Geostatistical Reservoir Modeling with Parameter Uncertainty in Presence of Limited Well Data

by

Mehdi Rezvandehy

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

Doctor of Philosophy
in
Mining Engineering

Department of Civil and Environmental Engineering
University of Alberta

Abstract

An important challenge in reservoir management is establishing reliable numerical geological models of all required flow parameters including facies, porosity and permeability. These numerical models are driven by conceptual geology, seismic, production and well data that are widely spaced early in exploration. Critical decisions are made early in the reservoir lifecycle where limited seismic and production data may be available. Geostatistical simulation is commonly used to construct these numerical models and quantify uncertainty. This thesis develops techniques that improve the uncertainty represented in the final geostatistical model.

The variogram is a key parameter for geostatistical simulation. In presence of preferential positioning of the wells to maximize production, variogram modeling is suboptimal. A novel technique is proposed to weight variogram pairs in order to compensate for preferential or clustered sampling. Weighting the variogram pairs helps remove noise and minimize artifacts in the experimental variogram.

A new approach of variogram uncertainty is developed since variogram declustering does not remove all uncertainty in the experimental variogram. Variogram realizations are drawn from the uncertainty interval of lag distances honoring the correlation between lags. The realizations are transferred to geostatistical simulation to incorporate variogram uncertainty in the numerical geological models.

A methodology to improve horizontal variogram inference from the widely spaced well data is developed considering seismic data and the vertical well variogram. Seismic data provide constraints on the unknown horizontal variogram of the well data. The vertical variogram of the well data can be scaled to scenarios of the horizontal variogram. Improved horizontal variogram realizations are achieved by considering these

constraints.

Uncertainty in the histogram of flow parameters affects resource/reserve estimation, investment and development decisions. A new simulation-based approach of quantifying histogram uncertainty is also established. A methodology to calculate the true histogram uncertainty for a single variable in certain circumstances is proposed. This allows checking the proposed spatial bootstrap methodology. Multivariate distribution uncertainty is implemented considering the correlation between variables.

The applicability of the proposed methodology for variogram and histogram uncertainties are shown with an offshore real reservoir located in the Dutch sector of the North Sea. This case study confirms that the histogram uncertainty has the highest impact on resource estimation.

Dedication

This thesis is dedicated to my beloved wife, Samaneh, for her unconditional love and being a constant source of support and encouragement when there was no one to answer my queries.

Acknowledgements

I would like to express my special appreciation and thanks to my supervisor Dr. Clayton V. Deutsch, the professor and director of the Centre for Computational Geostatistics (CCG) for the continuous support of my Ph.D study and related research, motivation, patience, and immense knowledge. His guidance helped me in all the time of research and writing of thesis.

I would like to thank the sponsors of the CCG for financial support. I am also grateful to all CCG students and research associates for their supports and providing a stimulating and fun environment in which to learn and grow.

I also want to thank my committee members Dr. Juliana Leung, Dr. Jeffery Boisvert, Dr. Japan Trivedi, and external examiner Dr. Deepak Devegowda for their brilliant comments and suggestions.

Last, but not the least, special thanks to my family for their love and encouragement. Words cannot express how grateful I am to my parents for all of the sacrifices that they have made on my behalf and supported me in all my pursuits. I would also like to thank my sibling who incented me to strive towards my goal.

Table of Contents

List of Tables

List of Figures

List of Symbols

Symbol	Definition	First Use
P90	proved reserves	4
P50	proved and probable reserves	4
P10	proved, probable and possible reserves	4
UI	uncertainty index	5
$\hat{\gamma}(\mathbf{h})$	experimental variogram for lag distance $\mathbf{h}$	16
$n(\mathbf{h})$	number of variogram pairs for lag distance $\mathbf{h}$	16
$z(\mathbf{u}_\alpha)$	value of random variable z at location $\mathbf{u}_\alpha$	20
A	stationary modeling domain	21
λ_α^{OGK}	ordinary global kriging declustering weight	21
$C(\mathbf{u}_\beta - \mathbf{u}_\alpha)$	covariance between locations $\mathbf{u}_\beta$ and $\mathbf{u}_\alpha$	21
μ	Lagrange parameter	21
$\overline{C(\mathbf{u}_\alpha - A)}$	average covariance between location $\mathbf{u}_\alpha$ and domain A	21
$\overline{C(A - A)}$	average covariance of domain A to itself	21
$\uparrow$	variogram pair	21
$X_{\mathbf{h}}(\mathbf{u})$	half of quadratic differences of pair value for lag distance $\mathbf{h}$	21
$X_{\mathbf{h}}^*$	declustered variogram for lag distance $\mathbf{h}$	22
$F(X_\beta - X_\alpha)$	fourth order covariance between pairs X_β and X_α	22
$\overline{F(X_\alpha - A)}$	average fourth order covariance between pair X_α and domain A	22
$\overline{F(A - A)}$	average fourth order covariance of domain A to itself	23
$\gamma(\mathbf{h})$	variogram model for lag distance $\mathbf{h}$	24
m	meter	25
$\lambda_{\alpha,\mathbf{h}}^{OLK}$	ordinary local kriging weight for variogram declustering	28
km	kilometer	36
$\overline{F_{i,j}}$	average fourth order covariance between variogram pairs	45
$\lambda_{\alpha,\mathbf{h}}^{SGK}$	simple global kriging weight for variogram uncertainty	47
χ^2_{DoF}	Chi-square distribution	52

List of Abbreviations

Abbrv.	Definition	First Use
EOR	enhanced oil recovery	1
OWC	oil water contact	2
HIIP	hydrocarbon initially in place	3
GRV	gross rock volume	3
NTG	net-to-gross	3
SPE	Society of Petroleum Engineers	4
JCPT	Journal of Canadian Petroleum Technology	4
OOIP	original oil in place	5
MCS	Monte Carlo simulation	5
CFD	conditional finite domain	6
SGS	sequential Gaussian simulation	7
OGK	ordinary global kriging	20
CPU	central processing unit	22
OLK	ordinary local kriging	28
TPCD	two-point cell declustering	36
FOM	fourth order moment approach for variogram uncertainty	45
SGK	simple global kriging	47
DoF	degree of freedom approach for variogram uncertainty	52
SB	spatial bootstrap	56
MSE	mean square error	87
SCK	simple cokriging	90
OCK	ordinary cokriging	90
SGCK	simple global cokriging	91
OGCK	ordinary global cokriging	91
LMC	linear model of coregionalization	102
GK	global kriging	157

Chapter 1

Introduction

Numerical reservoir models for the upstream petroleum exploration/production business are used for developing a recovery plan, reserves estimation, redevelopment of old fields, accurate management throughout the production period, monitoring and execution, EOR planning and abandonment of production. The process of reservoir modeling is cyclic and continues through the lifetime of the reservoir (Singh, Yemez, & Sotomayor, 2013).

Developing hydrocarbon reservoirs is risky. The industry aims to predict and mitigate risk (Kaleta et al., 2012; Wolff, 2010). According to Rose (2004), exploration and production companies had delivered about half of the predicted reserves in the last 20 years of the 20th century. Merrow (2011) shows a decreasing rate of success for exploration and production megaprojects. One of the main reasons for this underperformance is due to use of evaluation methods that do not account for the full uncertainty, which leads to inaccurate reservoir production forecasts (McVay & Dossary, 2014; Singh et al., 2013). In the past, the input statistics were held constant and relatively small fluctuations between realizations were used to characterize reservoir uncertainty (Pyrcz & Deutsch, 2014). This approach underestimates uncertainty (Wang & Wall, 2003). Uncertainty is small because local fluctuations above and below the average cancel out between locations (Babak & Deutsch, 2009). Thus, the uncertainty in each input parameter is important and must be integrated into final reservoir modeling.

According to Singh, Hegazy, and Fontanelli (2009), major sources of uncertainty in regionalized variables for production forecasts can be divided into four main categories: 1- Geophysical uncertainty (mainly seismic data) regarding conversion of time-to-depth, migration, picking seismic surface and finding correct position of faults and well ties. 2- Geological uncertainty regarding the sedimentary depositional environment, facies, spatial distribution and grain size. 3- Petrophysical uncertainties in porosity, permeability, water saturation, shale volume, net thickness of rock, and oil water contact (OWC), and 4- Dynamic uncertainties about relative and absolute permeabilities, horizontal barriers, thermodynamics, injectivity and productivity indexes. The uncertainty is due to limited data and measurement errors. Limited data leads to incomplete knowledge of petrophysical properties, subsurface structure, and fluid properties. Generating a model that represents the real reservoir is difficult. A realistic numerical model is required to understand the subsurface and quantify the inherent uncertainty.

The variogram is required by most geostatistical and uncertainty methods. Inferring a stable variogram model in presence of preferential sampling is a longstanding challenge. According to geological and geophysical data, the wells are likely located in the areas of higher quality for maximizing production (Pyrcz & Deutsch, 2014). This leads to noisy and unreliable experimental variogram due to unbalanced number of variogram pairs in low quality and high quality areas. Although weighting the data by declustering techniques corrects the statistics (Deutsch, 1989), this approach is not normally considered in variogram calculation. The challenge is to decluster the experimental variogram for removing noise and artifacts that occurs because of irregular and preferential sampling. Despite improving the experimental variogram by variogram declustering, there is unavoidable variogram uncertainty that could be calculated and incorporated in geostatistical modeling. The variogram uncertainty is very high in case of sparse well data.

Seismic data are more exhaustive than well data, so they can be used to improve the variogram and histogram of regionalized variables. There is the challenge of establishing a reliable horizontal variogram from well data. A poor horizontal variogram from well data can be improved by considering seismic data. The histogram of primary

variables constructed from few wells at the early stages of reservoir development will not be representative of the whole area. The histogram could be improved by calibration to seismic data (Deutsch, Frykman, & Xie, 1999, 2005). There is unavoidable uncertainty in the histogram due to the fact that multiple seismic attributes (Rezvandehy, Aghababaei, & Raissi, 2011) could lead to spurious correlation between well and seismic attributes (Chambers & Yarus, 2002; Kalkomey, 1997). Therefore, uncertainty in the representative histogram of variables (after calibration with seismic data) should be taken into consideration in geostatistical modeling.

The main objective of this research work is to propose techniques that estimate the correct uncertainty in input statistics such as the variogram and histogram of regionalized variables, and develop an approach for incorporating the uncertainty in geostatistical modeling. This leads to reservoir models with more realistic uncertainty for decision making.

1.1 Literature Review

1.1.1 The Importance of Reservoir Uncertainty

Uncertainty is an inevitable aspect of the upstream oil and gas business. There are many bad consequences of decisions made under uncertainty. The resources may be underestimated or overestimated and there may be unrealistic forecasted production.

For each reservoir, an estimate of the hydrocarbon initially in place (HIIP) is required. This estimation is often computed by gross rock volume (GRV), petrophysical properties including net-to-gross (NTG), porosity, fluid saturations, and hydrocarbon properties such as the formation volume factor. The uncertainty in HIIP depends on the uncertainty in each parameter. For instance, GRV is the volume of a reservoir trapped between stratigraphic surfaces and/or hydrocarbon-water contacts. Uncertainty in GRV is attributed to widely spaced well data and uncertainty in the interpreted structural surfaces from seismic data. In case of thin reservoirs, this uncertainty increases because of the seismic data resolution. Spectral and Cepstral decomposition approaches (Hall, 2006; Hall & Trouillot, 2004; Partyka, Gridley, & Lopez, 1999;

Rezvandehy et al., 2011) have been implemented by Fourier transform to improve resolution. There is no unique surface in units of depth because of uncertainty in the interpretation (in time) and uncertainty in the time-to-depth conversion. Therefore, the calculated GRV is uncertain. The oil water contact (OWC) can often be estimated from the available wells. The depth of this surface is subject to uncertainty. Uncertainty in all variables should be assessed and merged to compute the uncertainty in HIIP (Alshehri, 2010).

Deterministic approaches produce a single estimate. The deterministic approach usually includes sensitivity analysis. The analysis can be used to understand the critical variables by changing each parameter one at a time keeping the other parameters unchanged. The results could be presented by a tornado chart that ranks the impact of each parameter on the outcome. Then, the most important parameters can be considered for a more complete uncertainty analysis (Amudo, Graf, Dandekar, & Randle, 2009; Cebastiant & Osbon, 2011; Ehinola & Akinbodewa, 2014; Peng & Gupta, 2004; Salinas, Di Nezio, & Huerta Petroperu, 2014; Van Elk, Guerrera, Vijayan, & Gupta, 2000). This approach provides some information on the global uncertainty but it does not give a full picture of uncertainty. In addition, it does not account for local uncertainty (Alshehri, 2010). Individual parameters are better described by a probability distribution or different realizations instead deterministic values. This leads to a probability distribution for all response parameters of reserve estimation. The uncertainty in acquired data and derived parameters is incorporated in geostatistical models. Sensitivity analysis is performed to find the most uncertain variables. The final reserve estimation is reported based on a probability distribution (having an uncertainty interval) rather than a deterministic single value. Rose (2007) shows several merits of a probabilistic methodology over a deterministic estimate for exploration and production. According to Society of Petroleum Engineers (SPE) (Etherington, Pollen, & Zuccolo, 2005) and the Journal of Canadian Petroleum Technology (JCPT) (Robinson & Elliott, 2004), P90, P50, and P10 can be applied in reporting reserves using probabilistic methods to define proved, probable, and possible reserves respectively. The P90 (0.1 quantile) refers to proved reserves, P50 refers to proved and probable reserves, and P10

(0.9 quantile) refers to proved, probable and possible reserves.

A thorough assessment of uncertainty and tracking its evolution with time provides decision support information. Uncertainty in estimation resource/reserve (HIIP) will reduce with more wells. There are some examples of historical look-backs to demonstrate the evolution of different input parameters (porosity, permeability, saturation, etc) along with hydrocarbon in place estimates. Meddaugh, Barge, Todd, and Griest (2007); Meddaugh, Griest, and Barge (2009); Meddaugh, Gross, Griest, Todd, and Barge (2006) and Singh et al. (2013) present look-back studies for original oil in place (OOIP) as a function of time. They show that the uncertainty look-back is useful in tracking the impact of new data. The number of delineation wells could be optimized for OOIP uncertainty management. The slope of uncertainty index (UI) (UI=((P90 OOIP)-(P50 OOIP))/(P50 OOIP) vs. time provides useful information during drilling delineation. If this slope is remain constant, there is little value in drilling more wells. Although the history look back is useful in tracking the impact of new data, it can lead to unreliable decision if the global uncertainty in OOIP is inaccurate (overestimated or underestimated).

1.1.2 Quantifying Uncertainty

There are many papers applying Monte Carlo simulation (MCS) to estimate reserve volumes in early reservoir life (Murtha, 1997). Conditional simulation is proposed to assess and quantify uncertainty of hydrocarbon properties such as porosity, permeability, hydrocarbon pore volume with structural parameters. The methodology is based on simulation with conditioning data considering fixed statistical parameters (Samson, Dubrule, & Euler, 1996). Uncertainty in the univariate distribution of regionalized variables (input parameters) is found to be important. There are a number of approaches that have been proposed to get uncertainty in the univariate distribution.

The bootstrap is the first simplest method of quantifying uncertainty in the histogram developed by Efron (1979). This method uses MCS simulation to draw values from the data distribution to simulate different possible data sets; so, it can be easily applied to calculate the uncertainty in the mean and other statistical parameters.

There are two critical assumptions for applying the bootstrap: 1- The distribution of the data should be representative of the whole domain, and 2- The data are independent. The bootstrap may be useful when it is needed to measure the uncertainty in the mean early in appraisal with widely spaced well data (Pyrcz & Deutsch, 2014).

The spatial bootstrap is proposed by Solow (1985) in order to consider the spatial correlation of data. The spatial bootstrap in geostatistics (Deutsch, 2004; Journel & Bitanov, 2004) applies unconditional LU simulation at the data locations according to spatial correlation of the data (Deutsch & Journel, 1998). This approach considers neither the conditioning data nor the area of interest. Increasing spatial correlation leads to greater uncertainty because the data are more redundant (Khan, Deutsch, & Deutsch, 2014).

The conditional finite domain (CFD) is another stochastic approach that accounts for the conditioning data and the size of the domain. This technique permits evaluation of uncertainty by sampling multiple configurations of the data previously simulated. The configurations should be similar to the configuration of the original data. Generally, CFD is very difficult to operate and leads to low uncertainty because of the conditioning data (Babak & Deutsch, 2009).

The last technique for assessing uncertainty in the mean is using kriging for estimation of the entire domain. This technique called global kriging uses the kriging variance when estimating the entire domain. The variance will decrease when the domain size increases due to the support effect (Deutsch & Deutsch, 2010). This technique is independent of data values and lead to relatively low uncertainty.

Uncertainty in the variogram of regionalized variables should also be quantified and incorporated into final modeling. Different authors (Bogaert & Russo, 1999; Webster & Oliver, 1992) have considered sampling scheme for measuring variogram uncertainty. Cressie (1985), Pardo-Igúzquiza and Dowd (2001), Ortiz and Deutsch (2002), and Marchant and Lark (2004) propose similar expressions related to the covariance matrix of the experimental variogram for each lag distance. The diagonal elements of such a covariance matrix are variances. The covariance matrix can be resolved by quadratic covariance (Matheron, 1965).

Global uncertainty could be computed from the fluctuations between the realizations of simulation (Deutsch, 2005; Journel & Xu, 1994). This uncertainty should be calculated by considering correct parameter uncertainty (uncertainty in input statistics). Derakhshan and Deutsch (2008) proposed a methodology to directly incorporate parameter uncertainty into sequential Gaussian simulation (SGS). The central idea is to use a non-uniform distribution of random numbers for Monte Carlo simulation incorporating the uncertainty in the mean of the variable. The idea seems like an ad-hoc engineering approach with no statistical basis. Alshehri (2010) quantifies global uncertainty of HIIP by assessing parameter uncertainty of the reservoir surface, OWC and reservoir properties (porosity, water saturation etc). A 2D surface of seismic interpretation is assumed as a reference map for considering uncertainty in reservoir volume. This surface can even be attained by a few 2D seismic lines in the absence of 3D seismic (Rezvandehy, 2014). Monte Carlo simulation (MCS) is also used to quantify uncertainty in fluid contacts levels by using the triangular distribution. A cosimulation approach of SGS with super secondary data is utilized for quantifying uncertainty in petrophysical properties for assessing uncertainty in HIIP (Babak & Deutsch, 2007b). There are several issues in the methodology: 1- The parameter uncertainty is only calculated for the mean of the reference distribution without considering the uncertainty in the variance. 2- The parameter uncertainty is calculated by bootstrap, spatial bootstrap and CFD, but there is no guidance on the most accurate approach. Since each method has specific limitations, the correct method of quantifying parameter uncertainty is unknown. 3- The unreliable experimental variogram of each variable is used directly for calculating global uncertainty of HIIP without considering variogram declustering. 4- The variogram uncertainty is not discussed in this study. This uncertainty is important for flow simulation (Meddaugh, Champenoy, Osterloh, & Tang, 2011) and could be improved by auxiliary data and incorporated in the final model. 5- Multivariate parameter uncertainty and transforming this uncertainty in multivariate geostatistical modeling is discussed partially and incomplete in the end.

1.2 Problem Statement

One of the key parameters for geostatistical modeling of petrophysical properties is the variogram for each property. Preferential sampling may bias the spatial structure and often leads to noisy and unreliable variograms. Declustering techniques compensate for the geometric configuration of the data locations (Deutsch, 1989); however, they correct the histogram of data and not the experimental variogram. Some authors consider variogram declustering. There are drawbacks with each technique. A regular data configuration could be taken taken from the cluster data (Chiles & Delfiner, 1999). This leads to zero weights to discarded pairs and loss of information. The average of the univariate declustering weights for two data locations is proposed for variogram declustering (Kovitz & Christakos, 2004; Omre, 1984; Richmond, 2002). However, variogram declustering is only applied for the pairs originating and ending in the same cell. Emery and Ortiz (2005, 2007) discuss variogram declustering by ordinary local kriging. This approach has also some limitations since kriging of the local mean does not give correct global estimation of the mean (declustered mean).

Variogram uncertainty of regionalized variables could also be quantified and incorporated in geostatistical simulation. According to Meddaugh et al. (2011), variogram uncertainty could have a tremendous impact on flow simulation. The current approaches of variogram uncertainty based on sampling scheme (Bogaert & Russo, 1999; Webster & Oliver, 1992) or the covariance matrix of the experimental variogram (Cressie, 1985; Marchant & Lark, 2004; Ortiz & Deutsch, 2002; Pardo-Igúzquiza & Dowd, 2001) are not robust and lead to a very high variogram uncertainty which may bias the final reservoir model. Moreover, the method of incorporating variogram uncertainty in geostatistical simulation is not fully addressed.

There is always a challenge inferring the horizontal variogram in case of widely spaced well data. Applying the horizontal variogram of 3D seismic data is a naive solution to this problem (Wang & Dou, 2010) since the processed seismic data are not the same physical attribute as that under consideration; seismic data are acquired in time domain that should be transferred to depth and there is a much different in the

scale of measurement. Pyrcz and Deutsch (2014) propose an approach of scaling the vertical variogram based on horizontal-to-vertical anisotropy ratio. However, there is no unique ratio for a specific reservoir (Kupfersberger & Deutsch, 1999).

Uncertainty in univariate distributions for each regionalized variable should be quantified correctly. Geostatistical models require input parameters such as univariate distribution for continuous variables and proportions of categorical variables. These parameters are subject to uncertainty because of limited data and measurement error. Each current method of parameter uncertainty gives a different result and has specific drawbacks. The uncertainty in correlation coefficients between variables should be incorporated in the final model.

1.3 Problem Solution and Objectives

The goal of this research is quantifying correct parameter uncertainty and incorporating this uncertainty in geostatistical reservoir models for improved forecasting and management. Figure 1.1 shows a conceptual illustration of L realization of different sources of uncertainty in the reservoir with four well data, and Figure 1.2 shows incorporating all uncertainties (variogram plus histogram, surface, etc) in resource estimation or flow simulation. This leads to lower risk in reservoir decision making and prevents overestimating or underestimating production facilities. Uncertainty in the reservoir is divided into parameter and data uncertainties (see Figure 1.1). Parameter uncertainty includes the variogram, histogram of continuous and discrete (proportion) variables, surface (top and thickness), domain size, etc. Instead of using one fixed set of parameters, different realizations of parameters are used for geostatistical simulation: one realization of parameter uncertainty for one realization of the geostatistical simulation process. Data uncertainty could also be calculated and incorporated in the final model (Barnett, 2015). The focus of this thesis is quantifying correct variogram and histogram uncertainties.

