the next location.

The method can be applied to three scenarios. The first one is for single-layered deposits. The second one is for multi-layered deposits. The third one is for high-deviated long drill holes. The original method can be used for single-layered deposits and it can extended to multi-layered deposits. The difference is that there are more places being imputed and there are more opposite elevation values that can be used to transform thickness distribution into elevation distribution. In order to impute areas with highly deviated long drill holes, assumption about the opposite elevation and the distribution of thickness are made. Results with synthetic data show that the method can provide acceptable imputation results.

In order to achieve optimum performance of multiple imputation, many parameters must be considered. It is important to choose a good angle tolerance. If the angle tolerance is too low, there will be too few data to form the global distribution and conduct kriging; however, if the angle tolerance is too high, inappropriate thickness values would be included in the global distribution. It is important to consider the angle tolerance with drill holes numbers to form a sample with a proper size. Imputation Sequence is another important factor. The sequence of imputation always starts from the most constrained area to the least constrained area. For multiple imputation, it is shown that the imputation should start from the most vertical drill hole to the most inclined drill hole. Within a drill hole, it is better to use kriging variance to decide the sequence. During transformation, it is the best to have a global parametric distribution; empirical fitting by linear interpolation and Gaussian mixture models (GMM) also provide good results. Overall, the method of Bayesian updating and the method of error ellipses performance quite similarly. Bayesian updating performs better when the distribution of geometric variables are known. The later method performs better when the distribution of geometric data is not known and when the layers of the deposits increases. For multi-layered deposits, it is better to use the nearest opposite elevation value to conduct transformation. For imputation along long drill holes, it seems that it is possible to impute locations along drill holes

without sacrificing the accuracy.

There are some limitations of multiple imputation. The most important limitation is that during the transformation, there is no guarantee that the imputed values would be in the original range of the geometric variables. Another important limitation is that the imputed values can have a lower variance because the imputed values are sometimes based on the sample only and the calculated thickness distribution may not represent the true thickness distribution. Another characteristic is that because there is fewer data for thickness values, the variance of the thickness can be larger than the elevation. As a result, during the merging of the two distributions, the distribution of elevation can have more weights. For multi-layered deposits, the number of imputed values can increase dramatically. As a result, some areas can be highly constrained and do not need imputation. TLS can provide inclined results for multi-layered deposits. For imputation along drill holes, although the overall performance is acceptable, the imputed areas along drill holes are less accurate than the imputed areas near pierce points.

The main contribution of the thesis is the implementation of the method of multiple imputation for different scenarios. Multiple imputation now integrates both the distribution of elevation and the distribution thickness during imputation. Furthermore, the sensitivity analysis about different parameters that the method should choose allows a better performance of the method. The method can be applied to model the geological domain problematically.

## 7.2 Future Work

The complete workflow of multiple imputation can be integrated into the framework of probabilistic resource modeling of tabular vein deposits. After geometric data imputation to construct probabilistic geological domains, the surface simulation can be conducted as the next step of the probabilistic modeling of the deposits. Then, the boundary modeling, the grade simulation, and the post-processing including sensitivity analysis.

The proposed method of multiple imputation has some limitations.

It is important to have a representative thickness distribution. The current method

approximately calculates the thickness values. Furthermore, it is possible that there is no vertical drill hole that can provide thickness values near some locations. Additional research is required to determine the thickness distribution in these challenging circumstances. Some iterative methods can be used to get the thickness distribution by explicit modeling. Furthermore, it is assumed that the three geometric variables are independent. However, it is likely that there are some dependency between the elevation and the thickness. These correlation can be considered when modeling the geological domain.

The definition of simple tabular deposits can be extended to complex deposits with reverse folds and other complex geological units. Currently, some unfolding techniques can be applied to constrain deposits into this definition and normal faults can be handled. The assumption used in imputation along long drill holes can be used to reject possible imputed results and decide whether a point is within a certain surface. Additional research would be required for such complex deposits.

# Bibliography

J. Bai and C. Deutsch. The pairwise relative variogram. *Geostatistics Lessons*, 2020a. URL https://geostatisticslessons.com/lessons/pairwiserelative.

J. Bai and C. Deutsch. Multiple Imputation of Tabular Vein Deposit Geometry Data Using Error Ellipses. *Centre for Computational Geostatistics Annual Report*, 22: 1–12, 2020b.

J. Bai and C. Deutsch. Multiple Imputation of Multi-Layered Tabular Deposit Geometry Data. *Centre for Computational Geostatistics Annual Report*, 23:1–8, 2021a.

J. Bai and C. Deutsch. Parameters Influence Imputation and Other Developments. *Centre for Computational Geostatistics Annual Report*, 23:1–3, 2021b.

R. Barnett. *Managing Complex Multivariate Relations In The Presence Of Incomplete Spatial Data*. PhD thesis, University of Alberta, 2015.

R. M. Barnett and C. V. Deutsch. Multivariate imputation of unequally sampled geological variables. *Mathematical Geosciences*, pages 791–817, 2015.

M. Batty and J. Boisvert. Geometry Imputation of the Arrow Deposit. *Centre for Computational Geostatistics Annual Report*, 22:1–10, 2020a.

M. Batty and J. Boisvert. Imputation of Density Values and the Uncertainty in Grade and Density at the Arrow Deposit. *Centre for Computational Geostatistics Annual Report*, 22:1–13, 2020b.