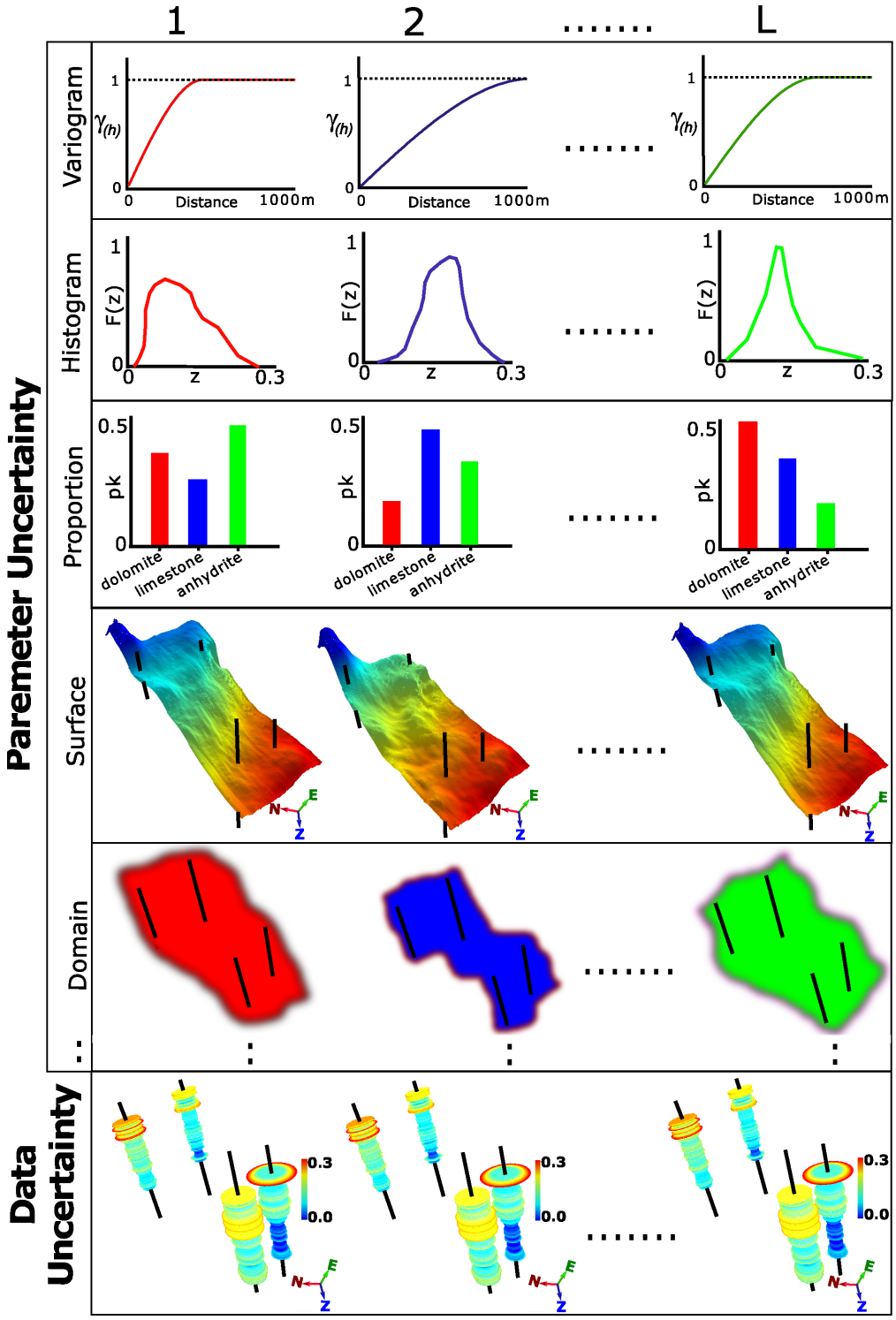

Figure 1.1: A conceptual illustration of parameter and data uncertainties in the reservoir modeling (in presence of four widely-spaced wells). L is number of simulation for each source of uncertainty.

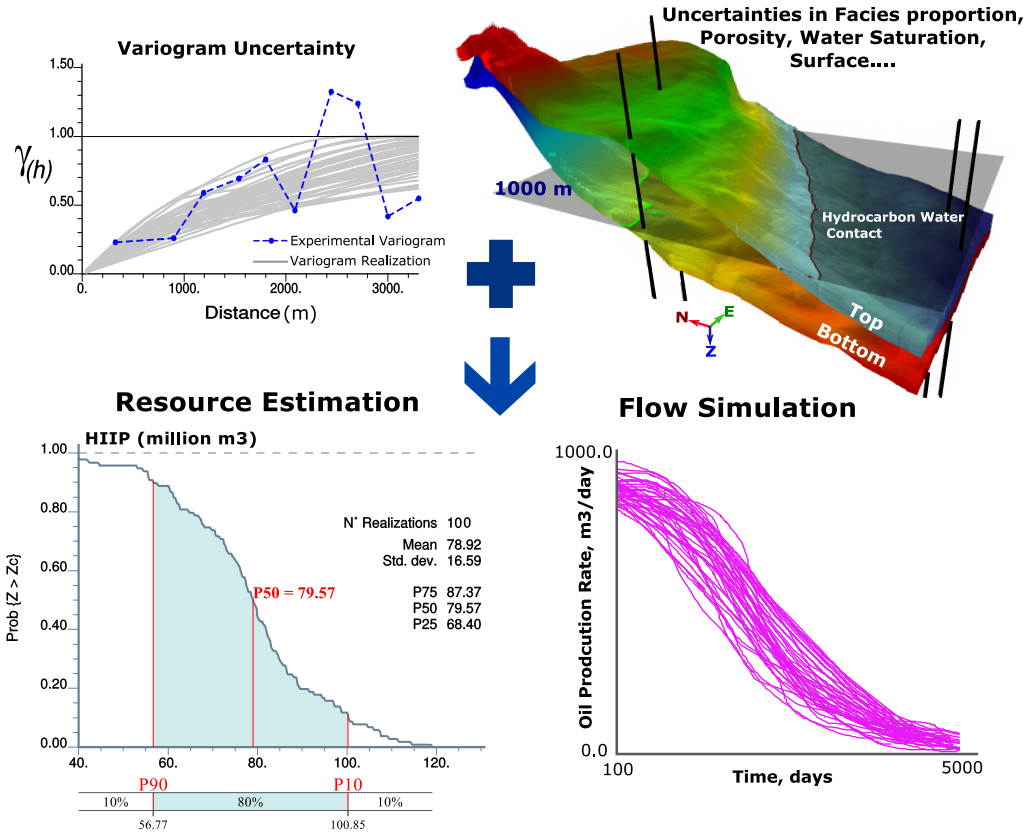

Figure 1.2: A conceptual illustration of resource estimation and flow simulation with all uncertainties.

1.3.1 Representative Variogram with Uncertainty

An accurate variogram model is important for geostatistical modeling and for quantifying parameter uncertainty. The spatial bootstrap needs a reference variogram model, the higher variogram range, the higher the prior histogram uncertainty. The approaches of quantifying variogram uncertainty also need a reference variogram model that should be as accurate as possible. In case of preferential sampling, variogram declustering should be considered. A novel approach of variogram declustering is proposed in this work that corrects the limitation of the previous techniques. The new fitted variogram to the declustered horizontal and vertical variograms (3D variogram model) could be used for geostatistical modeling and quantifying parameter uncertainty.

A new approach of quantifying variogram uncertainty and how to incorporate this uncertainty in geostatistical simulation is proposed in this thesis. In case of sparse

well data, there is high variogram uncertainty in the horizontal variogram. This work also proposes a novel technique to improve the horizontal variogram of well data. The vertical variogram from wells and the horizontal variogram from seismic data are well-defined in most cases, and could be used to improve the horizontal variogram inference from well data: the distributions of uncertainty for each lag distance are merged in order to achieve an improved horizontal variogram. This can be described by the following Steps (Figure 1.3):

1. The horizontal seismic variogram is enforced on the noisy horizontal well variogram. An updated variogram distribution for each lag distance is achieved.
2. The vertical variogram of the well data can be scaled to scenarios of the horizontal variogram using conceptual geological models. A variogram distribution for each lag distance is attained.
3. Merge the probability distributions of each lag distance from Steps 1 and 2.

After merging variogram distributions, improved horizontal variogram realizations are drawn from the uncertainty intervals of lag distances.

1.3.2 Histogram Uncertainty

The next challenge faced by this research is to establish uncertainty in the histogram of regionalized variable (distribution uncertainty). Quantifying this uncertainty is an important task of geostatistical modeling, due to the fact that it affects reservoir investment and development decisions. There are some techniques for quantifying uncertainty in histogram. Each method has limitations (see Section 1.1.2). This work proposes a new simulation-based approach of quantifying histogram uncertainty. An experimental framework is developed, where the true uncertainty in the histogram is known, to evaluate histogram uncertainty approaches. Univariate distribution uncertainty can be extended to multivariate parameter uncertainty for incorporating the correlation between variables in the final model.

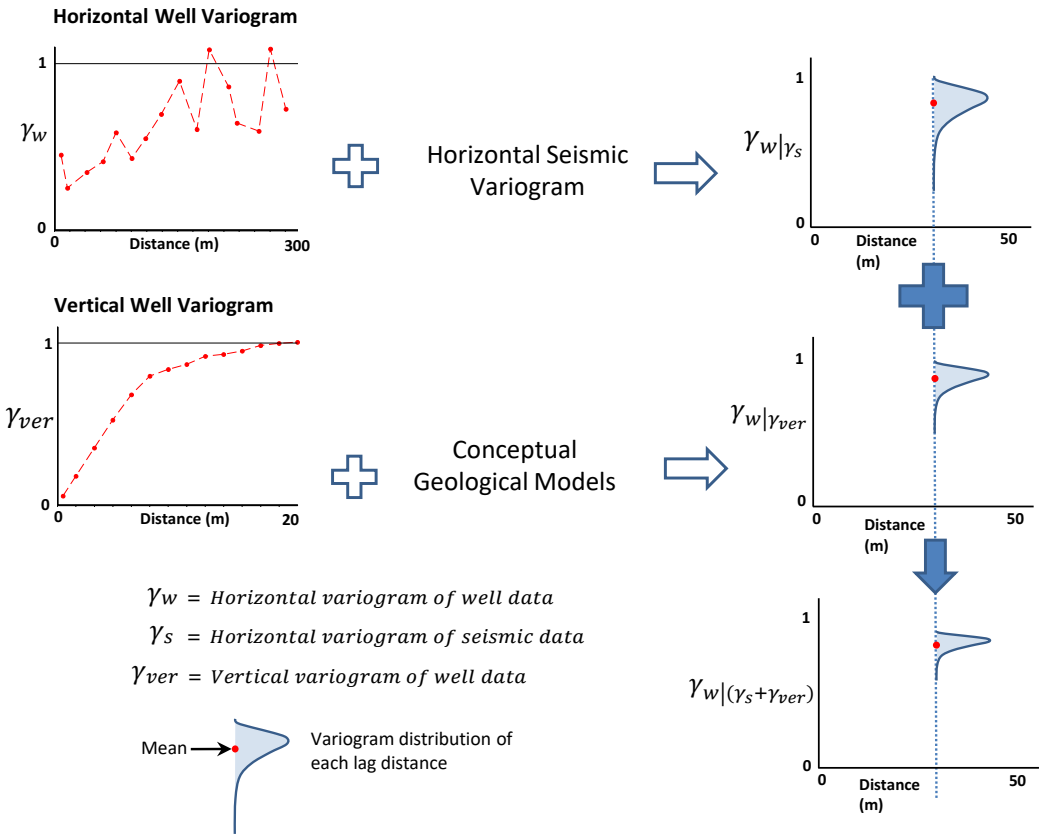

Figure 1.3: Illustrative work flow for improving horizontal variogram of the well data by the horizontal seismic variogram and the vertical well variogram. The distributions of uncertainty for each lag distance are merged to achieve the improved horizontal variogram.

1.4 Thesis Statement and Dissertation Outline

Thesis statement: *Development of techniques to capture correct parameter uncertainty in presence of sparse well data will improve the uncertainty represented in the final geostatistical model.*

The thesis contains eight Chapters. **Chapter** 1 is the introduction; literature review, problem statement and solution are presented. **Chapter** 2 discusses variogram declustering to improve the experimental variogram. **Chapter** 3 demonstrates variogram uncertainty and incorporating this uncertainty in geostatistical modeling. **Chapter** 4 shows implementation to improve the variogram of well data by the vari-

ogram of seismic data. **Chapter** 5 covers final improved variogram realizations (3D) for geostatistical modeling. Multivariate distribution uncertainty is provided in **Chapter** 6. Practical implementation of parameter uncertainty for a real case study is presented in **Chapter** 7. The last **Chapter** 8 wraps up the thesis with conclusions. The description of developed Fortran codes is provided in an **Appendix**.

Chapter 2

Declustering Experimental Variograms by Global Estimation with Fourth Order Moments

The following Chapter presents methodology and examples for variogram declustering in presence of irregular sampling. Preferential sampling may bias the spatial structure and often leads to noisy and unreliable variograms. A novel technique is proposed to weight variogram pairs in order to compensate for preferential or clustered sampling: weighting the variogram pairs gives each pair the appropriate weight, removes noise and minimizes artifacts in the experimental variogram.

The Chapter begins with a discussion on why variogram declustering is required. The principle of variogram declustering and previous works are discussed. A methodology is presented to derive declustering weights of variogram pairs that best compensate for preferential sampling. The methodology is assessed by synthetic and realistic examples.

2.1 Problem of the Equal Weighted Variogram

Geostatistical modeling is widely used to estimate and simulate properties in the petroleum and mining industries. One of the key parameters for geostatistical modeling of continuous variables is the variogram or covariance function for each property. The variogram model provides a three-dimensional definition for pairwise spatial co-

variances (Deutsch & Journel, 1998; Pyrcz & Deutsch, 2014). Variogram modeling is performed to fit an analytical model to a sample variogram computed from the data: the experimental variogram points are not directly used in geostatistical modeling since the variogram function is required for all distances and they must be positive definite (Deutsch & Journel, 1998; Pyrcz & Deutsch, 2014). The variogram model can be attained manually or by autofitting to achieve a variogram model. One conventional approach of autofitting is measuring the goodness of fit by an objective function, which is the sum of the squared difference between the experimental variogram and the modeled variogram. Random changes to the variogram model that decrease the objective function are accepted. This process is repeated many times (say more than 10000) to obtain a variogram model with the minimum objective function (Larrondo, Neufeld, & Deutsch, 2003). For more information of variogram modeling see Chiles and Delfiner (1999); Cressie (1985); Genton (1998); Gringarten and Deutsch (2001); Pardo-Igúzquiza and Dowd (2001).

In practice, variogram modeling is suboptimal in presence of irregular and preferential positioning of the wells (Emery & Ortiz, 2007; Kovitz & Christakos, 2004). Based on geological and geophysical data, the data are likely located in the areas of higher quality that would be developed first or that would maximize production (Pyrcz & Deutsch, 2014). Higher valued areas are often more variable, thus, equal-weighted experimental variograms are often noisy and biased. Weighting the data by declustering techniques compensates for the geometric configuration of the data locations and corrects the statistics (Deutsch, 1989); however, this approach is not normally considered in variogram calculation where equal weights are often considered for the variogram pairs (Emery & Ortiz, 2007). The experimental variogram for a particular lag vector is computed as:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2n(\mathbf{h})} \sum_{i=1}^{n(\mathbf{h})} \left[y(\mathbf{u}_i) - y(\mathbf{u}_i + \mathbf{h}) \right]^2, \quad i = 1, ..., n(\mathbf{h}) \tag{2.1}$$

where $n(\mathbf{h})$ is number of variogram pairs for lag distance $\mathbf{h}$. Each lag of the experimental variogram $\hat{\gamma}(\mathbf{h})$ is a variable comprised of pairs of well data $y(\mathbf{u}_i)$, $y(\mathbf{u}_i+\mathbf{h})$. Figure 2.1-a

shows a synthetic example of clustered well locations in areas of high quality, and sketch of its corresponding variogram for azimuth $0° \pm 10°$ (Figure 2.1-b). The variogram is unstable and has fluctuations because for the short lag distance (lag distance h_1 in Figure 2.1-a) the majority of pairs are located in the high quality area with greater variability, and for a larger distance (lag distance h_2 in Figure 2.1-a) the majority of pairs are located in low quality areas. This unbalanced number of pairs within low and high quality areas leads to a noisy variogram due to the fact that the equal weighted averaging as in Equation 2.1 does not account for preferential sampling. The objective of variogram declustering is to correct the effect of preferential sampling.

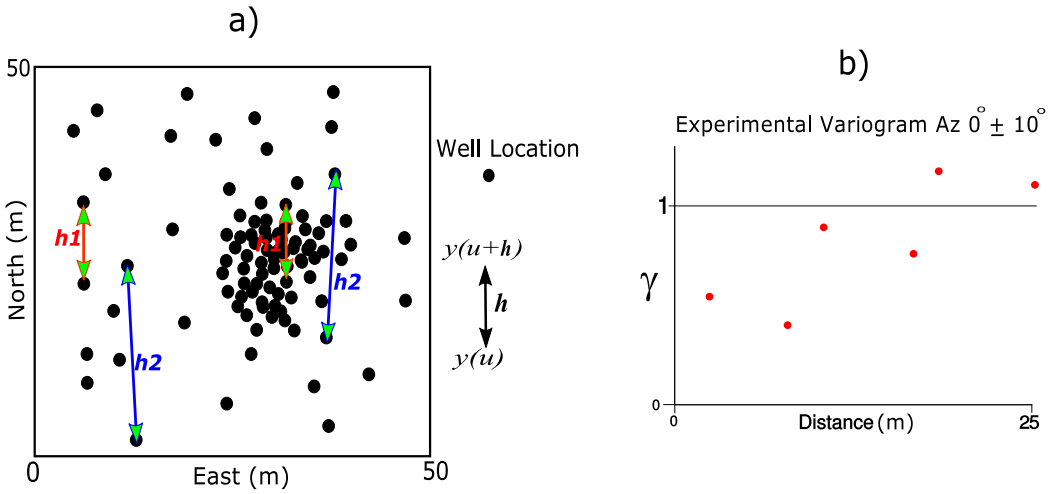

Figure 2.1: a) Synthetic example of clustered locations in areas of high quality (the center right of the sketch). b) Sketch of experimental variogram (not calculated) for azimuth $0° \pm 10°$. The noisy and unreliable variogram is due to clustering some variogram pairs in high valued areas.

2.2 Principle of Variogram Declustering

Estimation of the variogram could consider weighting the variogram pairs:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2n(\mathbf{h})} \sum_{i=1}^{n(\mathbf{h})} w_i \times [y(\mathbf{u}_i) - y(\mathbf{u}_i + \mathbf{h})]^2 \;\; , \;\; i = 1, ..., n(\mathbf{h}) \;\; , \;\; \sum_{i=1}^{n(\mathbf{h})} w_i = 1 \;\; , \;\; w_i \geq 0 \tag{2.2}$$

where w_i are scalar weights assigned to variogram pairs $[y(\mathbf{u}_i) - y(\mathbf{u}_i + \mathbf{h})]^2$. A few approaches based on the configuration of the data have been proposed to calculate these

weights; however, their efficiency has not been proved. Chiles and Delfiner (1999) recommend one data location from the cluster data in order to make a regular data configuration for variogram calculation. This procedure assigns zero weights to discarded pairs. This leads to a loss of information and the variogram calculation for short lag distance may be worse.

Some authors calculate the average of the univariate declustering weights for two data locations and used it in Equation 2.2 (Kovitz & Christakos, 2004; Omre, 1984). Richmond (2002) proposes variogram declustering by univariate cell declustering weight. This algorithm has drawbacks; it only down-weights pairs originating and ending in the same cell. However, in case of omnidirectional variogram, it is likely to have pairs originating and ending in another cell, and it is quite common to have pairs ending in cells where other pairs originate (Emery & Ortiz, 2007). Furthermore, it needs a fixed cell size for cell declustering. Different cell size gives different variogram declustering weights.

Emery and Ortiz (2005, 2007) propose to minimize the variance between the experimental variogram and the underlying true variogram. They use fourth order covariance to calculate optimum declustering weights. The differences between the methodology of this work and their technique is presented in Section 2.7. Olea (2007) selects subsamples of the data that do not have clusters. Although this method is simple to apply, there is a loss of information due to eliminating some variogram pairs. There are other techniques of variogram declustering such as kernel variogram estimator (Menezes, Garcia-Soidán, & Febrero-Bande, 2007) and Box-Cox transformation (Pu & Tiefelsdorf, 2015). None of those mentioned techniques give robust weights for experimental variogram pairs and each has drawbacks. Moreover, it is not feasible to directly calculate the variance of the declustered variogram by those techniques. The goal of this work is to extend the use of declustering techniques to improve the experimental variogram.

2.3 Methodology

Kriging is a linear estimator of a regionalized variable. The weights assigned to the data locations are computed based on the spatial relationship between the data locations and the point being estimated, and the data locations and themselves. Kriging estimates the value at an unsampled location by a maximum number of data within the search radius. Global kriging applies all available data for estimation at either an unsampled location or for the entire domain. If this unknown location is the entire domain, global kriging estimates the declustered global mean and variance. This global kriging approach is proposed for declustering the sample variogram. In this approach, global kriging is applied to the variogram pairs of each lag distance. The "data" value of each pair is half of the quadratic difference of the tail and head values. Global kriging estimates an average of all pairs for the entire area, which is the variogram value for the desired lag distance. This process is repeated for all lag distances to achieve a declustered variogram.

The covariance between variogram pairs should be used in the global kriging system since pairs are the data values. The covariance between two pairs is a fourth order covariance and can be written as a sum of second order moments (Matheron, 1965). This requires an input variogram model as considered by other authors including Ortiz and Deutsch (2002).

Figure 2.2 shows a schematic illustration (not calculated) of variogram declustering by global kriging, which has three Steps: 1) Experimental variogram of the data is calculated and a preliminary model is fitted (Figure 2.2-a). The fitted variogram model is used for computing the fourth order covariance matrix between variogram pairs. 2) Variogram pairs of each lag distance are assembled from the data (Figure 2.2-b). 3) Global kriging is applied on the variogram pairs of each lag distance separately (Figure 2.2-c). The declustered variogram is then fit again and used for geostatistical modeling. This approach also gives a measure of variogram uncertainty at each lag distance. This variogram uncertainty could be incorporated in geostatistical models. Variogram uncertainty is fully discussed in Chapter 3. The main objective of this Chapter is

variogram declustering.

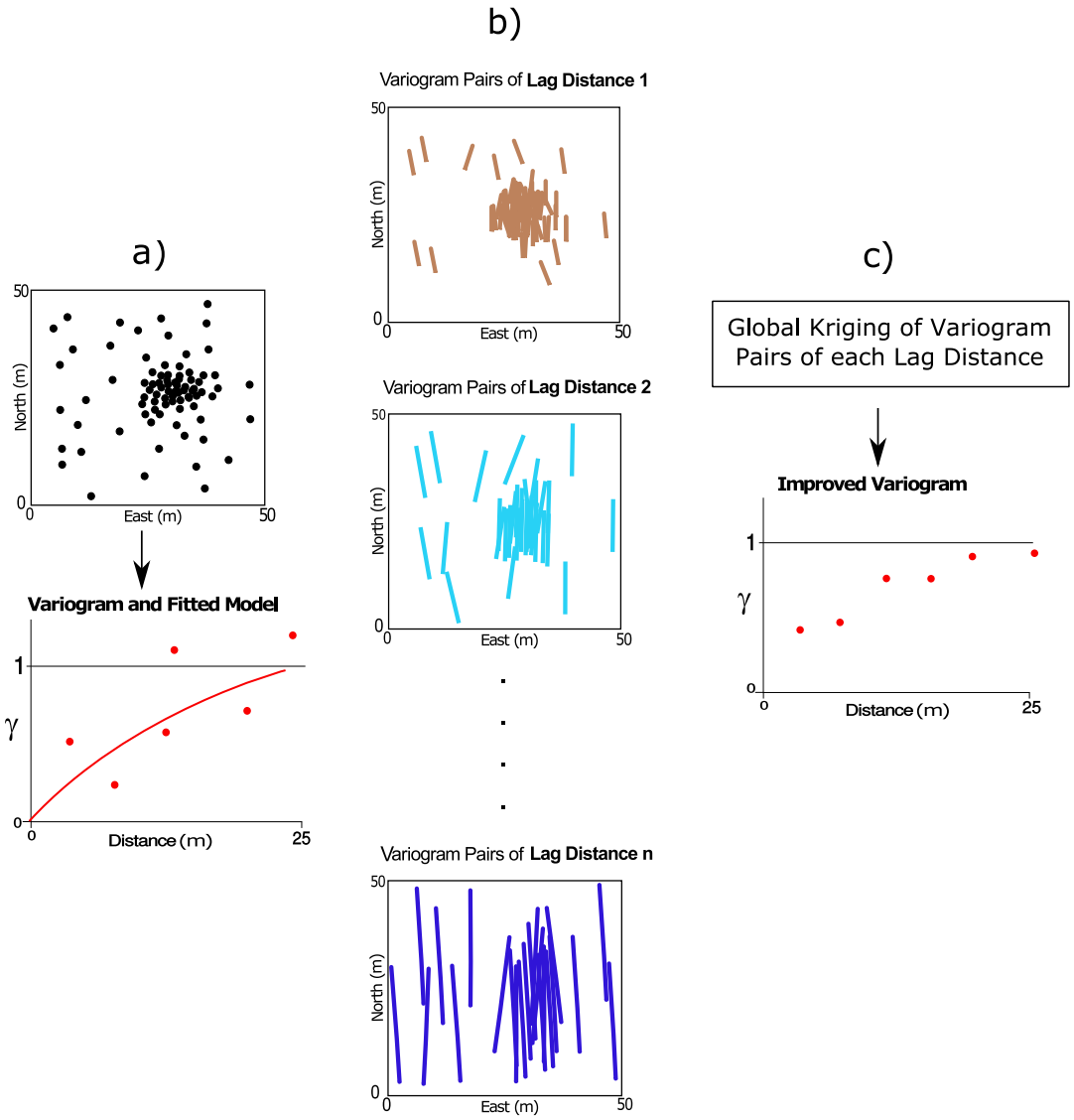

Figure 2.2: Schematic illustration of variogram declustering by global kriging. a) A variogram model is fitted to the experimental variogram of data. b) Draw variogram pairs of each lag distance from the data. c) Apply global kriging for variogram pairs of each lag distance to improve and decluster noisy variogram.

2.4 Global Kriging of Variogram Pairs

Global kriging estimates the declustered global mean and variance provided the domain is assumed stationary (Deutsch & Deutsch, 2010). Let $z(\mathbf{u}_\alpha), \alpha = 1, ..., n$ be the values of variable z at n locations $\mathbf{u}_\alpha$. The ordinary global kriging (OGK) estimator of z for

the entire domain A is written:

$$Z^*_{OGK} = \sum_{\alpha=1}^{n} \lambda_\alpha^{OGK}[Z(\mathbf{u}_\alpha)] \tag{2.3}$$

where λ_α^{OGK} is the declustering weight of global mean assigned to each datum $z(\mathbf{u}_\alpha)$. There is no need to have a stationary mean in the kriging system because ordinary kriging enforces the sum of the weights to 1 by the Lagrange formalism:

$$\sum_{\beta=1}^{n} \lambda_\beta^{OGK} C(\mathbf{u}_\beta - \mathbf{u}_\alpha) + \mu = \overline{C(\mathbf{u}_\alpha - A)} \quad , \quad \alpha = 1....n \quad , \quad \sum_{\alpha=1}^{n} \lambda_\alpha^{OGK} = 1 \tag{2.4}$$

where $C(\mathbf{u}_\beta - \mathbf{u}_\alpha)$ is the covariance between locations $\mathbf{u}_\beta$ and $\mathbf{u}_\alpha$. μ is the Lagrange parameter and $\overline{C(\mathbf{u}_\alpha - A)}$ is average covariance between each data location $\mathbf{u}_\alpha$ and the whole area A. Estimation variance of OGK is written as:

$$Var\left\{Z^*_{OGK}\right\} = \overline{C(A-A)} - \sum_{\alpha=1}^{n} \lambda_\alpha^{OGK} \overline{C(\mathbf{u}_\alpha - A)} - \mu \tag{2.5}$$

where $\overline{C(A-A)}$ is the average covariance of the entire area to itself. Theoretically, the average covariances in Equations 2.4 and 2.5 could be estimated by linear averaging of the calculated covariances after discretization the entire study area A into small blocks (Goovaerts, 1997).

OGK is proposed for declustering variogram pairs. In this approach, global kriging weights are applied to the quadratic differences of pair values:

$$\begin{matrix} Z(\mathbf{u}+\frac{\mathbf{h}}{2}) \\ \uparrow \\ Z(\mathbf{u}-\frac{\mathbf{h}}{2}) \end{matrix} \quad , \quad X_{\mathbf{h}}(\mathbf{u}) = \frac{[Z(\mathbf{u}-\frac{\mathbf{h}}{2}) - Z(\mathbf{u}+\frac{\mathbf{h}}{2})]^2}{2} \tag{2.6}$$

where $Z(\mathbf{u}-\frac{\mathbf{h}}{2})$ is the tail value, and $Z(\mathbf{u}+\frac{\mathbf{h}}{2})$ is the head value of each variogram pair $\uparrow$ for lag distance $\mathbf{h}$. The expected value of the $X_{\mathbf{h}}(\mathbf{u})$ is the variogram value for each lag distance. The expected value for the entire area is estimated by the global kriging

where declustering weights for each lag distance $\mathbf{h}$ are computed by:

$$X^*_{\mathbf{h}} = \sum_{\alpha=1}^{n(\mathbf{h})} \lambda^{OGK}_{\alpha,\mathbf{h}} [X_{\alpha,\mathbf{h}}(\mathbf{u})] \tag{2.7}$$

where $X^*_{\mathbf{h}}$ is the declustered variogram, $n(\mathbf{h})$ is number of variogram pairs for each lag distance $\mathbf{h}$ and $\lambda^{OGK}_{\alpha,\mathbf{h}}$ is variogram declustering weight for pair α and lag distance $\mathbf{h}$ and $X_{\alpha,\mathbf{h}}(u)$ is half of the quadratic differences of the pair values (see Equation 2.6). Global kriging (Equation 2.7) should be applied on variogram pairs of each lag distance in order to decluster and improve the variogram. Since the data are pairs instead of locations, fourth order covariances are used in Equation 2.4 (Ortiz & Deutsch, 2002). The fourth order covariance is briefly discussed in Section 2.5. The system of linear equations of OGK for obtaining $\lambda^{OGK}_{\alpha,\mathbf{h}}$ can be written as:

$$\begin{aligned} &\sum_{\beta=1}^{n(\mathbf{h})} \lambda^{OGK}_{\beta,\mathbf{h}} F(X_{\beta,\mathbf{h}}(\mathbf{u}) - X_{\alpha,\mathbf{h}}(\mathbf{u})) + \mu = \overline{F(X_{\alpha,\mathbf{h}}(\mathbf{u}) - A)} \;\;,\;\; \alpha = 1....n(\mathbf{h}) \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad , \sum_{\alpha=1}^{n(\mathbf{h})} \lambda^{OGK}_{\alpha,\mathbf{h}} = 1 \end{aligned} \tag{2.8}$$

where $F(X_{\beta,\mathbf{h}}(\mathbf{u}) - X_{\alpha,\mathbf{h}}(\mathbf{u}))$ signifies fourth order covariance between pairs $X_{\beta,\mathbf{h}}(\mathbf{u})$ and $X_{\alpha,\mathbf{h}}(\mathbf{u}))$ (left hand side), and $\overline{F(X_{\alpha,\mathbf{h}}(\mathbf{u}) - A)}$ implies average fourth order covariance between each pair $X_{\alpha,\mathbf{h}}(\mathbf{u})$ and the entire domain A (right hand side). The average fourth order covariance could be estimated by discretization of the entire study area A into many variogram pairs for lag distance $\mathbf{h}$. Figure 2.3 shows a discretization of the entire domain A to 39 variogram pairs for lag distance $\mathbf{h}$. The average fourth order covariance is achieved by linear averaging of all covariances between the experimental variogram pair and the discretized variogram pairs (red arrows). In case of exhaustive discretization (more than 2000), the CPU time is an issue to calculate the average fourth order covariances (right hand side covariance matrix).

Uncertainty in the variogram is very important and should be incorporated in the final model. Variogram uncertainty can also be calculated by OGK. The minimized

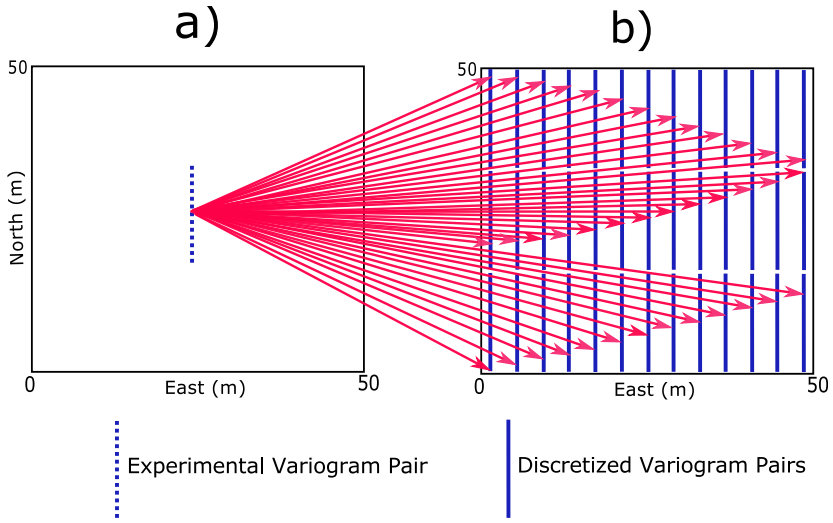

Figure 2.3: Average fourth order covariance between each variogram pair and the entire domain A. a) The experimental variogram pair of lag distance $\mathbf{h}$. b) Discretization of the entire domain to 39 variogram pairs of lag distance $\mathbf{h}$. Average fourth order covariance is achieved by linear averaging of all covariances. Note that there is only one 50m by 50m area; it has been redrawn twice to illustrate the discretization.

estimation variance of each lag distance $\mathbf{h}$ is written as:

$$Var\left\{2X^*_{\mathbf{h}}\right\} = \overline{F(A-A)} - \sum_{\alpha=1}^{n(\mathbf{h})} \lambda^{OGK}_{\alpha,\mathbf{h}} \overline{F(X_{\alpha,\mathbf{h}}(\mathbf{u}) - A)} - \mu = Var\left\{2\hat{\gamma}(\mathbf{h})\right\}$$
$$\rightarrow \quad Var\left\{\hat{\gamma}(\mathbf{h})\right\} = \frac{1}{4} \times \left(Var\left\{2X^*_{\mathbf{h}}\right\}\right) \tag{2.9}$$

where $\overline{F(A-A)}$ is the average fourth order covariance of the entire domain A to itself. Variogram uncertainty is discussed in Chapter 3.

2.5 Fourth Order Covariance

Ortiz and Deutsch (2002) show the fourth order covariance between variogram pairs by second order moments (variogram) under a multi-Gaussian assumption of univariate distribution. As an example, this covariance is calculated with a synthetic example. Figure 2.4 shows two variogram pairs "Pair p" and "Pair q" with lag distances $\mathbf{h}_p$ and $\mathbf{h}_q$. These pairs are a $\mathbf{h}$ distance from each other. The covariance between quadratic differences of variogram pairs $Cov\left\{(y_1-y_2)^2,(y_3-y_4)^2\right\}$ that is the so called fourth

order covariance is to be calculated (Ortiz & Deutsch, 2002). Where y_1 and y_2 are tail and head locations of "Pair p", and y_3 and y_4 are tail and head locations of "Pair q", respectively. According to the definition of the covariance:

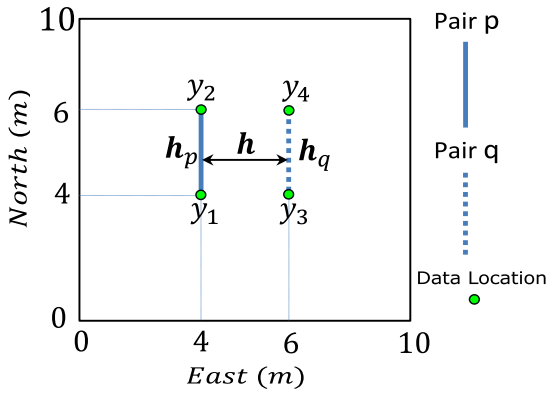

Figure 2.4: Two variogram pairs p and q with lag distances $\mathbf{h}_p$ and $\mathbf{h}_q$. The distance between pairs is $\mathbf{h}$. y_1 and y_2 are tail and head locations of "Pair p", and y_3 and y_4 are tail and head locations of "Pair q", respectively.

$$Cov(Y_i, Y_j) = E\{Y_i.Y_j\} - E\{Y_i\}.E\{Y_j\} \tag{2.10}$$

now, replacing Y_i and Y_j by the quadratic differences $(y_1 - y_2)^2$ and $(y_3 - y_4)^2$):

$$\begin{aligned} Cov\left\{(y_1 - y_2)^2, (y_3 - y_4)^2\right\} = E\left\{(y_1 - y_2)^2.(y_3 - y_4)^2\right\} - \\ E\left\{(y_1 - y_2)^2\right\}.E\left\{(y_3 - y_4)^2\right\} \end{aligned} \tag{2.11}$$

where $E\left\{(y_1 - y_2)^2\right\}$ and $E\left\{(y_3 - y_4)^2\right\}$ are equal $2.\gamma(\mathbf{h}_p)$ and $2.\gamma(\mathbf{h}_q)$, respectively. $\gamma(\mathbf{h})$ is a variogram model fitted to the experimental variogram. Expanding Equation 2.11 leads to a sum of fourth order moments:

$$\begin{aligned} Cov\left\{(y_1 - y_2)^2, (y_3 - y_4)^2\right\} = E\left\{y_1^2.y_3^2\right\} + E\left\{y_1^2.y_4^2\right\} - E\left\{2.y_1^2.y_3.y_4\right\} \\ +E\left\{y_2^2.y_3^2\right\} + E\left\{y_2^2.y_4^2\right\} - E\left\{2.y_2^2.y_3.y_4\right\} - E\left\{2.y_1.y_2.y_3^2\right\} - \\ E\left\{2.y_1.y_2.y_4^2\right\} + E\left\{4.y_1.y_2.y_3.y_4\right\} - 2.\gamma(\mathbf{h}_p).2.\gamma(\mathbf{h}_q) \end{aligned} \tag{2.12}$$

under the multivariate Gaussian distribution, any fourth order moment can be com-

puted by the pairwise covariances (Matheron, 1965):

$$E\{y_1.y_2.y_3.y_4\} = Cov\{y_1, y_2\}.Cov\{y_3, y_4\} + Cov\{y_1, y_3\}.Cov\{y_2, y_4\} + Cov\{y_1, y_4\}.Cov\{y_2, y_3\} \tag{2.13}$$

second order moments (variogram model) are used to estimate the fourth order covariance. Table 2.1 shows the covariances between "Pair p" and "Pair q" (fourth order covariance $Cov(p, q)$) of Figure 2.4 for lag distance $\mathbf{h}_p \approx \mathbf{h}_q \approx 2m$ and $\mathbf{h}$ from 0 to 5m. $\mathbf{h}$ is distance between "Pairs p" and "Pairs q". The initial variogram model for this example is assumed to be an isotropic spherical model with practical range of 5m and sill 1 (no nugget effect). The highest fourth order covariance is for $\mathbf{h}$=0, which is the covariance of each pair to itself. The covariance goes to zero for lag $\mathbf{h}$=5m, which means that there is no linear correlation between these pairs. By replacing fourth order covariance in global kriging approach, variogram declustering weights can be calculated.

Table 2.1: The covariances between "Pair p" and "Pair q" (fourth order covariance) of Figure 2.4 for $\mathbf{h}$ from 0 to 5m.

	h= 0	h = 1m	h= 2m	h= 3m	h= 4m	h= 5m
$Cov(p, q)$	2.5810	0.8717	0.2889	0.0835	0.0127	0.0

2.6 Small Example of Variogram Declustering

A small data set is considered to show how declustering weights can be achieved by solving a fourth order covariance matrix of variogram pairs. Figure 2.5 shows a synthetic data set with nine data locations which leads to six variogram pairs for lag distance approximately $10m$ and azimuth $0° \pm 10°$. Pairs 2 and 3 plus pairs 4, 5 and 6 are clustered. Fourth order covariance matrix for declustering these pairs by OGK

approach is written as:

$$\begin{bmatrix} F(1-1) & F(1-2) & F(1-3) & F(1-4) & F(1-5) & F(1-6) & 1 \\ F(2-1) & F(2-2) & F(2-3) & F(2-4) & F(2-5) & F(2-6) & 1 \\ F(3-1) & F(3-2) & F(3-3) & F(3-4) & F(3-5) & F(3-6) & 1 \\ F(4-1) & F(4-2) & F(4-3) & F(4-4) & F(4-5) & F(4-6) & 1 \\ F(5-1) & F(5-2) & F(5-3) & F(5-4) & F(5-5) & F(5-6) & 1 \\ F(6-1) & F(6-2) & F(6-3) & F(6-4) & F(6-5) & F(6-6) & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ \mu \end{bmatrix} = \begin{bmatrix} \overline{F(1-A)} \\ \overline{F(2-A)} \\ \overline{F(3-A)} \\ \overline{F(4-A)} \\ \overline{F(5-A)} \\ \overline{F(6-A)} \\ 1 \end{bmatrix} \quad (2.14)$$

where left hand side covariance implies fourth order covariance matrix between six

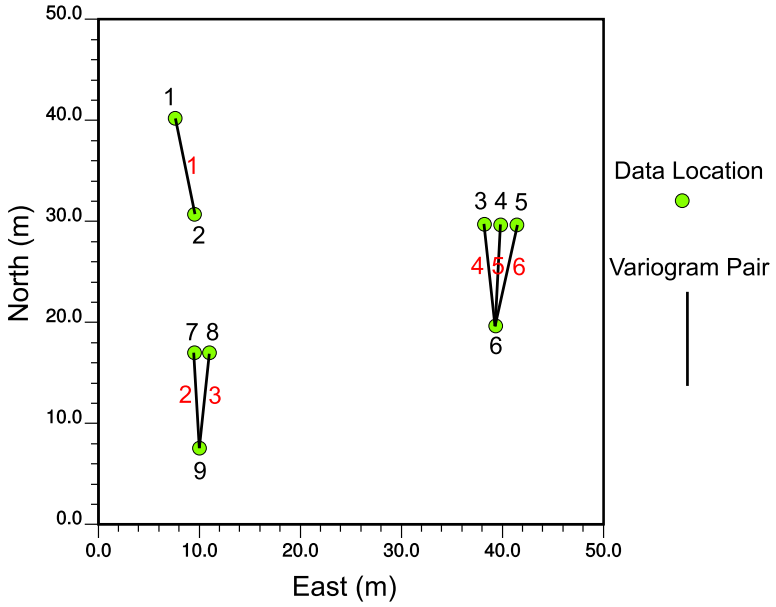

Figure 2.5: Synthetic data set with nine well locations and six variogram pairs for lag distance approximately $10m$ and azimuth $0° \pm 10°$. The numbers signify data locations and variogram pairs. There are nine data locations and six pairs.

pairs written in ordinary kriging format. The right hand side covariance is the average fourth order covariance of each pair to the entire domain. This covariance could be calculated by discretizing the entire study area for the desired lag distance (see Figure 2.3) and calculate linear averaging of the whole fourth order covariances. λ_1 to λ_6 are declustering weights that must sum to 1 by the ordinary kriging approach, and μ is Lagrange parameter. These parameters are calculated by solving the covariance matrix

Equation 2.14 with fourth order covariances calculated from a fitted variogram model. If it is assumed the fitted variogram model is isotropic exponential model with range of 25m and sill 1 (no nugget effect), the left hand side covariance is calculated as:

$$\begin{bmatrix} 3.78 & 0.016 & 0.0159 & 0.285\times10^{-7} & 0.899\times10^{-5} & 0.297\times10^{-4} & 1 \\ 0.016 & 3.71 & 2.87 & 0.376\times10^{-4} & 0.125\times10^{-4} & 0.195\times10^{-5} & 1 \\ 0.0159 & 2.87 & 3.72 & 0.184\times10^{-4} & 0.833\times10^{-6} & 0.416\times10^{-5} & 1 \\ 0.285\times10^{-7} & 0.376\times10^{-4} & 0.184\times10^{-4} & 3.97 & 3.02 & 2.39 & 1 \\ 0.899\times10^{-5} & 0.125\times10^{-4} & 0.833\times10^{-6} & 3.02 & 3.91 & 3.03 & 1 \\ 0.297\times10^{-4} & 0.195\times10^{-5} & 0.416\times10^{-5} & 2.39 & 3.03 & 3.99 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \tag{2.15}$$

and by calculating right hand side covariance, the declustering weights for each pair and Lagrange parameter are achieved:

$$\begin{bmatrix} 0.0346 \\ 0.0343 \\ 0.0342 \\ 0.0341 \\ 0.0365 \\ 0.0355 \\ 1 \end{bmatrix} \Rightarrow \lambda_1 = 0.299\,, \lambda_2 = 0.172\,, \lambda_3 = 0.170\,, \lambda_4 = 0.154\,, \lambda_5 = 0.054\,, \lambda_6 = 0.151\,, \mu = -1.099 \tag{2.16}$$

The calculated declustering weights appear quite reasonable: pair 1 which is not clustered with other pairs receives the highest weight ($\lambda_1 = \mathbf{0.299}$) and pair 5 which is located between pairs 4 and 6 receives the lowest weight ($\lambda_5 = \mathbf{0.054}$). The weights to pairs 2 and 3 are a little higher than weights to pairs 4 and 6 due to smaller number of clustered pairs (two pairs versus three pairs). The weights are positive although there is no constraint that the weights are no negative for OGK (Equation 2.8).