M. Batty and J. Boisvert. Multivariate Spatial Bootstrap of the Arrow Deposit. *Centre for Computational Geostatistics Annual Report*, 22:1–11, 2020c.

M. Batty and J. Boisvert. Boundary Uncertainty of the Arrow Deposit. *Centre for Computational Geostatistics Annual Report*, 23:1–10, 2021a.

M. Batty and J. Boisvert. Uncertainty Review of the Arrow Deposit. *Centre for Computational Geostatistics Annual Report*, 23:1–12, 2021b.

D. Carvalho. Probabilistic resource modeling of vein deposits. Master's thesis, University of Alberta, 2018.

D. Carvalho and C. V. Deutsch. A Framework for Tabular Vein Type Deposit Modeling. *Centre for Computational Geostatistics Annual Report*, 19:1–9, 2017a.

D. Carvalho and C. V. Deutsch. Developments on Tabular Vein Geometry Modeling. *Centre for Computational Geostatistics Annual Report*, 19:1–9, 2017b.

D. Carvalho and C. V. Deutsch. Imputation of Tabular Vein Geometry Data for Tonnage Uncertainty Assessment. *Centre for Computational Geostatistics Annual Report*, 19:1–9, 2017c.

D. Carvalho and C. V. Deutsch. Data and Parameter Uncertainty for Tabular Deposits. *Centre for Computational Geostatistics Annual Report*, 20:1–13, 2018a.

D. Carvalho and C. V. Deutsch. Implementation of Probabilistic Resource Modeling for Tabular Deposits Results. *Centre for Computational Geostatistics Annual Report*, 20:1–6, 2018b.

D. Carvalho and C. V. Deutsch. Probabilistic Resource Modeling for Tabular Deposits. *Centre for Computational Geostatistics Annual Report*, 20:1–21, 2018c.

D. Carvalho and C. V. Deutsch. Vein Deposit Resource Uncertainty: A Case Study. *Centre for Computational Geostatistics Annual Report*, 20:1–14, 2018d.

J. Cowan, R. Beatson, H. Ross, W. Fright, T. McLennan, T. Evans, J. Carr, R. Lane, D. Bright, A. Gillman, P. Oshurst, and M. Titley. Practical implicit geological modelling. 11 2003.

F. M. D'Affonseca, M. Finkel, and O. A. Cirpka. Combining implicit geological modeling, field surveys, and hydrogeological modeling to describe groundwater flow in a karst aquifer. *Hydrogeology Journal*, 28(8):2779–2802, 2020.

C. K. Enders. *Applied missing data analysis*. Guilford press, 2010.

O. Erten and C. Deutsch. Combination of multivariate gaussian distributions through error ellipses. *Geostatistics Lessons*, 2020. URL https://geostatisticslessons.com/lessons/errorellipses.

E. H. Isaaks and M. R. Srivastava. *An Introduction to Applied Geostatistics*. Oxford Unviersity Press, 1989.

R. J. Little and D. B. Rubin. *Statistical analysis with missing data*, volume 793. John Wiley & Sons, 2002.

J. Manchuk and C. Deutsch. Geometric Modeling of Irregular Tabular Deposits. *Centre for Computational Geostatistics Annual Report*, 17:1–20, 2015.

P. McInerney, A. Goldberg, P. Calcagno, G. Courrioux, A. Guillen, and R. Seikel. Improved 3d geology modelling using an implicit function interpolator and forward modelling of potential field data. In *Proceedings of exploration*, volume 7, pages 919–922, 2007.

J. McLennan and C. Deutsch. Implicit boundary modeling (BOUNDSIM). *Centre for Computational Geostatistics Annual Report*, 08, 2006.

B. Ostenberg and C. Deutsch. Transforming Vein Deposit Coordinates for Improved Resource Estimation. *Centre for Computational Geostatistics Annual Report*, 19: 1–8, 2017.

M. Rossi and C. Deutsch. *Mineral Resource Estimation*. 01 2014. ISBN 978-1-4020-5716-8. doi: 10.1007/978-1-4020-5717-5.

D. Silva and C. Deutsch. Well / Drill Hole Prediction Tool. *Centre for Computational Geostatistics Annual Report*, 17:1–19, 2015.

S. A. Vollgger, A. R. Cruden, and J. Cowan. 3d implicit geological modeling of a gold deposit from a structural geologist's point of view. In *12th SGA Biennial Meeting-Mineral Deposit Research for a High-Tech World*, pages 1–4. SGA, Uppsala, Sweden, 2013.

L. Yang, P. Achtziger-Zupančič, and J. Caers. 3d modeling of large-scale geological structures by linear combinations of implicit functions: Application to a large banded iron formation. *Natural Resources Research*, 30(5):3139–3163, 2021.

H. Zhang, O. Erten, and C. Deutsch. Bayesian updating for combining conditional distributions. *Geostatistics Lessons*, 2020. URL https://geostatisticslessons.com/lessons/bayesianupdating.

Ítalo Gomes Gonçalves, S. Kumaira, and F. Guadagnin. A machine learning approach to the potential-field method for implicit modeling of geological structures. *Computers Geosciences*, 103:173–182, 2017. ISSN 0098-3004. doi: https://doi.org/10.1016/j.cageo.2017.03.015. URL https://www.sciencedirect.com/science/article/pii/S0098300416304848.