2.7 Comparison with the Methodology of Emery and Ortiz (2005, 2007)

The idea of ordinary global kriging for declustering variogram pairs by fourth order covariance is discussed. The difference between this technique and the methodology applied by Emery and Ortiz (2005, 2007) for variogram declustering is presented in this Section. Emery and Ortiz (2005, 2007) use ordinary kriging of the local mean (Goovaerts, 1997) by fourth order covariance for estimating the global mean (variogram). Kriging of the local mean is used for mapping a smooth picture of the trend (Goovaerts, 1997) and it does not give correct global estimation of the mean (declustered mean). The difference between ordinary kriging of the local mean and the global mean is the right hand side covariance, which is zero for estimating the local mean. The system of linear equations of this approach for obtaining variogram declustering pairs is written as:

$$\sum_{\beta=1}^{n(\mathbf{h})} \lambda_{\beta,\mathbf{h}}^{OLK} F(X_{\beta,\mathbf{h}}(\mathbf{u}) - X_{\alpha,\mathbf{h}}(\mathbf{u})) + \mu = 0 \;\; , \;\; \alpha = 1, ..., n(\mathbf{h}) \;\; , \;\; \sum_{\alpha=1}^{n(\mathbf{h})} \lambda_{\alpha,\mathbf{h}}^{OLK} = 1 \quad (2.17)$$

where $\lambda_{\alpha,\mathbf{h}}^{OLK}$ denotes the ordinary local kriging (OLK) weights for variogram declustering approach by Emery and Ortiz (2005, 2007). In comparison with Equation 2.8, the right-hand side covariance is zero. It means that the estimation location, which is the entire domain A, does not appear in the ordinary kriging system (Goovaerts, 1997). There are some limitations in their work:

1. The approach does not account for the average fourth order covariance between each pair and the entire domain A. The higher average fourth order covariance in left hand side for a variogram pair, the higher variogram declustering weight.

2. According to Goovaerts (1997), ordinary kriging of the local mean approach is aimed at mapping the local mean without considering of the estimation variance. However, Emery and Ortiz (2005, 2007) derive the estimation variance in terms of disjunctive kriging to $\frac{-\mu}{2}$, or $\frac{-(\mu+\nu)}{2}$ if there is another Lagrange parameter

ν for additional constraint. Compared with the variance of OGK (Equation 2.9), the calculated variance is independent from the area of interest. This may lead to overestimation of the variance. For example, the estimation variance for variogram declustering in Figure 2.5 is 0.267 by OGK, and 0.547 by OLK using only one constrain μ.

3. There is no comparison between the declustered and the real variograms in their work. Variogram declustering is not reliable unless it is evaluated by the true variogram of the exhaustive sampling of the data set. This can be done by synthetic examples, or realistic examples such as the data "cluster data" (Deutsch & Journel, 1998) in which the exhaustive sampling "true.dat" is available.

4. The sill of the declustered variogram is not defined. After declustering variogram, another variogram model should be fitted to the declustered variogram with a known sill. Otherwise, the fitted variogram model might be biased because it might violate the correct stationary variance of data after declustering. In Section 2.8, the sill of the declustered variogram is defined.

OGK improves the limitations of OLK by Emery and Ortiz (2005, 2007) for variogram declustering. Synthetic and realistic examples confirm this improvement.

2.8 Sill of the Declustered Variogram

The sill of the declustered variogram may be different than the sill of the experimental variogram (Figure 2.6) due to the unequal weighting of the variogram pairs. The sill of the experimental variogram is important for variogram standardization and interpretation.

Figure 2.7 shows a methodology for calculating the sill of the declustered variogram. There is a data location $\mathbf{u}$ in this Figure. This location could be tail or head locations of variogram pairs for different lag distances. The corresponding variogram pairs of this location receive the declustering weights $w_1, ..., w_n$. Half of each variogram declustering weight could be assigned to location $\mathbf{u}$ $(\frac{w_1}{2},...,\frac{w_n}{2})$. The final received weight for

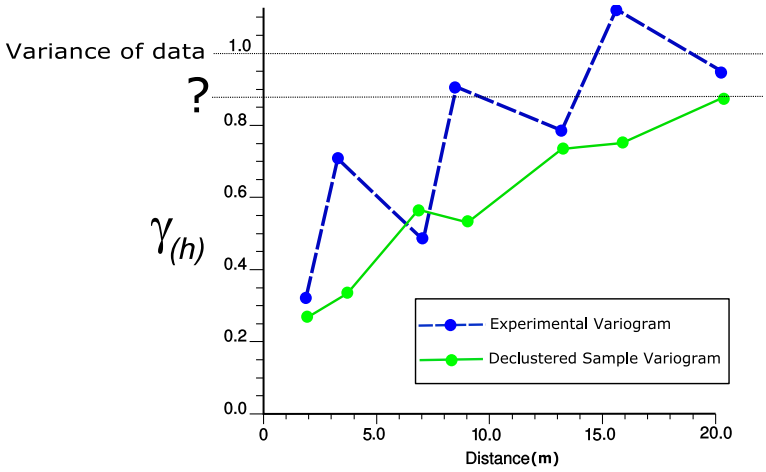

Figure 2.6: A schematic of unknown sill of the declustered sample variogram that is likely to be less than the variance of data.

location **u** could be written as the average of all received weights:

$$W = \frac{\sum_{i=1}^{n} \frac{w_i}{2}}{n} \tag{2.18}$$

where n is number of all variogram pairs and W is final assigned weight to location **u** calculated by all variogram declustering weights. This should be done for all locations of data. So, declustering weight for each data location comes from variogram declustering. These weights should sum to 1. The resulting variance using these declustering weights is the approximate sill of the declustered sample variogram:

$$m = \sum_{j=1}^{nd} W_j z_j \quad , \quad \sigma^2 = \sum_{j=1}^{nd} W_j (z_j - m)^2 \tag{2.19}$$

where z_j , $j = 1, ..., nd$ are data values and W_j are the weights from Equation 2.18 for each location **u** of data. σ^2 is the variance of data by the variogram declustering weights. This is the sill of the declustered variogram.

2.9 Numerical Approach

The applicability of the proposed variogram declustering approach is now verified by considering some numerical examples. Figure 2.8 shows a two-dimensional sampling

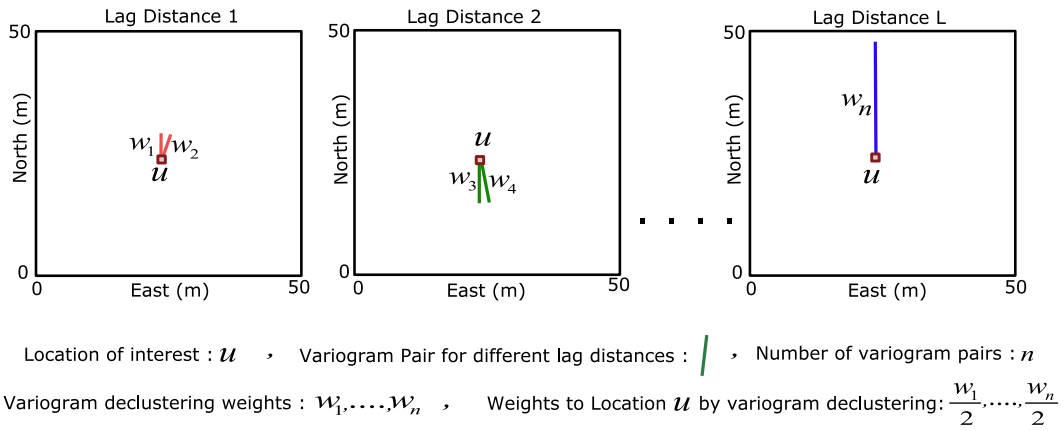

Figure 2.7: Schematic of variogram declustering weights assigned to data location **u**.

data configuration with 100 irregularly spaced data locations. Data values at these locations are drawn by LU unconditional simulation (Davis, 1987; Deutsch & Journel, 1998) of a stationary random field having an isotropic exponential variogram model with range of 25m and sill 1.

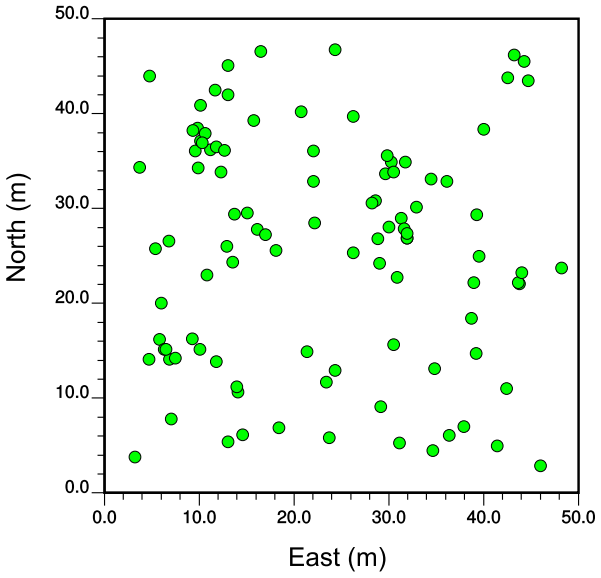

Figure 2.8: Synthetic data set with 100 data locations.

To assess the quality of variogram estimators (experimental and declustered sample variograms), 100 realizations at each data location are constructed. Then, the experimental and declustered sample variogram are calculated for each realization, and the distribution of each lag is built. The variance of each lag distance of these variograms

are compared with each other since the means are approximately equal, converging to the pre-specified variogram model after 100 realizations. The lower variance, the more precise the variogram estimator.

For comparison, both the experimental and declustered sample variograms are calculated after weighting the data locations by cell declustering technique (Deutsch, 1989; Journel, 1983) and normal score the data using declustering weights. The experimental variogram is standardized by the variance of data with the declustering weights. The fitted variogram model for each realization is attained by autofitting (auto variogram modeling) a variogram model to the experimental variogram. The declustered variogram should be standardized based on the calculated variance by the variogram declustering weights for each location. This is for making the experimental variogram and the declustered variogram consistent (both have sill 1) to verify the variogram declustering technique. Cell declustering of the data location leads to the variance lower than 1. This variance is used to standardize the experimental variogram of normal score data using cell declustering weights. By declustering the experimental variogram, the variance will be lower again. This variance could be calculated by Equation 2.19 and used to standardize the declustered variogram to sill 1. Figure 2.9 shows the variance of the experimental and declustered sample variograms for seven lag distances 1m, 3m, 6m, 9m, 12m, 15m, 18m after LU unconditional simulation of data locations in Figure 2.8 for 100 realizations. The variance of the declustered variograms for all lag distances is much lower than the experimental variograms. This difference is going to increase from short to large lag distances. Hence, variogram declustering leads to a more accurate variogram. According to this example, the variance of the declustered variogram decreases as the lag distance increases, and it reaches to a maximum variance of 0.032 at lag distance 6m and then it stays constant until a lag distance 18m (see Figure 2.9). It can be concluded that in general the declustered variogram is more reliable.

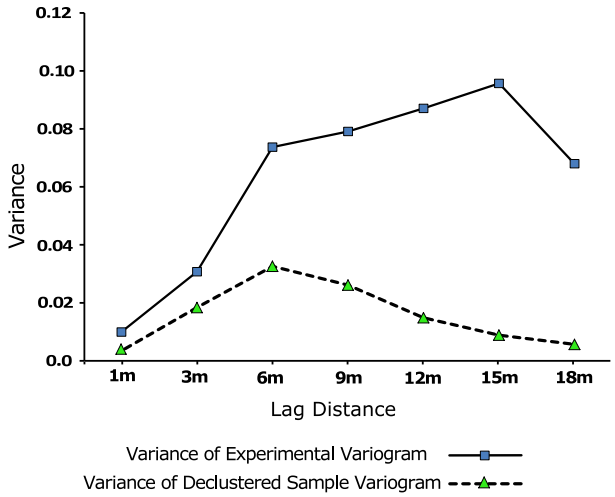

Figure 2.9: Variance of the experimental and declustered sample variograms for seven lag distances 1m, 3m, 6m, 9m, 12m, 15m, 18m achieved by 100 realizations of LU unconditional simulation for data locations in Figure 2.8.

2.10 Synthetic Examples

Two synthetic two-dimensional data sets are considered to assess the proposed variogram declustering technique by comparing the experimental and the declustered sample variograms. The methodology of Emery and Ortiz (2005, 2007) is compared with OGK. For this assessment, reference normal score data is first simulated and then some data are sampled from the reference data. The experimental and declustered variograms of the sampled data are compared with the variogram of the reference data, which is the true variogram of the sampled data. Both the experimental and declustered variograms are calculated after weighting the well data (sampled data) by cell declustering technique (Deutsch, 1989; Journel, 1983) and normal score the data using declustering weights. The variograms are standardized.

Figure 2.10 shows the first example. Figure 2.10-a is the simulated reference data and Figure 2.10-b is location map of 100 data sampled from the reference data. The experimental variogram and the declustered variogram of the sample data in Figure 2.10-b are calculated for azimuth $0^\circ \pm 10^\circ$. A variogram model is fitted to the experimental variogram. This fitted variogram model is used for variogram declustering. Figure 2.11 shows the comparison between the experimental and declustered sample variograms

with the real variogram of the sample data. The real variogram is the variogram of the reference data in Figure 2.10-a for azimuth 0°. Figure 2.11 demonstrates that the declustered variogram by OGK is closer to the real variogram, it eliminates noise and unrealistic fluctuations in experimental variogram especially after a lag distance 12m. It shows the trend in the reference data at an azimuth of 0°. The declustered sample variogram by Emery and Ortiz (2005, 2007), which is standardized by the variance of normal score data after cell declustering, is far from the real variogram.

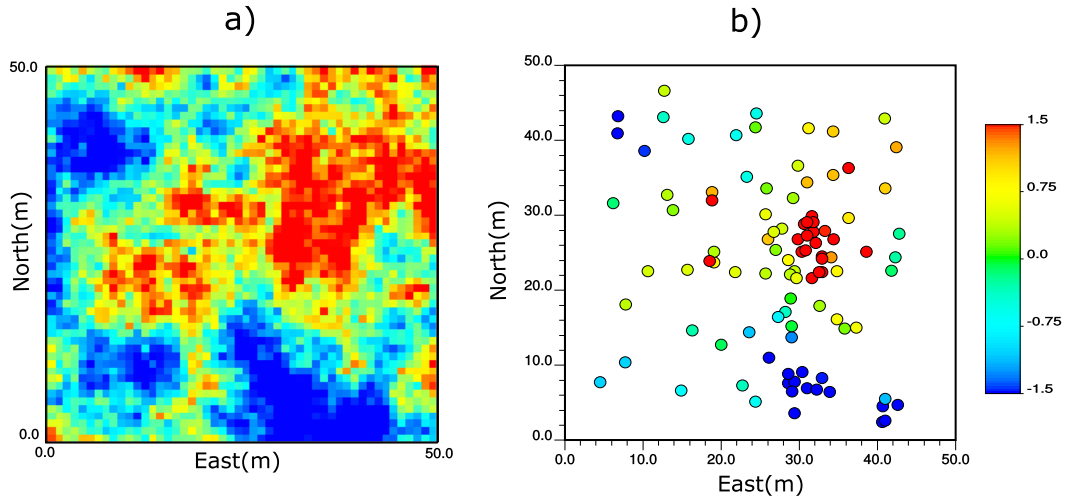

Figure 2.10: a) Reference data. b) Location map of 100 samples taken from the reference data (a).

Figure 2.12 shows a second example with 150 samples. Similar to the previous example, variograms are calculated for azimuth $0° \pm 10°$. Figure 2.13 shows the comparison between the experimental and declustered variograms with the real variogram. Figure 2.13 confirms that the declustered variograms work and it shows that for this example, OGK and the approach by Emery and Ortiz (2005, 2007) are quite similar although OGK is slightly closer to the real variogram. The right hand side covariance matrix in this case are similar or close to zero. Moreover, the sill of the declustered variogram is approximately equal the variance of the normal score data after cell declustering. This data set shows that under certain conditions OGK approach may give similar results as approach of Emery and Ortiz (2005, 2007).

The declustered variogram (both OGK and methodology of Emery and Ortiz (2005,

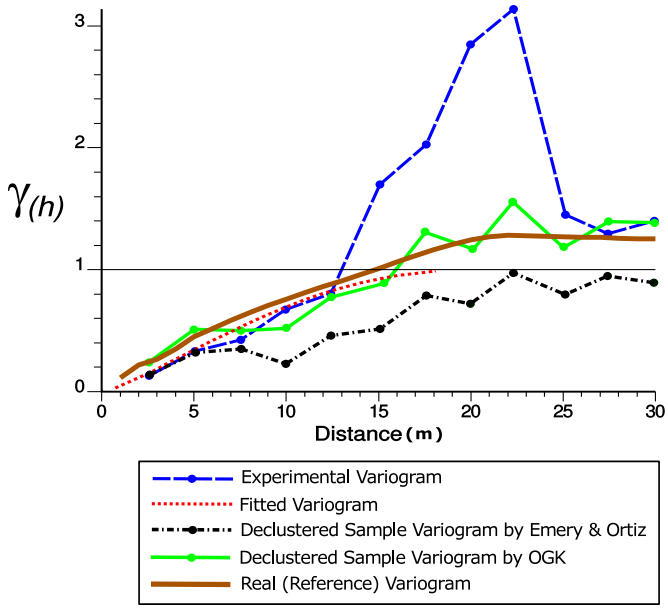

Figure 2.11: Comparison between the experimental and declustered sample variograms of data (normal score) in Figure 2.10-b for azimuth $0° \pm 10°$ with their real variogram (variogram of the reference data in Figure 2.10-a).

2007)) shows the zonal anisotropy in the reference data for azimuth 0°; and it is different from the fitted variogram. Therefore, the declustered variogram does not necessarily return the fitted variogram just because it is used for the declustering weights.

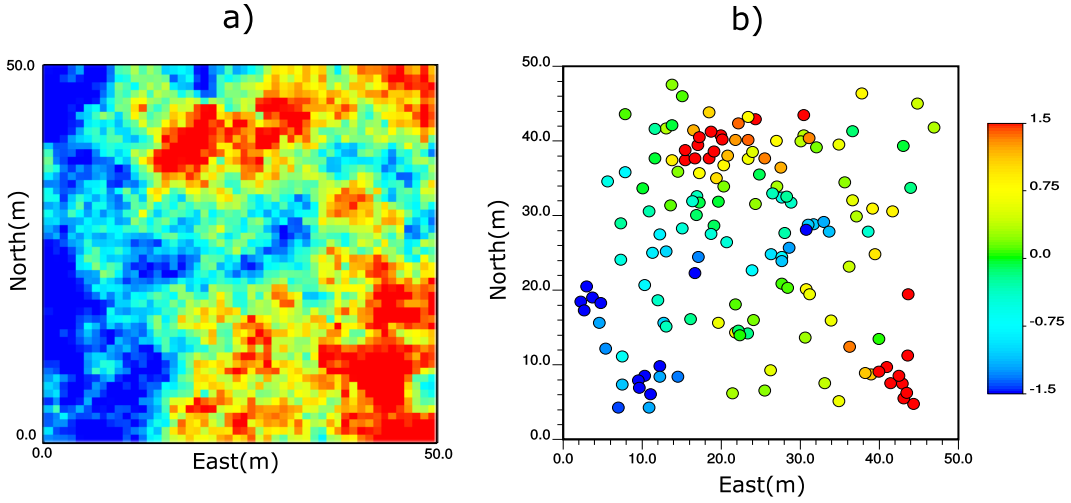

Figure 2.12: a) Reference data. b) Location map of 150 samples taken from the reference data (a).

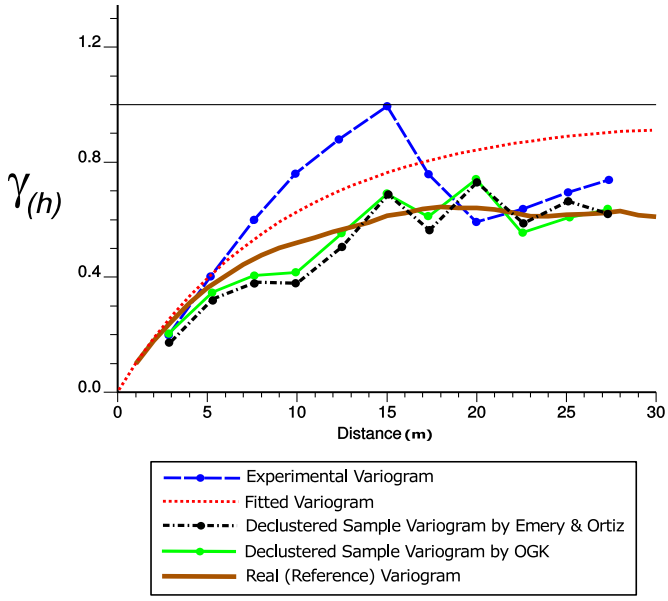

Figure 2.13: Comparison between the experimental and declustered variograms of data (normal score) in Figure 2.12-b for azimuth $0° \pm 10°$ with their real variogram (variogram of the reference data in Figure 2.12-a).

2.11 Realistic Examples

2.11.1 The Cluster Data Set

The data "cluster.dat" taken from the literature (Deutsch & Journel, 1998) is now considered. The exhaustive sampling of this data set called "true.dat" is available. Hence, it can be compared with the experimental and declustered variograms. The experimental and declustered variograms are calculated and standardized after weighting the data by the cell declustering technique and normal score transferring the data using declustering weights. For this example, variogram declustering by two-point cell declustering approach (Richmond, 2002) is also calculated although not recommended by Emery and Ortiz (2007). Two-point cell declustering (TPCD) weights for each variogram pair is calculated by averaging cell declustering weights of tail and head locations of that pair. The sum of declustering weights is 1.

Figure 2.14-a shows location of cluster data with 140 samples. Figure 2.14-b shows exhaustive sampling of cluster data with 2500 samples. The unit of measurements for coordinate system is assumed kilometer (km). Figure 2.15 shows the declustered vari-

ograms by Emery and Ortiz (2005, 2007), OGK, TPCD and the experimental variogram versus real variogram for azimuth $0° \pm 20°$. The fitted variogram model is isotropic with two spherical structures with variance contributions 0.44 and 0.55, and ranges of 10km and 20km for each structure, and nugget effect 0.01. The result is similar to the synthetic examples. The declustered sample variogram by OGK is a better estimate of the true variogram. TPCD approach is close to the experimental variogram. The declustered sample variogram by Emery and Ortiz (2005, 2007) is lower than the real variogram due to the fact that it is standardized with the variance of normal score data, which is higher than the variance of the declustered variogram.

Uncertainty in the estimated variogram (declustered variogram) is very important and should be incorporated into final modeling. The variance of each lag distance could be calculated by Equation 2.9 for OGK, and $\frac{-\mu}{2}$ by Emery and Ortiz (2005, 2007) derived the estimation variance in terms of disjunctive kriging. μ is Lagrange parameter in Equation 2.17. Figure 2.16 shows variogram uncertainties (variances) of each lag distance for these two approaches. OGK gives a reasonable variogram uncertainty while the methodology of Emery and Ortiz (2005, 2007) leads to a very high uncertainty. Incorporating this high uncertainty into geostatistical modeling could bias the results.

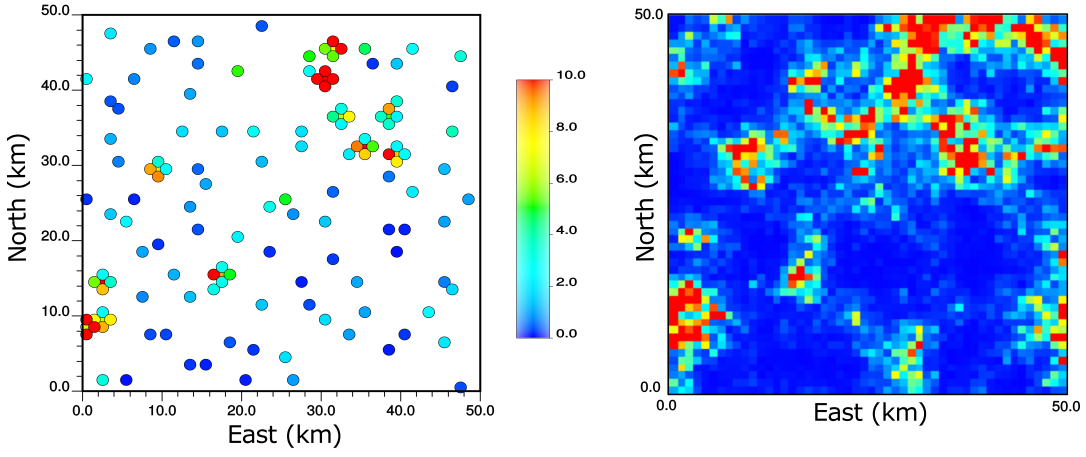

Figure 2.14: a) Location map of cluster data (140 samples). b) Exhaustive sampling of cluster data with 2500 samples.

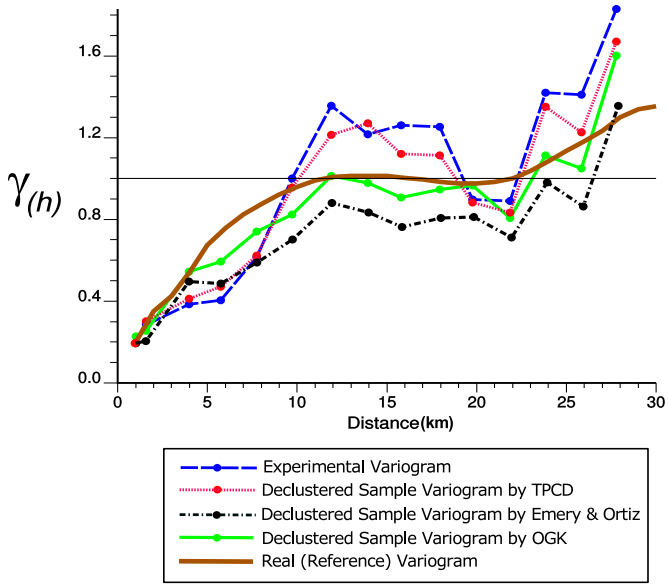

Figure 2.15: The experimental and real variograms (normal score) versus the declustered sample variograms by Emery and Ortiz (2005, 2007), OGK (ordinary global kriging) and TPCD (two-point cell declustering) for azimuth $0° \pm 20°$.

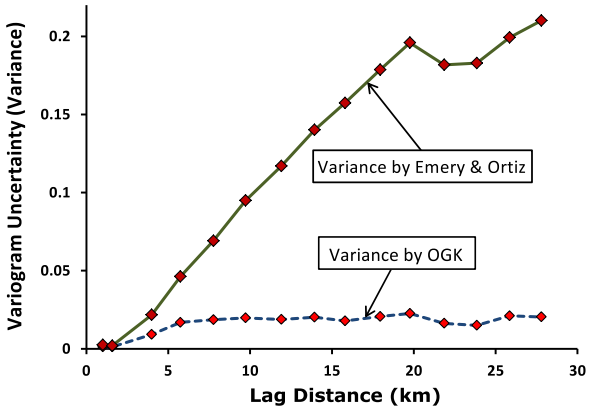

Figure 2.16: Variogram uncertainty (variance) of each lag distance for cluster data resulted by Emery and Ortiz (2005, 2007) and OGK (Equation 2.9).

2.11.2 The Jura Data Set

Another realistic example is considered although there is no exhaustive sampling to compare the declustered variogram with the real variogram. The Jura data set was collected by Swiss Federal Institute of Technology at Lausanne. There are 359 data locations for concentration of seven heavy metal in the topsoil (Goovaerts, 1997). Variogram declustering is considered for the variables nickel (Ni), lead (Pb) and zinc (Zn)

for azimuth $0° \pm 20°$ and azimuth $90° \pm 20°$. The experimental and declustered variograms are calculated and standardized after weighting the data by cell declustering technique and normal score the data using declustering weights. Only OGK approach is considered for variogram declustering of the Jura data set as it was showed that this approach is a better estimation of the real variogram in comparison with other techniques.

Figure 2.17-a shows location map for nickel (Ni) concentration and Figures 2.17-b and c show the experimental, fitted and declustered variograms for azimuth $0° \pm 20°$ and azimuth $90° \pm 20°$, respectively. The declustered variograms especially for azimuth $0°$ (Figures 2.17-b) improves the experimental variogram. The declustered variogram is fit again and used for geostatistical modeling. Figures 2.18 and 2.19 show location maps (a) and the declustered variograms for lead (Pb) and zinc (Zn) concentrations for azimuth $0° \pm 20°$ (b) and azimuth $90° \pm 20°$ (c). The new fitted variograms to the declustered variograms for azimuth $0°$ will have lower ranges than the primary fitted variograms, and higher ranges for azimuth $90°$ for all metal concentrations. It demonstrates again that the declustered variogram does not necessarily return the fitted variogram.

2.12 Remarks

The weighted experimental variogram is a useful tool to improve the determination of the spatial structure of a data set in the presence of preferential sampling. Ordinary global kriging (OGK) of the entire domain is proposed for declustering variogram pairs. Global kriging is applied on the variogram pairs of each lag distance by fourth order covariance. The value of each pair is assigned as half of the quadratic differences of tail and head values. Global Kriging estimates the average of all variogram pairs for the entire domain. This process is repeated for variogram pairs of all lag distances to decluster the variogram. Although only 2D data sets are considered to assess the methodology, it can be applied to 3D data with vertical or horizontal wells. The declustered 3D variogram is achieved by declustering the horizontal and vertical variograms.

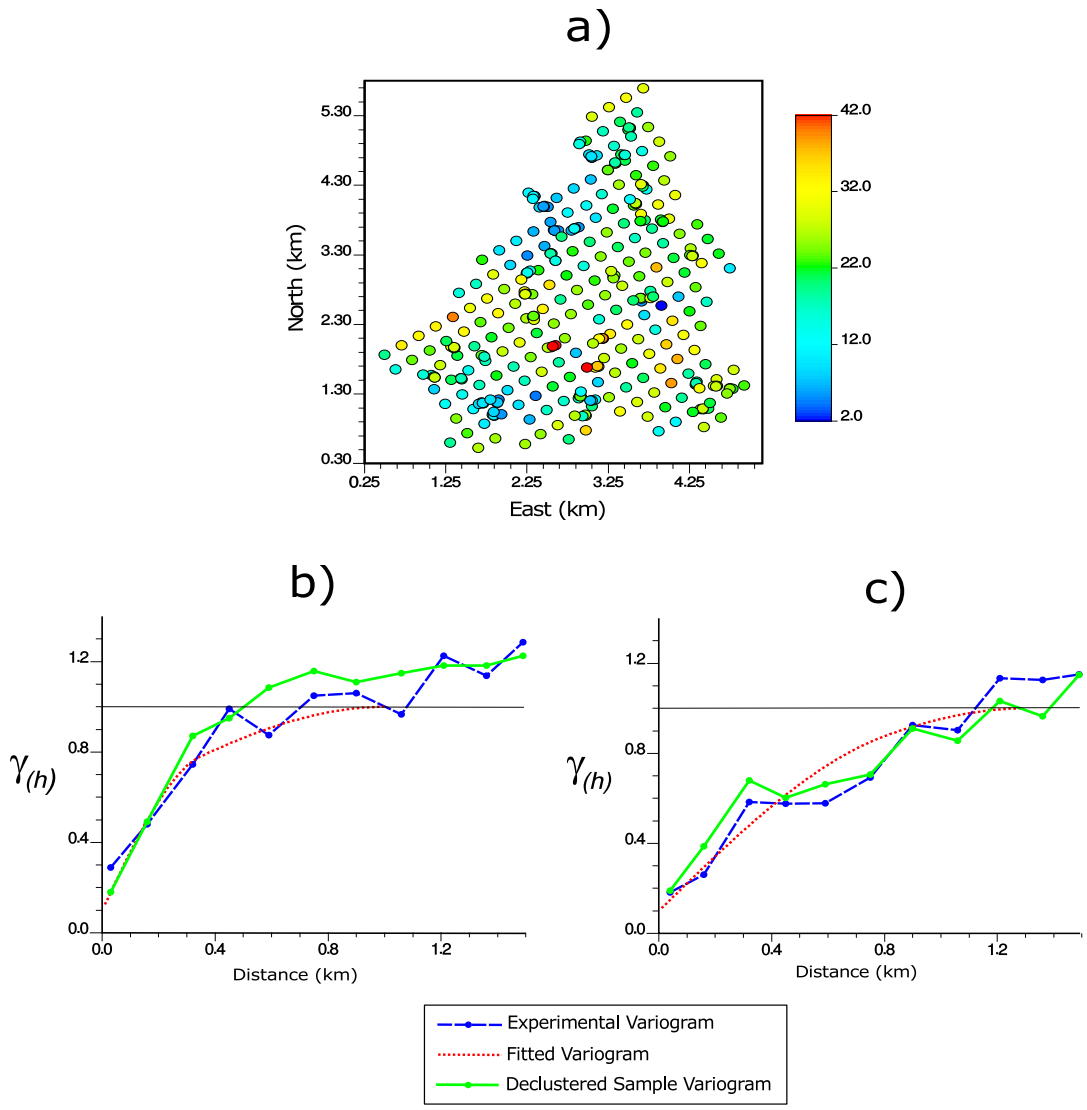

Figure 2.17: a) Location map of nickel (Ni) concentration for the Jura data set. b) The experimental and declustered variograms (normal score) for azimuth $0° \pm 20°$. c) The experimental and declustered variograms (normal score) for azimuth $90° \pm 20°$.

The new fitted variogram to the declustered horizontal and vertical variograms (3D variogram model) is used in geostatistical modeling.

The following Steps are required for the proposed framework of variogram declustering by global kriging: 1) Univariate declustering of the data (cell declustering is recommended) and transform to normal score using declustering weights. 2) Check multiGaussianity (Von Eye & Bogat, 2004). Gaussianity of the univariate distribution is not enough. Fourth order moments are calculated based on multigausianity

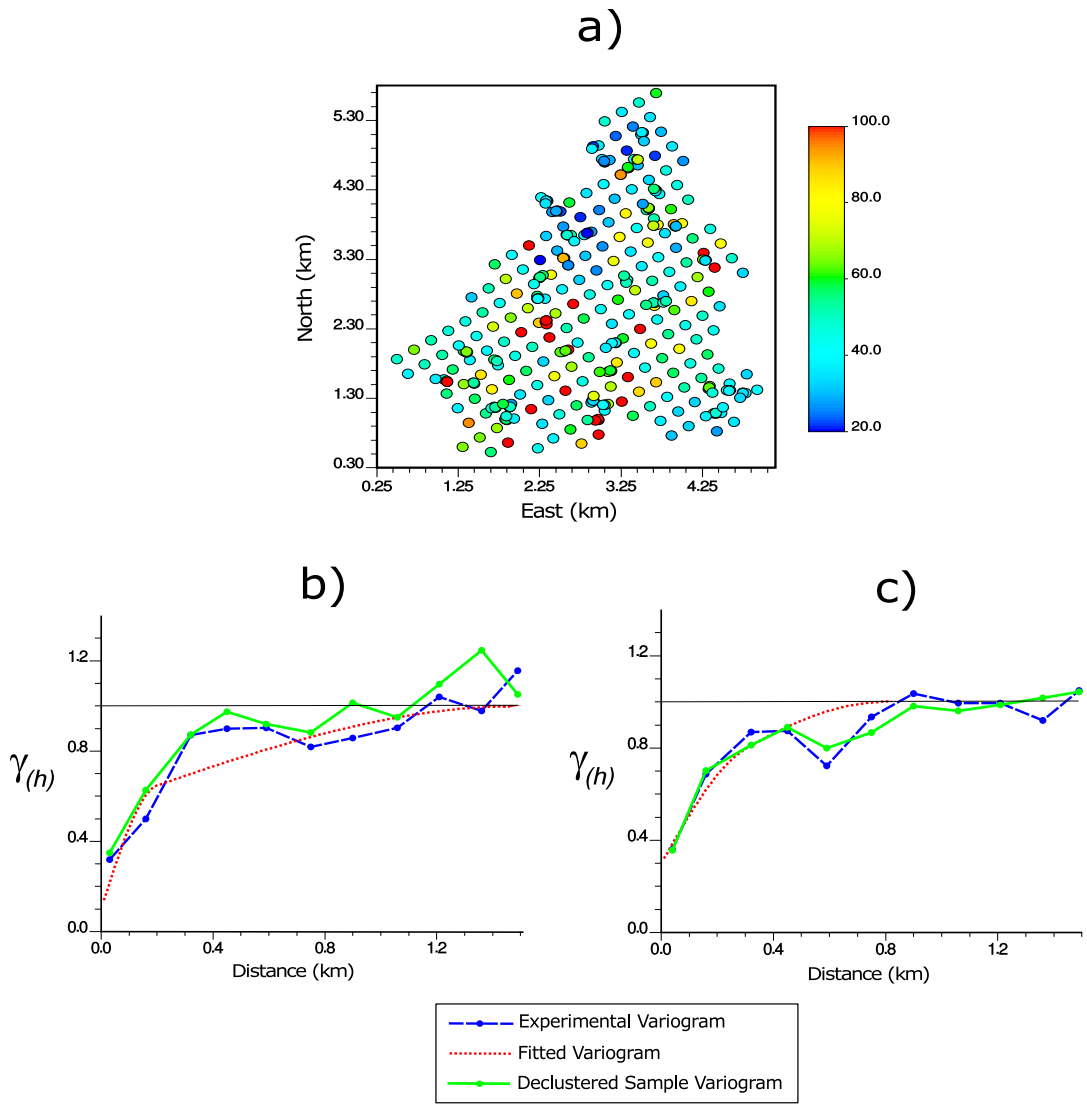

Figure 2.18: a) Location map of lead (Pb) concentration for the Jura data set. b) The experimental and declustered variograms (normal score) for azimuth $0° \pm 20°$. c) The experimental and declustered variograms (normal score) for azimuth $90° \pm 20°$.

assumption although this is not a strong assumption. There are some tests to check biGaussianity (Deutsch & Deutsch, 2009). 3) Calculate the experimental variogram for the desired direction and fit a preliminary variogram model to that direction. It is recommended to have a directional fitted model for variogram declustering of each direction. The variogram declustering weights to each pair are influenced by the variogram ranges as well as the configuration of pairs. Using a fitted variogram seems circular for variogram estimation. However, the synthetic and realistic examples show

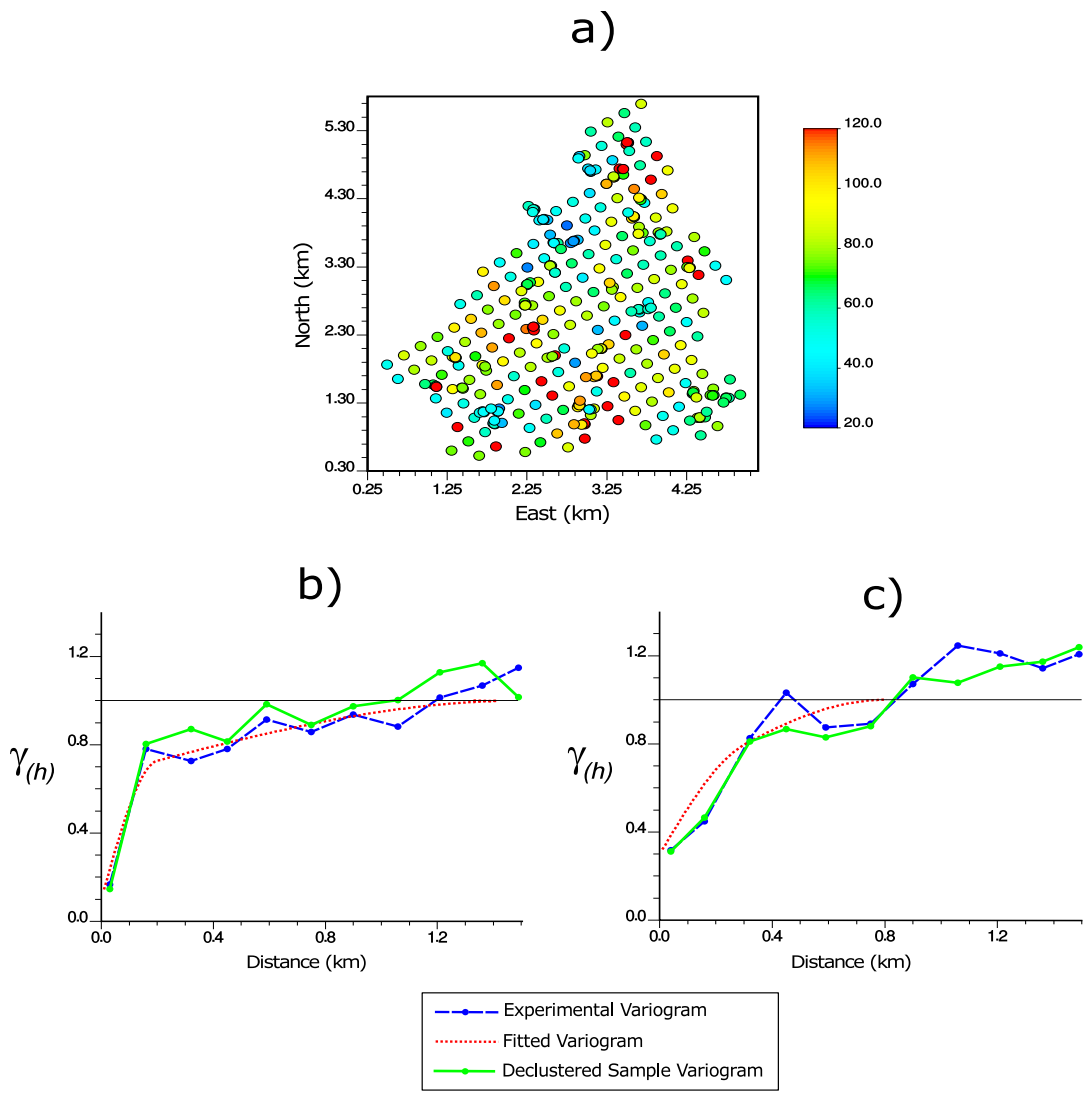

Figure 2.19: a) Location map of zinc (Zn) concentration for the Jura data set. b) The experimental and declustered variograms (normal score) for azimuth $0° \pm 20°$. c) The experimental and declustered variograms (normal score) for azimuth $90° \pm 20°$.

that the declustered variogram is not necessarily close to that fitted variogram; the fitted variogram is only used for the weighting and is not used directly as a target for declustering.

The fourth order covariance matrix between variogram pairs is positive definite. However, as far as programming is concerned, this covariance may be close to a singular matrix in presence of exhaustive variogram pairs (more than 5000 pairs). This happens because of numerical precision. One solution is to use Higham's algorithm (Higham,

1988) to compute a nearest symmetric positive semidefinite matrix. Another simple solution is to add a very small value to the diagonal elements before solving the ordinary kriging. This small value will stabilize the results with no significant affect on the declustering weights.

The global kriging is CPU intensive with very large sets of data because of matrix inversion (Neufeld & Wilde, 2005). This problem is worse for variogram declustering because of using the fourth order covariance and solving the covariance matrix for each lag distance. This method is currently applicable with less than a few thousand variogram pairs. It would not be advisable to consider variogram declustering in presence of more than 10000 variogram pairs for each lag distance. Furthermore, the CPU time goes up with increasing the number of lag distances and discretization of the area to variogram pairs for calculating the average fourth order covariance (see Figure 2.3). In general, less than 10 lag distances and 1000 discretization points are reasonable for variogram declustering.

Chapter 3

Estimation of Variogram Uncertainty and Transfer to Geostatistical Modeling

Variogram modeling fits an analytical model to the experimental variogram computed from the data to filter noise in the experimental points and provide a model for all possible distances and directions (Chiles & Delfiner, 1999; Cressie, 1985; Genton, 1998; Gringarten & Deutsch, 2001; Pardo-Igúzquiza & Dowd, 2001). This analytical model is used for kriging or simulation. Although variogram declustering can remove some noise and artifacts in the experimental variogram (Chapter 2), there is unavoidable uncertainty associated with the experimental variogram. This variogram uncertainty will be calculated and incorporated into geostatistical models.

Although there are many references on calculating and using the variogram, few authors have considered variogram uncertainty. Bogaert and Russo (1999); Webster and Oliver (1992) measure the variogram uncertainty by sampling schemes. Cressie (1985); Marchant and Lark (2004); Ortiz and Deutsch (2002); Pardo-Igúzquiza and Dowd (2001) have suggested calculating variogram uncertainty by the covariance matrix of the experimental variogram resolved by quadratic covariances (Matheron, 1965).

This Chapter presents a new approach of quantifying variogram uncertainty. The Chapter begins by discussing two variogram uncertainty approaches and their limitations. A new approach of quantifying variogram uncertainty is proposed, and a methodology to transfer this uncertainty through geostatistical modeling is developed.

Synthetic and realistic examples are presented.

3.1 Variogram Uncertainty Approaches and Limitations

3.1.1 Fourth Order Moments (FOM)

Marchant and Lark (2004); Ortiz and Deutsch (2002) have derived the variance of each lag distance for a multivariate Gaussian field in terms of fourth order moments. Figure 3.1 shows the fourth order covariance between two variogram pairs. Fourth order covariance between two variogram pairs "Pair p" and "Pair q" with lag distances $\mathbf{h}_p$ and $\mathbf{h}_q$ in Figure 3.1 is written as:

$$F(p-q) = Cov\left\{(y_1 - y_2)^2, (y_3 - y_4)^2\right\} \tag{3.1}$$

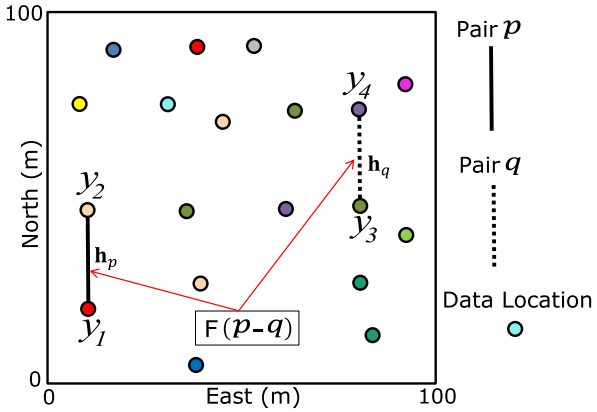

Figure 3.1: Fourth order covariance F(p-q) between two variogram pairs "Pair p" and "Pair q" with lag distances $\mathbf{h}_p$ and $\mathbf{h}_q$.

under the multivariate Gaussian distribution, any fourth order moment can be computed by the pairwise covariances (Matheron, 1965). The fourth order covariance was discussed in Chapter 2, Section 2.5. The average fourth order covariance between variogram pairs of each lag distance is proportional to variogram uncertainty (variance). This is the diagonal elements of the average fourth order covariance matrix between lag distances (see Equation 3.3). A formula for the average fourth order covariance

between the pairs of the experimental variogram is written as:

$$\overline{F} = \frac{1}{n(\mathbf{h}_p).n(\mathbf{h}_q)} \sum_{i=1}^{n(\mathbf{h}_p)} \sum_{j=1}^{n(\mathbf{h}_q)} F(i-j) \quad , i = 1, ..., n(\mathbf{h}_p) \quad , \quad j = 1, ..., n(\mathbf{h}_q) \tag{3.2}$$

where $n(\mathbf{h}_p)$ and $n(\mathbf{h}_q)$ are number of variogram pairs for lag distances $\mathbf{h}_p$ and $\mathbf{h}_q$. $F(i-j)$, $i = 1, ..., n(\mathbf{h}_p)$, $j = 1, ..., n(\mathbf{h}_q)$ are the fourth order covariances. The average fourth order covariance matrix between lag distances is written as:

$$\left[\overline{F}\right] = \begin{bmatrix} \overline{F_{1,1}} & .. & \overline{F_{1,n}} \\ : & \ddots & : \\ \overline{F_{n,1}} & .. & \overline{F_{n,n}} \end{bmatrix} \tag{3.3}$$

where n is number of lag distances. $\overline{F_{i,j}}$, $i, j = 1, ..., n$ are the average fourth order covariance. For example, $\overline{F_{1,2}}$ is the average fourth order covariance between variogram pairs of lag distance 1 and lag distance 2. The diagonal elements of this covariance matrix $\overline{F_{1,1}}$, ... , $\overline{F_{n,n}}$ are the average fourth order covariances between variogram pairs to themselves. Since the experimental variogram (semivariogram) is calculated as half the average squared difference between points, the variance of each lag distance is calculated by:

$$Var\left\{2\widehat{\gamma}(\mathbf{h})\right\} = \overline{F_{i,i}} \quad \rightarrow \quad Var\left\{\widehat{\gamma}(\mathbf{h})\right\} = \frac{\overline{F_{i,i}}}{4} \quad , \quad i = 1, ..., n \tag{3.4}$$

where $Var\left\{\widehat{\gamma}(\mathbf{h})\right\}$ is the variance of each lag distance. The mean of each lag distance is the fitted variogram model $\gamma(\mathbf{h})$. The calculated variogram uncertainty by this approach is very high (for more details and examples see Ortiz and Deutsch (2002)). The high variogram uncertainty is because of no conditioning and not considering the degrees of freedom of variogram pairs for calculating variogram uncertainty. The proposed methodology developed below calculates the degrees of freedom of variogram pairs for quantifying variogram uncertainty.

3.1.2 Global Kriging of Variogram Pairs

The ordinary global kriging (OGK) was discussed in Chapter 2 to decluster the experimental variogram and calculate variogram uncertainty (variance). Variogram uncertainty by simple global kriging (SGK) can also be calculated (Rezvandehy & Deutsch, 2016). Compared to OGK (see Equation 2.3), SGK requires a stationary mean and it does not enforce sum of the weights to 1. The simple global kriging (SGK) estimator of z for the entire domain A is written:

$$Z^*_{SGK} - m = \sum_{\alpha=1}^{n} \lambda_\alpha^{SGK} [Z(\mathbf{u}_\alpha) - m] \tag{3.5}$$

m is a stationary mean in the kriging system and n is number of data. λ_α^{SGK} is the weight of global mean and variance assigned to each datum $z(\mathbf{u}_\alpha)$. The weight is obtained by solving the following system of linear equations:

$$\sum_{\beta=1}^{n} \lambda_\beta^{SGK} C(\mathbf{u}_\beta - \mathbf{u}_\alpha) = \overline{C(\mathbf{u}_\alpha - A)} \quad , \quad \alpha = 1....n \tag{3.6}$$

where $C(\mathbf{u}_\beta - \mathbf{u}_\alpha)$ is the covariance between locations $\mathbf{u}_\beta$ and $\mathbf{u}_\alpha$, and $\overline{C(\mathbf{u}_\alpha - A)}$ is the average covariance between each data location $\mathbf{u}_\alpha$ and the whole area A. The estimation variance of SGK is written as:

$$Var\{Z^*_{SGK}\} = \overline{C(A-A)} - \sum_{\alpha=1}^{n} \lambda_\alpha^{SGK}(A) \overline{C(\mathbf{u}_\alpha - A)} \tag{3.7}$$

where $\overline{C(A-A)}$ is the average covariance of the entire area to itself (Goovaerts, 1997). SGK of variogram pairs for lag distance $\mathbf{h}$ is computed by:

$$X^*_{\mathbf{h}} - \gamma(\mathbf{h}) = \sum_{\alpha=1}^{n(\mathbf{h})} \lambda_{\alpha,\mathbf{h}}^{SGK} [X_{\alpha,\mathbf{h}}(\mathbf{u}) - \gamma(\mathbf{h})] \tag{3.8}$$

where $n(\mathbf{h})$ is number of variogram pairs for lag distance $\mathbf{h}$, and $\gamma(\mathbf{h})$ is the fitted variogram model assumed as the stationary mean (see Equation 3.5) of lag distance $\mathbf{h}$. $\lambda_{\alpha,\mathbf{h}}^{SGK}$ is the weight for pair α and lag distance $\mathbf{h}$, and $X_{\alpha,\mathbf{h}}(\mathbf{u})$ is half of the quadratic

differences of the pair values (see Equation 2.6). The system of linear equations of SGK for obtaining $\lambda_{\alpha,\mathbf{h}}^{SGK}$ can be written as:

$$\sum_{\beta=1}^{n(\mathbf{h})} \lambda_{\beta,\mathbf{h}}^{SGK} F(X_{\beta,\mathbf{h}}(\mathbf{u}) - X_{\alpha,\mathbf{h}}(\mathbf{u})) = \overline{F(X_{\alpha,\mathbf{h}}(\mathbf{u}) - A)} \ , \ \alpha = 1....n(\mathbf{h}) \tag{3.9}$$

where $F(X_{\beta,h}(\mathbf{u}) - X_{\alpha,h}(\mathbf{u}))$ signifies fourth order covariance between pairs $X_{\alpha,\mathbf{h}}(\mathbf{u})$ and $X_{\beta,\mathbf{h}}(\mathbf{u})$ (left hand side), and $\overline{F(X_{\alpha,\mathbf{h}}(\mathbf{u}) - A)}$ implies the average fourth order covariance between each pair $X_{\alpha,h}(\mathbf{u})$ and the entire domain A (right hand side) (see Figure 2.3). Variogram uncertainty of each lag distance $\mathbf{h}$ is written as:

$$\begin{aligned} Var\left\{2X_{\mathbf{h}}^{*}\right\} = \overline{F(A-A)} - \sum_{\alpha=1}^{n(\mathbf{h})} \lambda_{\alpha,\mathbf{h}}^{SGK} \overline{F(X_{\alpha,h}(\mathbf{u}) - A)} = Var\left\{2\widehat{\gamma}(\mathbf{h})\right\} \\ \rightarrow \quad Var\left\{\widehat{\gamma}(\mathbf{h})\right\} = \frac{1}{4} \times \left(Var\left\{2X_{\mathbf{h}}^{*}\right\}\right) \end{aligned} \tag{3.10}$$

where $\overline{F(A-A)}$ is the average fourth order covariance of the entire domain A to itself. The variogram distribution of each lag distance has a mean of $X_{\mathbf{h}}^{*}$ and variance of $Var\left\{\widehat{\gamma}(\mathbf{h})\right\}$ (see Equation 3.10). The simple kriging estimator is unbiased (Goovaerts, 1997):

$$E\left\{X_{\mathbf{h}}^{*}(\mathbf{u}) - X_{\mathbf{h}}(\mathbf{u})\right\} = X_{\mathbf{h}}^{*} - \gamma(\mathbf{h}) = 0 \quad \rightarrow \quad X_{\mathbf{h}}^{*} = \gamma(\mathbf{h}) \tag{3.11}$$

where $X_{\mathbf{h}}^{*}(\mathbf{u})$ is the estimated variogram pair for each location $\mathbf{u}$ and $X_{\mathbf{h}}(\mathbf{u})$ is the unknown value for lag distance $\mathbf{h}$. The expected value of the estimated variogram pair $X_{\mathbf{h}}^{*}$ is calculated by global kriging (see Equation 3.8). The expected value of the unknown value is the fitted variogram model $\gamma(\mathbf{h})$ for each lag distance $\mathbf{h}$. The fitted variogram model $\gamma(\mathbf{h})$ is assumed as the mean of the variogram distribution for each lag distance instead of $X_{\mathbf{h}}^{*}$.

3.1.2.1 Small Example of Variogram Uncertainty by Global Kriging

A small 2D data set is considered to calculate variogram uncertainty by global kriging approach (OGK and SGK). Figure 3.2 shows a synthetic data set with twelve data locations which leads to seven variogram pairs for lag distance approximately $10m$ and

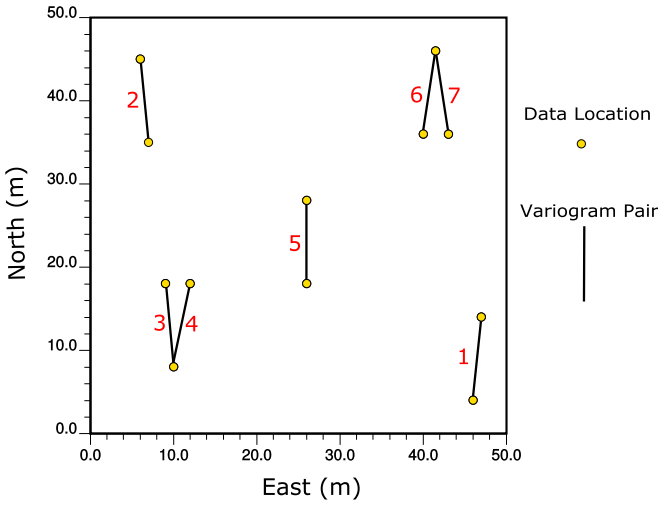

Figure 3.2: Synthetic data set with twelve well locations and seven variogram pairs for lag distance approximately $10m$ and azimuth $0^\circ \pm 10^\circ$.

azimuth $0^\circ \pm 10^\circ$. The fourth order covariance matrix of these pairs for calculating SGK weights can be written as:

$$\begin{bmatrix} F(1-1) & F(1-2) & F(1-3) & F(1-4) & F(1-5) & F(1-6) & F(1-7) \\ F(2-1) & F(2-2) & F(2-3) & F(2-4) & F(2-5) & F(2-6) & F(2-7) \\ F(3-1) & F(3-2) & F(3-3) & F(3-4) & F(3-5) & F(3-6) & F(3-7) \\ F(4-1) & F(4-2) & F(4-3) & F(4-4) & F(4-5) & F(4-6) & F(4-7) \\ F(5-1) & F(5-2) & F(5-3) & F(5-4) & F(5-5) & F(5-6) & F(5-7) \\ F(6-1) & F(6-2) & F(6-3) & F(6-4) & F(6-5) & F(6-6) & F(6-7) \\ F(7-1) & F(7-2) & F(7-3) & F(7-4) & F(7-5) & F(7-6) & F(7-7) \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ \lambda_7 \end{bmatrix} = \begin{bmatrix} \overline{F(1-A)} \\ \overline{F(2-A)} \\ \overline{F(3-A)} \\ \overline{F(4-A)} \\ \overline{F(5-A)} \\ \overline{F(6-A)} \\ \overline{F(7-A)} \end{bmatrix} \tag{3.12}$$

where the left hand side covariance implies the fourth order covariance matrix between seven pairs written in simple kriging format. The right hand side covariance is the average fourth order covariance of each pair to the entire domain. This covariance could be calculated by discretizing the entire study area for the desired lag distance and calculating the fourth order covariances by a discretized approximation (see Figure 2.3). λ_1 to λ_7 are the weights by SGK approach. These are calculated by solving Equation 3.12. The fitted variogram model is assumed an isotropic exponential model

with range of 25m and sill 1, the left hand side covariance is calculated as:

$$\begin{bmatrix} 3.93 & 0.137\times10^{-5} & 0.224\times10^{-4} & 0.367\times10^{-4} & 0.383\times10^{-5} & 0.00163 & 0.0022 \\ 0.137\times10^{-5} & 3.93 & 0.00792 & 0.0631 & 0.000513 & 0.598\times10^{-4} & 0.329\times10^{-4} \\ 0.224\times10^{-4} & 0.00792 & 3.93 & 2.44 & 0.000917 & 0.198\times10^{-4} & 0.279\times10^{-5} \\ 0.367\times10^{-4} & 0.0631 & 2.44 & 3.99 & 0.687\times10^{-5} & 0.778\times10^{-4} & 0.189\times10^{-4} \\ 0.383\times10^{-5} & 0.000513 & 0.000917 & 0.687\times10^{-5} & 3.91 & 0.00375 & 0.000601 \\ 0.0163 & 0.598\times10^{-4} & 0.198\times10^{-4} & 0.778\times10^{-4} & 0.00375 & 3.95 & 2.43 \\ 0.022 & 0.329\times10^{-4} & 0.279\times10^{-5} & 0.189\times10^{-4} & 0.000601 & 2.43 & 3.95 \end{bmatrix} \tag{3.13}$$

and by calculating right hand side covariance, SGK weights for variogram pairs are achieved:

$$\begin{bmatrix} 0.00250 \\ 0.00285 \\ 0.00298 \\ 0.00300 \\ 0.00316 \\ 0.00288 \\ 0.00283 \end{bmatrix} \Rightarrow \lambda_1 = 0.006\,, \lambda_2 = 0.007\,, \lambda_3 = 0.005\,, \lambda_4 = 0.005\,, \lambda_5 = 0.008\,, \lambda_6 = 0.005\,, \lambda_7 = 0.004 \tag{3.14}$$

the difference between the weights is quite small and do not sum to 1. Therefore, SGK is not a viable technique for variogram declustering. Variogram uncertainty by SGK is calculated as:

$$Var\left\{\widehat{\gamma}(\mathbf{h})\right\} = \frac{1}{4}\times Var\left\{2X^{*}_{\mathbf{h}}\right\} = \frac{1}{4}\times(\overline{F(A-A)} - \sum_{\alpha=1}^{7}\lambda^{SGK}_{\alpha,\mathbf{h}}(A)\overline{F(X_{\alpha,h}(\mathbf{u})-A)}) = 2.6\times10^{-3} \tag{3.15}$$

where $\overline{F(A-A)} = 1.1568\times10^{-2}$, which is the average fourth order covariance of the entire domain A to itself. If OGK is applied on the data set of Figure 3.2, the weights and Lagrange parameter are achieved as:

$$\lambda_1 = 0.182\,, \lambda_2 = 0.183\,, \lambda_3 = 0.115\,, \lambda_4 = 0.111\,, \lambda_5 = 0.185\,, \lambda_6 = 0.113\,, \lambda_7 = 0.113\,, \mu = -0.69 \tag{3.16}$$

the calculated weights are reasonable in terms of declustering: pair 5 which is not

clustered with other pairs receives the highest weight ($\lambda_5 = 0.185$) followed by pairs 2 ($\lambda_2 = 0.183$) and 1 ($\lambda_1 = 0.182$). According to these weights, variogram uncertainty by OGK is calculated as:

$$Var\left\{\widehat{\gamma}(\mathbf{h})\right\} = 1.37 \times 10^{-2} \tag{3.17}$$

the variance of OGK is much higher than SGK (more than five times) due to not assuming a known mean value. The big difference between the calculated variogram uncertainty by OGK and SGK signifies that the global kriging approach is not robust. Global kriging is dependent to the size of domain. The higher domain, the lower variogram uncertainty. This arises a challenge for choosing the correct size of the domain. Another challenge faced by global kriging is the CPU time for calculating the average fourth order covariance for the entire domain (right hand side covariance matrix). A new robust approach of quantifying variogram uncertainty independent of the domain size is proposed in this Chapter.

3.2 Shape of the Variogram Distribution

The shape of the distribution of the variogram value for a particular lag vector is important for sampling and transferring uncertainty to downstream calculations (see Section 3.4, Figure 3.6). Ortiz and Deutsch (2002) assume the shape to be approximately Gaussian. Koushavand, Ortiz, and Deutsch (2008) assume a Gamma distribution. Khan and Deutsch (2016) and Marchant and Lark (2004) consider a Chi-square distribution. A numerical approach is implemented in this Section to define the shape of the variogram distribution. Three distributions are considered: Gaussian, Gamma and Chi-square distributions are fitted to the direct observation of variogram distribution (target distribution). All distributions have the same mean and variance. The Gaussian distribution is defined by the mean and variance.

The following notation for the Gamma distribution is from the NIST (National Institute of Standards and Technology) Engineering and Statistics Handbook available

on the web (see reference list (NIST, n.d.)):

$$f(x) = \frac{(\frac{x-\mu}{\beta})^{\gamma-1}.exp(-\frac{x-\mu}{\beta})}{\beta\Gamma(\gamma)} \qquad x \geq \mu; \quad \gamma, \beta > 0 \ \ where \ \ \Gamma(\gamma) = \int_0^\infty t^{\alpha-1}e^{-t} \tag{3.18}$$

where β is the scale parameter, μ is the location parameter, γ is the shape parameter and Γ is the Gamma function. The Gamma distribution is defined by these three parameters (β, γ, μ). The Gamma distribution is closely related to the Chi-square and Wishart distributions (Johnson, Kotz, & Balakrishnan, 1994; Johnson & Wichern, 2002).

According to Khan and Deutsch (2016), the Chi-square distribution of each lag distance ($f(\widehat{\gamma}(\mathbf{h}))$) is proportional to:

$$f(\widehat{\gamma}(\mathbf{h})) \sim \chi^2_{DoF} \times \frac{\gamma(\mathbf{h})}{DoF} \tag{3.19}$$

where χ^2_{DoF} is the Chi-square distribution with a degree of freedom (DoF) and $\gamma(\mathbf{h})$ represents the base case or reference variogram model fitted to the experimental variogram. The only unknown parameter is DoF.

A simulation approach is considered to directly observe the variogram distribution (target distribution). Twenty variogram pairs with lag distance 4m are taken from an area assumed to be 18m by 18m (see Figure 3.3). 10000 unconditional realizations are generated by sequential Gaussian simulation (SGS) (Deutsch & Journel, 1998) of a stationary random field having an isotropic exponential variogram model with range of 12m and sill 1. The experimental variogram of each realization is computed to build the direct observation of variogram distribution for lag distance 4m. The best Gaussian, Gamma and Chi-square distributions are fit. All distributions should have approximately the same mean and the same variance as experimentally observed. The Gamma distribution is fit iteratively (20000 iterations) with different β, γ, μ to find the closest fit to the target distribution. The Chi-square distribution is calculated by Equation 3.19 with $\gamma(\mathbf{h}) = 0.63$, which is the fitted variogram model for lag distance 4m, and for different degrees of freedom (from 1 to 100) to find the distribution that

has the lowest difference with the direct variogram distribution.

Figures 3.4-a, b and c show the best fit Gaussian, Gamma and Chi-square distributions. The Gamma distribution is defined by $\beta = 0.116$, $\gamma = 3.528$, $\mu = 0.217$. The Chi-square distribution is defined by a DoF=17. The red curve is the fitted distribution. The error (mismatch) between the direct variogram distribution and the fitted distribution is shown, which is calculated by the sum of squared differences between the fitted and direct variogram distributions. The error is high (1.285) for the Gaussian distribution. In general, the square of a Gaussian statistic is not Gaussian. The Gamma and Chi-square distribution are similar. However, the error of the Chi-square distribution (0.011) is lower than the Gamma distribution (0.263). Moreover, the Chi-square distribution is simpler to build because of only one parameter (see Equation 3.19) whereas the Gamma distribution requires three parameters β, γ, μ. Thus, the shape of the variogram distribution is assumed to be a Chi-square distribution achieved by only DoF of each lag. Variogram uncertainty (variance) of each lag distance only requires the DoF to construct the Chi-square distribution for calculating variogram realizations (see Figure 3.6).

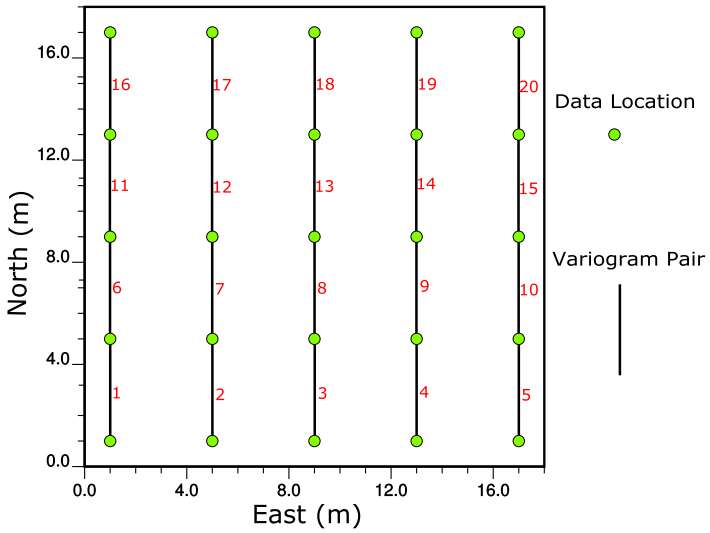

Figure 3.3: 2D synthetic data set with twenty variogram pairs for lag distance 4m.

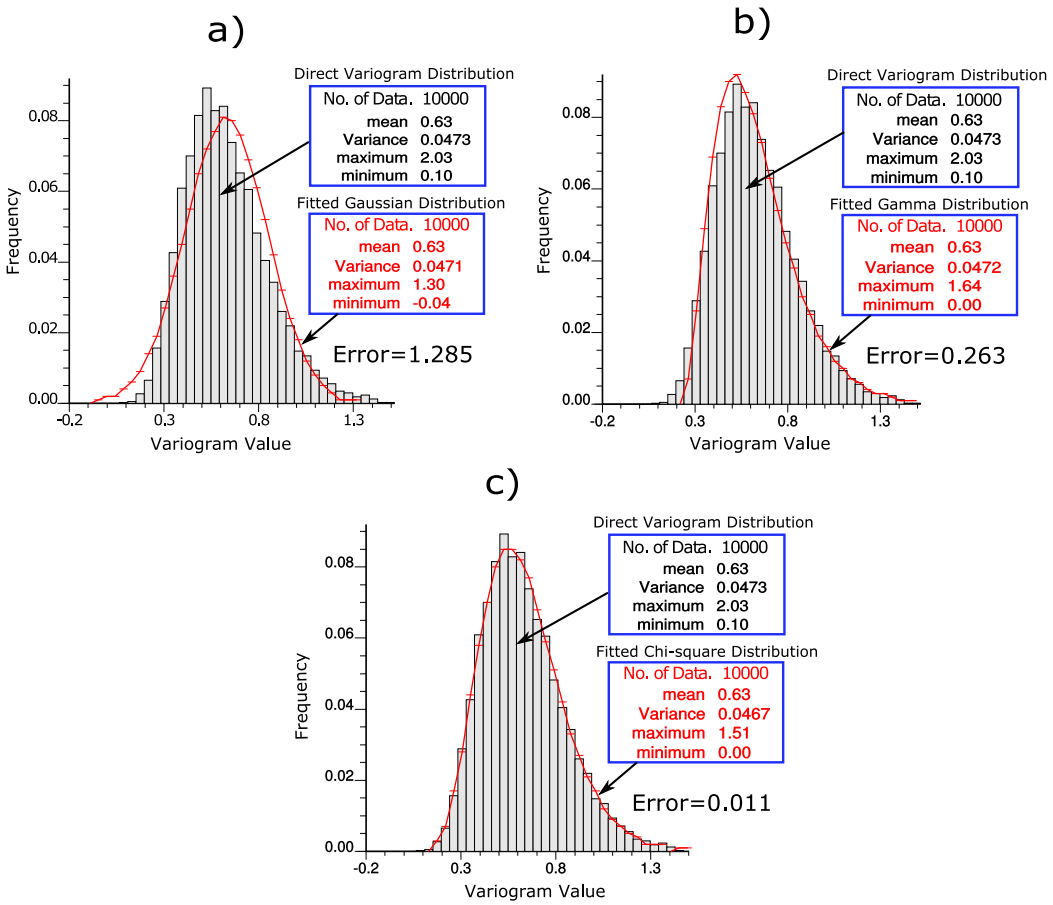

Figure 3.4: Gaussian (a), Gamma (b) and Chi-square (c) distributions fitted to the direct variogram distribution. The Chi-square distribution is calculated by DoF= 17, and the Gamma distribution is calculated by $\beta = 0.116$, $\gamma = 3.528$, $\mu = 0.217$.

3.3 Variogram Uncertainty by Direct Calculation of Degrees of Freedom (DoF)

Since the shape of the variogram distribution for each lag distance is reasonably a Chi-square distribution, the new approach of quantifying variogram uncertainty is proposed by the theoretical calculation of degree of freedom (DoF) for variogram pairs. This is called the DoF approach for quantifying variogram uncertainty. The mean of a Chi-square distribution for variogram distribution of each lag distance is $\gamma(\mathbf{h})$:

$$E\left\{f(\widehat{\gamma}(\mathrm{h}))\right\} \sim E\left\{\chi^2_{DoF} \times \frac{\gamma(\mathbf{h})}{DoF}\right\} = \gamma(\mathbf{h}) \tag{3.20}$$

because the mean of χ^2_{DoF} is DoF. $\gamma(\mathbf{h})$ is second order moments (variogram) from fitting a variogram model to a sample variogram computed from the data (Chiles & Delfiner, 1999; Cressie, 1985; Pardo-Igúzquiza & Dowd, 2001). The variance of the Chi-square distribution of each lag distance ($Var\{f(\widehat{\gamma}(\mathrm{h}))\}$) is written as:

$$Var\{f(\widehat{\gamma}(\mathrm{h}))\} \sim Var\left\{\chi^2_{DoF} \times \frac{\gamma(\mathbf{h})}{DoF}\right\} = (\frac{\gamma(\mathbf{h})}{DoF})^2 \times Var\left\{\chi^2_{DoF}\right\} = \frac{2 \times \gamma(\mathbf{h})^2}{DoF} \quad (3.21)$$

because $Var\left\{\chi^2_{DoF}\right\}$ is $2 \times DoF$. The only unknown parameter for calculating variogram uncertainty (variance) is DoF, which is the effective or independent number of variogram pairs. A theoretical derivation for calculating DoF is given in Bretherton, Widmann, Dymnikov, Wallace, and Bladé (1999) achieved by covariance matrix of data locations:

$$DoF = \frac{(\sum_{i=1}^{n} C_{ii})^2}{\sum_{i=1}^{n}\sum_{j=1}^{n} C_{ij}^2} \quad (3.22)$$

where C_{ii} and C_{ij} are diagonal elements and off-diagonal elements of a covariance matrix between data locations (n), respectively. The diagonal elements of a covariance matrix are the covariance of each data location to itself that are the variance of data. Equation 3.22 gives theoretical DoF for data locations. Since there are pairs instead of locations for the experimental variogram, fourth order covariances are used in Equation 3.22 to calculate DoF of each lag distance:

$$DoF = \frac{(\sum_{i=1}^{n(\mathbf{h})} F(i-i))^2}{\sum_{i=1}^{n(\mathbf{h})}\sum_{j=1}^{n(\mathbf{h})} F(i-j)^2} \quad (3.23)$$

where $n(\mathbf{h})$ is number of variogram pairs for lag distance $\mathbf{h}$ and $F(i-i)$ is the fourth order covariance of pair i to itself, and $F(i-j)$ is the fourth order covariance between pairs i and j. By calculating DoF of each lag distance, the variogram distribution and variogram uncertainty is achieved.

A synthetic 2D data set is considered to calculate variogram uncertainty for a lag distance by DoF approach. Figure 3.5 shows a 2D data set with twenty eight data locations which leads to twenty one variogram pairs for a lag distance approximately $10m$ and azimuth $0° \pm 15°$. DoF (Equation 3.23) and variogram uncertainty (variance)

(Equation 3.21) of each lag distance are calculated for four variogram models with

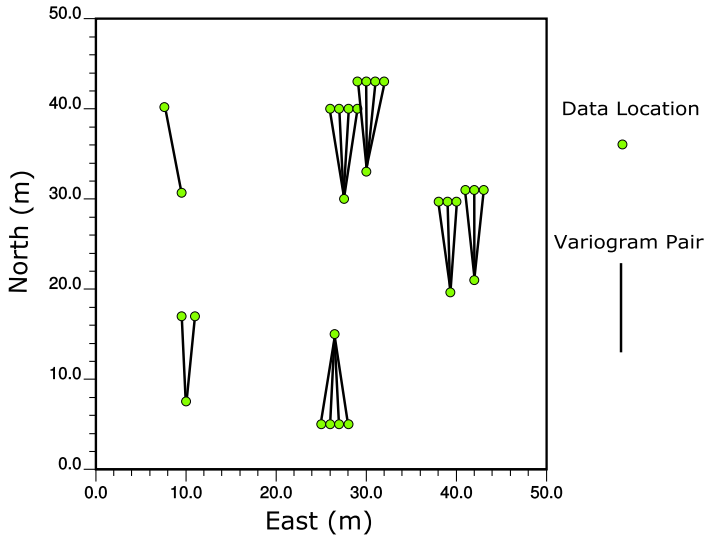

Figure 3.5: 2D synthetic data set with forty two data locations and twenty one variogram pairs for lag distance approximately $10m$ and azimuth $0° \pm 15°$.

ranges of 20m to 50m to analysis the effect of variogram range on variogram uncertainty. The variogram model is an isotropic spherical model with sill 1 (no nugget effect). Table 3.1 shows the results. As the range of the variogram model increases, the effective number of pairs (DoF) decreases. This leads to a higher variogram uncertainty for a high range variogram, and a lower variogram uncertainty for a low range variogram (see Equation 3.21). This is similar to the spatial bootstrap (SB) technique for quantifying histogram uncertainty. The spatial bootstrap gives high and low histogram uncertainties (uncertainty in the mean and variance) for high and low variogram ranges, respectively; however, Equation 3.21 shows that small number of DoF for high variogram range and large number of DoF for low variogram range are relatively compensated by lower and higher $\gamma(\mathbf{h})$. $\gamma(\mathbf{h})$ is the mean of the Chi-square distribution for each lag distance: higher variogram range leads to lower mean of the Chi-square distribution, the lower the mean of a distribution, the lower the variance and vice versa. This is the reason the calculated DoFs change from 22 (range 20m) to 3 (range 50m) which is a noticeable difference while the calculated variogram uncertainties change from 0.04297 (range 20m) to 0.05841 (range 50m).

DoF approach of quantifying variogram uncertainty is more efficient than FOM

Table 3.1: Variogram uncertainties (variances) for variogram pairs of Figure 3.5 achieved by variogram models with ranges of 20m to 50m.

Variogram Ranges (m)	20	30	40	50
$\gamma(\mathbf{h})$	0.68750	0.48148	0.36719	0.29600
DoF	22	10	5	3
$Var\{f(\widehat{\gamma}(\mathrm{h}))\} \sim \frac{2\times\gamma(\mathbf{h})^2}{DoF}$	0.04297	0.04636	0.05393	0.05841

approach (Section 3.1.1) (Marchant & Lark, 2004; Ortiz & Deutsch, 2002) and it accounts for the degree of freedom of variogram pairs to directly build the Chi-square distribution. Compared with global kriging technique (Section 3.1.2), DoF approach is independent of the domain size and has less CPU time. Although the calculated variogram uncertainty by DoF approach is reasonable, it is not conditioned to well data. Thus, it leads to high variogram uncertainty and can be assumed as the prior uncertainty in the variogram (similar to spatial bootstrap as the prior uncertainty in the histogram (Chapter 6)). The high variogram uncertainty is decreased in geostatistical modeling because of conditioning (see next Section).

3.4 Variogram Uncertainty in Geostatistcial Modeling

The DoF approach of quantifying variogram uncertainty calculates the degrees of freedom of each lag distance to build the Chi-square distribution. After building the Chi-square distribution of each lag, variogram realizations should be drawn from the uncertainty interval of the variogram lags (Figure 3.6) and used in geostatistical simulation to incorporate variogram uncertainty in the final model. These realizations should preserve the correlation between lags. Otherwise, the variogram realizations will be noisy and unrealistic. LU simulation (Deutsch & Journel, 1998) can be applied with the correlation matrix of lag distances. The correlated Gaussian realizations are back-transformed to the marginal Chi-square distribution of each lag distance. This approach was partially discussed by Koushavand et al. (2008).

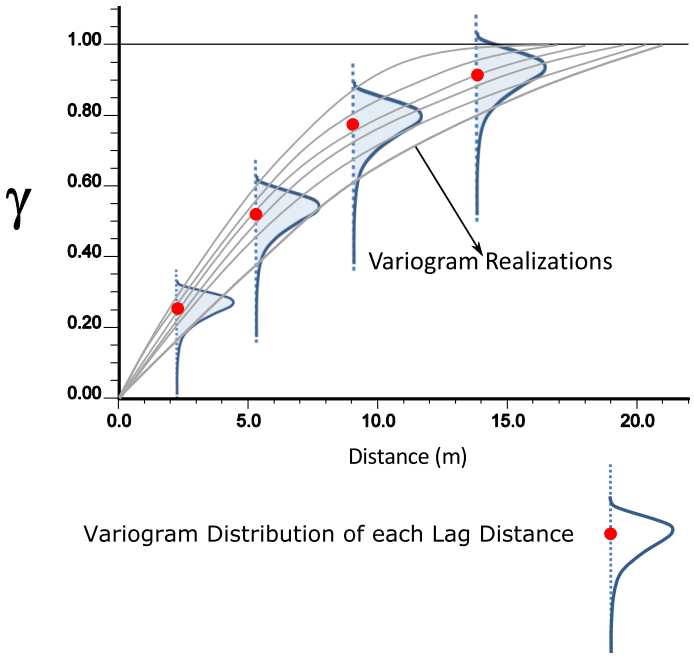

Figure 3.6: Schematic of six variogram realizations drawn from the variogram distribution of lag distances: the blue curve is the Chi-square distribution for each lag distance, the red dot is the mean of the each Chi-square distribution (the fitted variogram model) and the gray lines are the correlated variogram realizations.

3.4.1 LU Simulation and Back-transformation

Realizations of LU simulation (Deutsch & Journel, 1998) preserve the correlation between variables. In the present context, LU simulation can be applied with the correlation matrix of the variogram lag distances. LU realizations have a multivariate Gaussian distribution and can be back-transformed to the marginal Chi-square distribution for each lag distance approximately preserving the final correlation values. Figure 3.7-a shows the correlation matrix $\boldsymbol{\rho}$ between lags. The diagonal elements of this correlation matrix is 1, which is the correlation of each lag to itself. Figure 3.7-b shows LU simulation of the correlation matrix. Cholesky decomposition of the correlation matrix $\mathbf{L}$ is multiplied by an uncorrelated standard normal deviate $\mathbf{w} = [\mathbf{w}_1 ..... \mathbf{w}_n]$ (mean=0, standard deviation=1) to achieve the correlated Gaussian realization $\mathbf{y} = [\mathbf{y}_1 ..... \mathbf{y}_n]$. Where n is the number of lags. The Gaussian distribution of each lag distance is computed by generating many LU realizations (say 100 realizations). Then, the values from the Gaussian distribution are back-transformed to the Chi-square distribution of

each lag distance by quantile-quantile transformation (see Figure 3.8). The procedure is summarized by:

1. Define a stationary covariance function C($\mathbf{h}$) that comes from the fitted variogram model to the experimental variogram ($\gamma(\mathbf{h})$) (after variogram declustering in case of preferential sampling).
2. Define the Chi-square reference distribution of each lag distance $F_{l,DoF}(z),\ l = 1, ..., n$. Where n is the number of lag distance.
3. Construct the spatial lag-to-lag correlation matrix $\boldsymbol{\rho}$.
4. Compute the Cholesky decomposition of the correlation matrix as $\boldsymbol{\rho} = \mathbf{L}\mathbf{L}^T$.
5. Simulate a vector of uncorrelated standard normal deviate $\mathbf{w}$.
6. Generate a vector of correlated Gaussian realization $\mathbf{y} = \mathbf{L}\mathbf{w}$.
7. Lookup through the standard Gaussian CDF value of each lag distance $\mathbf{p} = G(\mathbf{y})$. Where G is the standard normal distribution.
8. Lookup the Chi-square distribution of each lag distance $\mathbf{z} = F^{-1}_{l,DoF}(\mathbf{p})$. Where $\mathbf{z}$ is the Chi-square correlated realization.

Steps 5 to 6 are repeated for multiple realizations (see Figure 3.7-b). Then, Steps 7 to 8 calculate the correlated variogram realizations in Chi-square distribution (see Figure 3.8). This approach ensures the variogram realizations respect the correlation between lag distances.

3.4.2 Correlation Matrix of Lag Distances

The correlation matrix between lag distances $\boldsymbol{\rho}$ is required to simulate the correlated variogram realizations by LU simulation. This correlation matrix could be achieved by calculating the average fourth order covariance matrix of variogram pairs (see Equation

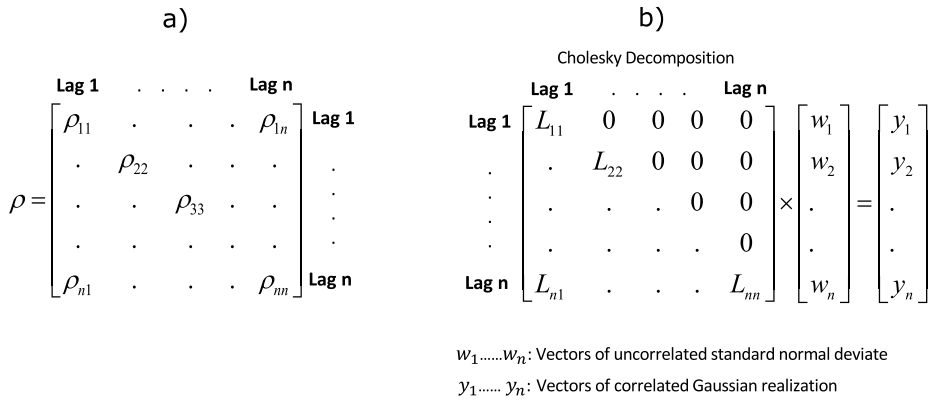

Figure 3.7: a) Correlation matrix $\boldsymbol{\rho}$ between lags. The diagonal elements of this covariance matrix is 1. b) LU simulation with the correlation matrix. Cholesky decomposition of the correlation matrix is multiplied by an uncorrelated standard normal deviate to achieve the correlated Gaussian realizations.

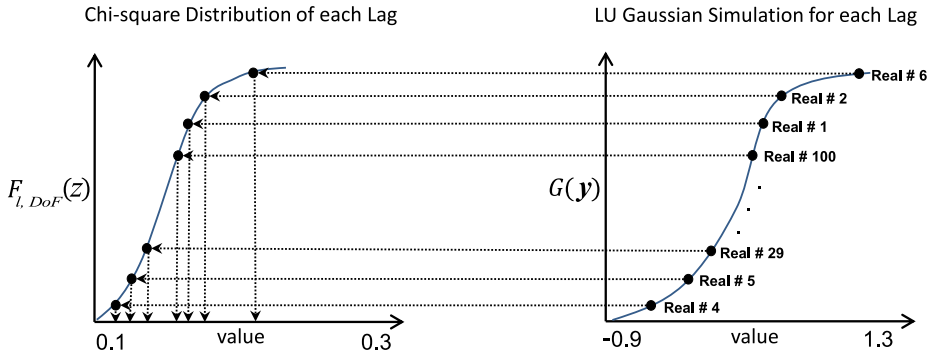

Figure 3.8: A schematic illustration of quantile-quantile transformation of each Gaussian realization (Real) to the Chi-square realization. The Gaussian distribution of each lag distance ($G(\mathbf{y})$) is back-transformed to the Chi-square distribution ($F_{l,DoF}(z)$) to calculate the correlated variogram realizations.

3.3) and converted to the Pearson correlation matrix:

$$\boldsymbol{\rho} = \begin{bmatrix} \frac{\overline{F_{1,1}}}{\sqrt{\overline{F_{1,1}}}\times\sqrt{\overline{F_{1,1}}}} & .. & \frac{\overline{F_{1,n}}}{\sqrt{\overline{F_{1,1}}}\times\sqrt{\overline{F_{n,n}}}} \\ : & \ddots & : \\ \frac{\overline{F_{n,1}}}{\sqrt{\overline{F_{n,n}}}\times\sqrt{\overline{F_{1,1}}}} & .. & \frac{\overline{F_{n,n}}}{\sqrt{\overline{F_{n,n}}}\times\sqrt{\overline{F_{n,n}}}} \end{bmatrix} \tag{3.24}$$

where $\overline{F_{i,j}}$, $i,j = 1,...,n$ are the average fourth order covariances between lag distances (n is number of lag distance). Equation 3.24 is the correlation matrix for isotropic or omnidirectional variogram where only one experimental variogram is cal-

culated. In case of an anisotropic variogram, multiple directions are considered: the major direction of continuity and the minor direction of continuity. In this case, variogram realizations should respect the cross correlation between lag distances in one direction and the other direction. A correlation matrix between lag distances of two directions should be calculated. This correlation matrix can be constructed by the average fourth order covariance between variogram pairs of two different directions and converted to a correlation matrix:

$$\boldsymbol{\rho} = \begin{bmatrix} \frac{\overline{F_{1,1}}}{\sqrt{\overline{F_{1,1}}\times\sqrt{F_{1,1}}}} & .. & \frac{\overline{F_{1,n}}}{\sqrt{\overline{F_{1,1}}\times\sqrt{F_{n,n}}}} & \frac{\overline{CF_{1,1}}}{\sqrt{\overline{CF_{1,1}}\times\sqrt{CF_{1,1}}}} & .. & \frac{\overline{CF_{1,m}}}{\sqrt{\overline{CF_{1,1}}\times\sqrt{CF_{m,m}}}} \\ : & \ddots & : & : & \ddots & \\ \frac{\overline{F_{n,1}}}{\sqrt{\overline{F_{n,n}}\times\sqrt{F_{1,1}}}} & .. & \frac{\overline{F_{n,n}}}{\sqrt{\overline{F_{n,n}}\times\sqrt{F_{n,n}}}} & \frac{\overline{CF_{n,1}}}{\sqrt{\overline{CF_{n,n}}\times\sqrt{CF_{1,1}}}} & .. & \frac{\overline{CF_{n,m}}}{\sqrt{\overline{CF_{n,n}}\times\sqrt{CF_{m,m}}}} \\ \frac{\overline{CF_{1,1}}}{\sqrt{\overline{CF_{1,1}}\times\sqrt{CF_{1,1}}}} & .. & \frac{\overline{CF_{1,m}}}{\sqrt{\overline{CF_{1,1}}\times\sqrt{CF_{m,m}}}} & \frac{\overline{F_{1,1}}}{\sqrt{\overline{F_{1,1}}\times\sqrt{F_{1,1}}}} & .. & \frac{\overline{F_{1,m}}}{\sqrt{\overline{F_{1,1}}\times\sqrt{F_{m,m}}}} \\ : & \ddots & : & : & \ddots & : \\ \frac{\overline{CF_{m,1}}}{\sqrt{\overline{CF_{m,m}}\times\sqrt{CF_{1,1}}}} & .. & \frac{\overline{CF_{m,n}}}{\sqrt{\overline{CF_{m,m}}\times\sqrt{CF_{n,n}}}} & \frac{\overline{F_{m,1}}}{\sqrt{\overline{F_{m,m}}\times\sqrt{F_{1,1}}}} & .. & \frac{\overline{F_{m,m}}}{\sqrt{\overline{F_{m,m}}\times\sqrt{F_{m,m}}}} \end{bmatrix} \tag{3.25}$$

where $\overline{F}$ is the average fourth order covariance for variogram pairs of the same direction, and $\overline{CF}$ is the average fourth order covariance for variogram pairs of cross directions. n is number of lags for direction 1 and m is number of lags for direction 2. For example, $\overline{F_{1,3}}$ signifies the average fourth order covariance between lag 1 and lag 3 of either direction 1 or direction 2. $\overline{CF_{1,3}}$ denotes the average fourth order covariance between lag 1 of direction 1 and lag 3 of direction 2. LU simulation should be applied with this correlation matrix to produce the correlated variogram realizations that not only preserve the correlation for the lags of the same direction but also the correlation for the lags of different directions.

3.4.3 Variogram Realizations for Omnidirectional Variogram

A synthetic example is considered to illustrate the calculation of variogram realizations for an ominidirectional (azimuth $0^\circ \pm 90^\circ$) variogram. Variogram realizations in case of an anisotropic variogram is discussed in Section 3.4.4. Figure 3.9-a shows a synthetic 2D example with the area of $50m \times 50m$ and eighty data locations. Figure 3.9-b shows omnidirectional (azimuth $0^\circ \pm 90^\circ$) experimental variogram and the fitted variogram

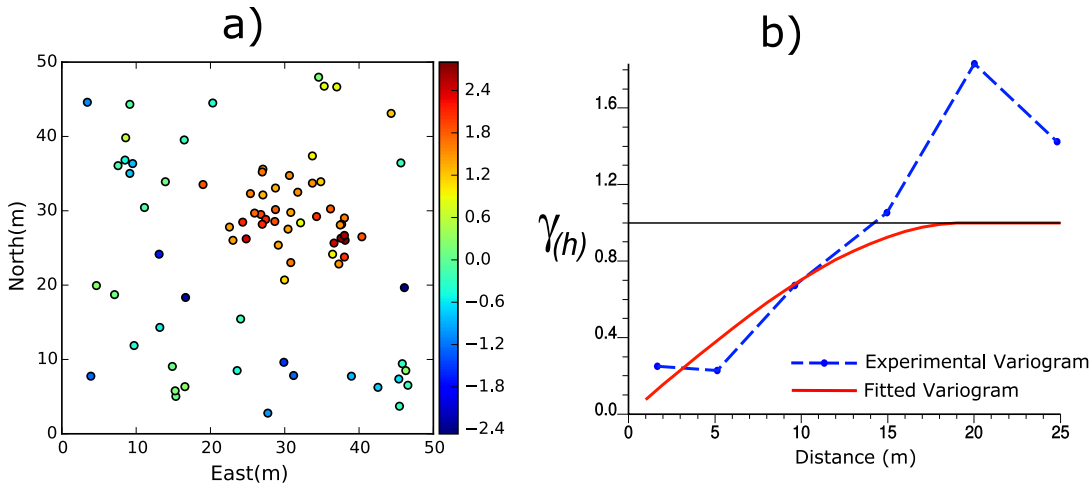

Figure 3.9: a) Synthetic example with 80 data locations. b) The experimental and fitted variogram model for azimuth $0° \pm 90°$.

model with a spherical model with range of 19.5m and sill 1 (no nugget effect). The DoF approach is applied to calculate variogram uncertainty. Figure 3.10 shows the correlation matrix of lags by Equation 3.24. The diagonal elements of this correlation matrix is 1 which is the correlation of each lag to itself. Figure 3.11 shows 100 correlated experimental variogram realizations after LU simulation and back transformation to the marginal Chi-square distribution of each lag. Figure 3.12 shows the correlation matrix between the sampled variogram realizations for six lag distances. The diagonal elements of this matrix shows the marginal Chi-square distribution of each lag. This matrix clearly demonstrates that the variogram realizations honor the correlation between lags (see Figures 3.10).

Variogram models should be fit to the experimental variogram realizations. These variogram models can be achieved by auto variogram modeling software (Larrondo et al., 2003). The fitted variogram models are constrained to have the same variogram structure number, type, and nugget effect as the base case variogram model (fitted variogram). Figure 3.13-a shows the fitted variogram realizations to the experimental realizations. Figure 3.13-b shows 100 standardized fitted variogram realizations. These realizations are used in geostatistical modeling to incorporate variogram uncertainty: one standardized variogram realization (Figure 3.13-b) is used to simulate one realization of SGS. This process is repeated to simulate 100 realizations.

	Lag...1	Lag...2	Lag...3	Lag...4	Lag...5	Lag...6
Lag...1	1.00	0.49	0.34	0.21	0.15	0.14
Lag...2	0.49	1.00	0.84	0.49	0.31	0.31
Lag...3	0.34	0.84	1.00	0.63	0.40	0.40
Lag...4	0.21	0.49	0.63	1.00	0.74	0.67
Lag...5	0.15	0.31	0.40	0.74	1.00	0.82
Lag...6	0.14	0.31	0.40	0.67	0.82	1.00

Figure 3.10: Correlation matrix between six lag distances of the experimental variogram of Figure 3.9 by Equation 3.24.

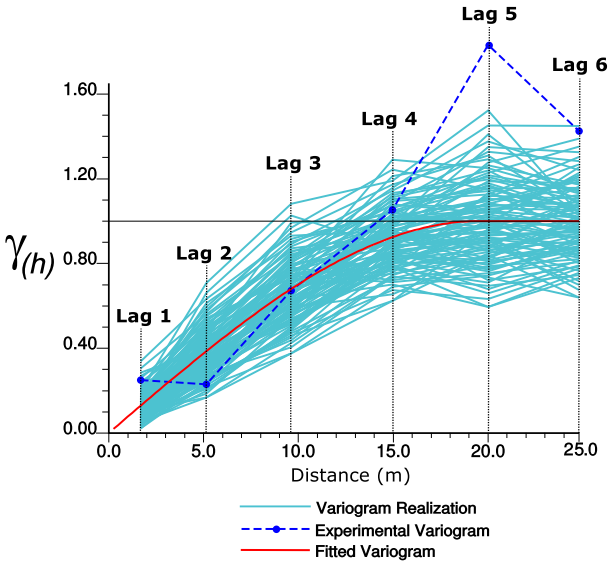

Figure 3.11: 100 correlated experimental variogram realizations.

The sill of the variogram realizations signifies uncertainty in the variance of the data. The uncertainty in the variance of data is accounted for in geostatistical modeling workflow by parameter uncertainty approaches such as spatial bootstrap (Khan & Deutsch, 2016). Figure 3.14 shows the cross plot between the ranked variance uncertainty by the spatial bootstrap and the ranked sills of the variogram realizations. There is a very high correlation which signifies that the sill of the variogram realizations are associated to the sampled variances. Incorporating variance uncertainty in geostatistical modeling by both spatial bootstrap and variogram uncertainty leads to

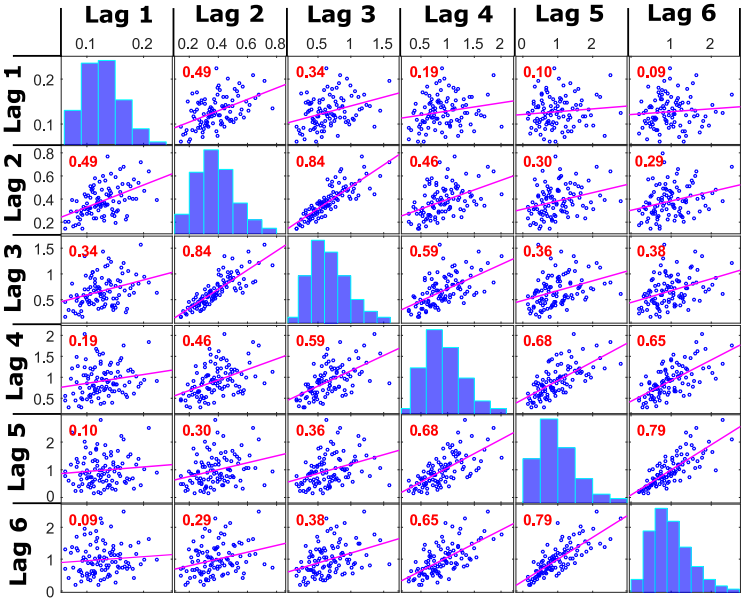

Figure 3.12: Correlation matrix between variogram realizations of six lag distances after LU simulation and back transformation to the marginal Chi-square distribution of each lag distance. The number on each cell signifies the correlation coefficient between lag distances. The diagonal elements are the marginal Chi-square distribution of each lag distance.

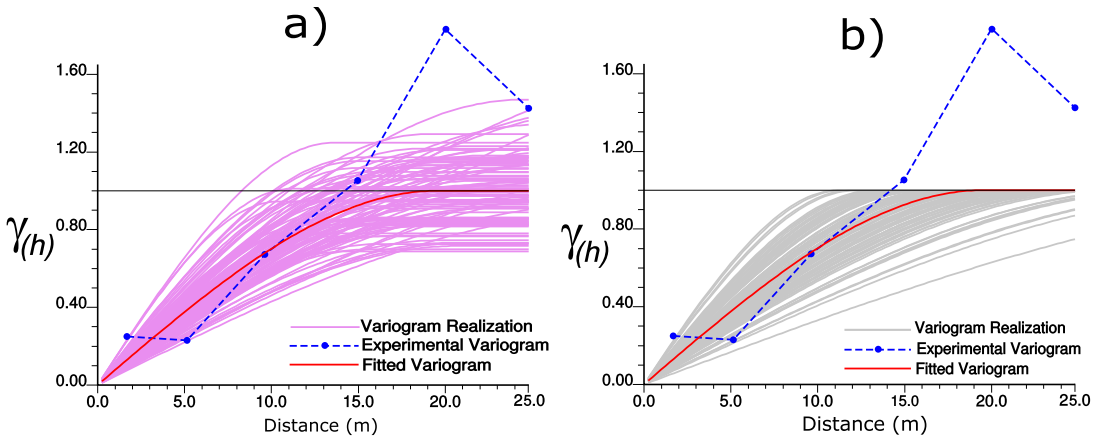

Figure 3.13: a) 100 fitted variogram realizations to the experimental realizations. b) Standardized to sill 1.

bias in the final model. Moreover, most geostatistical simulation techniques such as SGS needs standard normal data with variance 1, which needs the sill of the variogram model to be 1. Thus, all variogram realizations should be standardized to sill 1: all variance contributions of each variogram realization are divided by the total sill of that realization after variogram modeling.

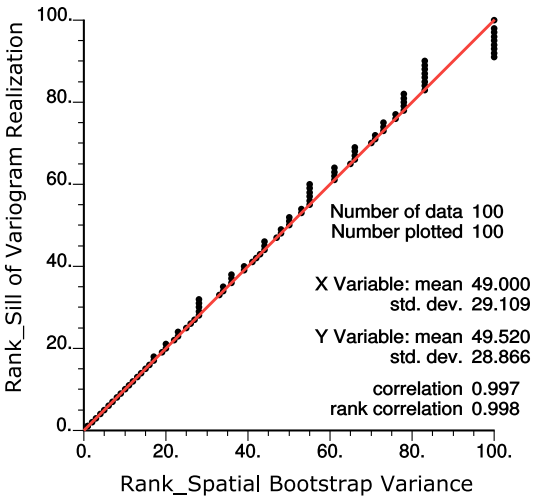

Figure 3.14: Cross plot between the ranked variance uncertainty by the spatial bootstrap and the ranked sills of the variogram realizations.

Figure 3.15 shows posterior variogram uncertainty: variogram of SGS realization for nx=50, ny=50 of data in Figure 3.9-a after incorporating variogram uncertainty. Posterior variogram uncertainty is lower than the prior variograms uncertainty calculated by the DoF approach (Figure 3.11 and 3.13). The conditioning of data in geostatistical modeling decreases high prior variogram uncertainty and makes the results conditioned and more realistic.

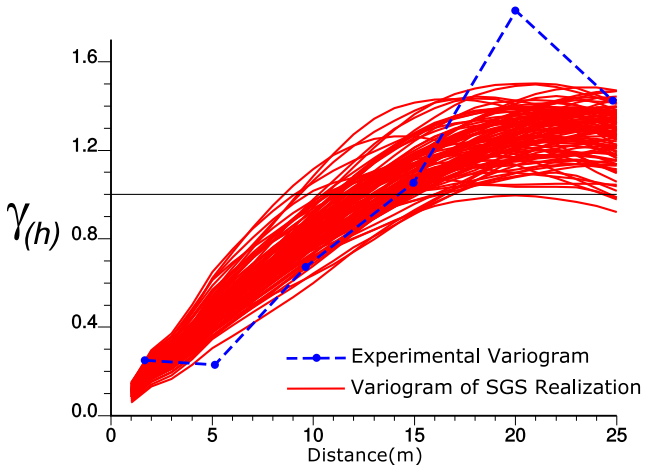

Figure 3.15: Posterior variogram uncertainty (variogram of SGS realization) after incorporating variogram uncertainty.

3.4.4 Variogram Realizations for Anisotropic Variogram

Another 2D synthetic example is considered to generate variogram realizations in presence of anisotropy. In case of an anisotropic variogram, the major direction of continuity and the minor direction of continuity are considered: variogram realizations should respect the cross correlation between lag distances in the minor and the minor directions. Figure 3.16 shows an anisotropic data set; there is more continuity for azimuth 0° than azimuth 90°. The experimental variograms are calculated for these directions.

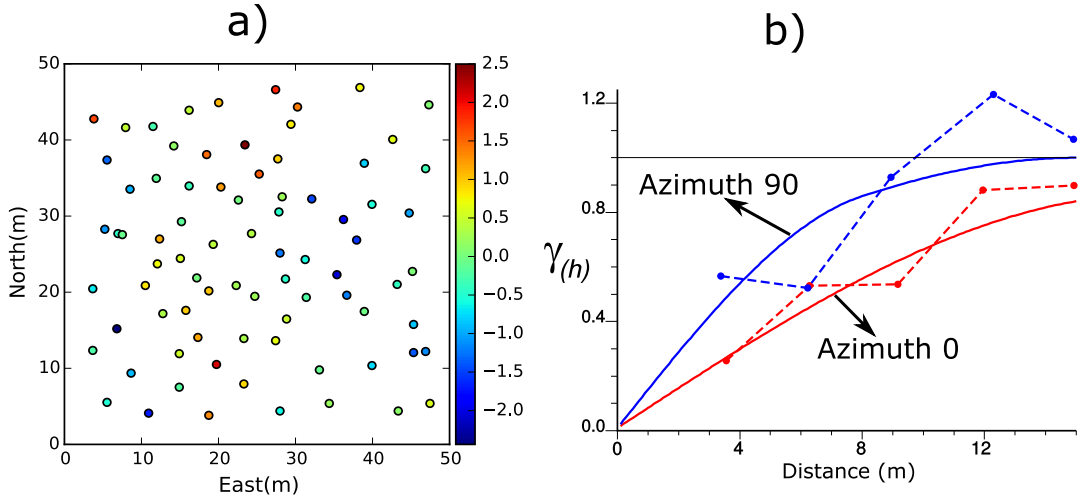

Figure 3.16: a) 2D synthetic example with 80 data locations. b) The experimental variogram and fitted variogram models (dashed lines) for azimuths $0° \pm 8°$ and $90° \pm 8°$.

The fitted variogram model for this example has two spherical structures with variance contributions of 0.48 and 0.52 (with no nugget effect). Ranges for structure 1 are 15m and 8m. Ranges for structure 2 are 30m and 15m. Figure 3.17 shows the correlation matrix between five lag distances of two directions by Equation 3.25. Similar to the previous example, the DoF approach is applied to calculate variogram uncertainty. Figure 3.18 shows 100 correlated experimental variogram realizations after LU simulation and back transformation to the marginal Chi-square distribution of each lag for azimuth 0° (Figure 3.18-a) and azimuth 90° (Figure 3.18-b). Figure 3.19 shows the correlation matrix between the sampled variogram realizations for five lag distances of azimuth 0° and azimuth 90°. The diagonal elements of this matrix shows the marginal Chi-square dis-

		Azimuth 0					Azimuth 90				
		Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5
Azimuth 0	Lag 1	1.00	0.13	0.22	0.20	0.15	0.03	0.10	0.07	0.08	0.08
	Lag 2	0.13	1.00	0.46	0.52	0.44	0.11	0.22	0.25	0.26	0.24
	Lag 3	0.22	0.46	1.00	0.62	0.68	0.17	0.28	0.34	0.37	0.36
	Lag 4	0.20	0.52	0.62	1.00	0.71	0.17	0.32	0.41	0.42	0.40
	Lag 5	0.15	0.44	0.68	0.71	1.00	0.16	0.33	0.40	0.45	0.44
Azimuth 90	Lag 1	0.03	0.11	0.17	0.17	0.16	1.00	0.10	0.25	0.21	0.15
	Lag 2	0.10	0.22	0.28	0.32	0.33	0.10	1.00	0.41	0.41	0.41
	Lag 3	0.07	0.25	0.34	0.41	0.40	0.25	0.41	1.00	0.67	0.59
	Lag 4	0.08	0.26	0.37	0.42	0.45	0.21	0.41	0.67	1.00	0.74
	Lag 5	0.08	0.24	0.36	0.40	0.44	0.15	0.41	0.59	0.74	1.00

Figure 3.17: Correlation matrix between lag distances of azimuth 0° and azimuth 90° in Figure 3.16.

tribution of each lag and each direction. This matrix demonstrates that the variogram

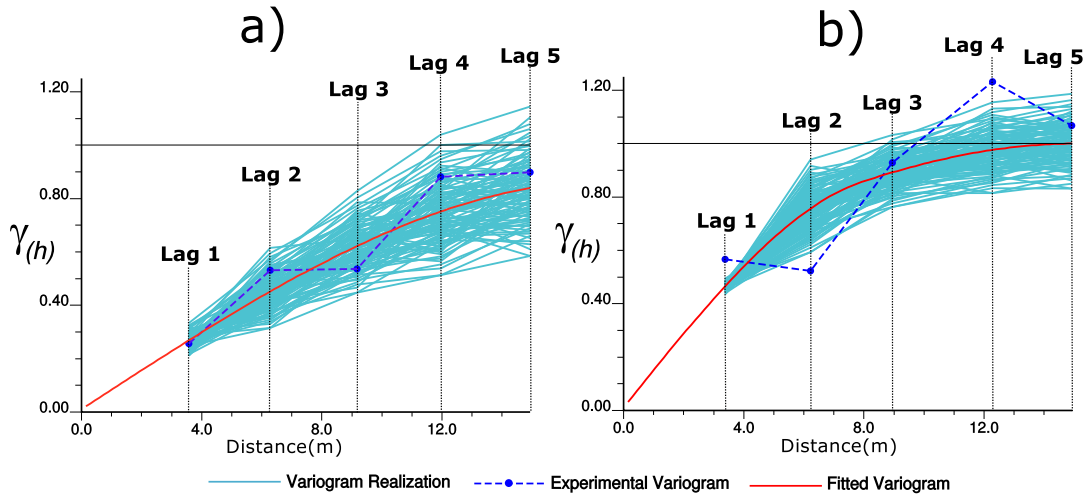

Figure 3.18: 100 correlated experimental variogram realizations for azimuth 0° (a) and azimuth 90° (b).

realizations respect the correlation between the lag distances of the same and different directions (see Figures 3.17). The final 2D variogram realizations are achieved by auto

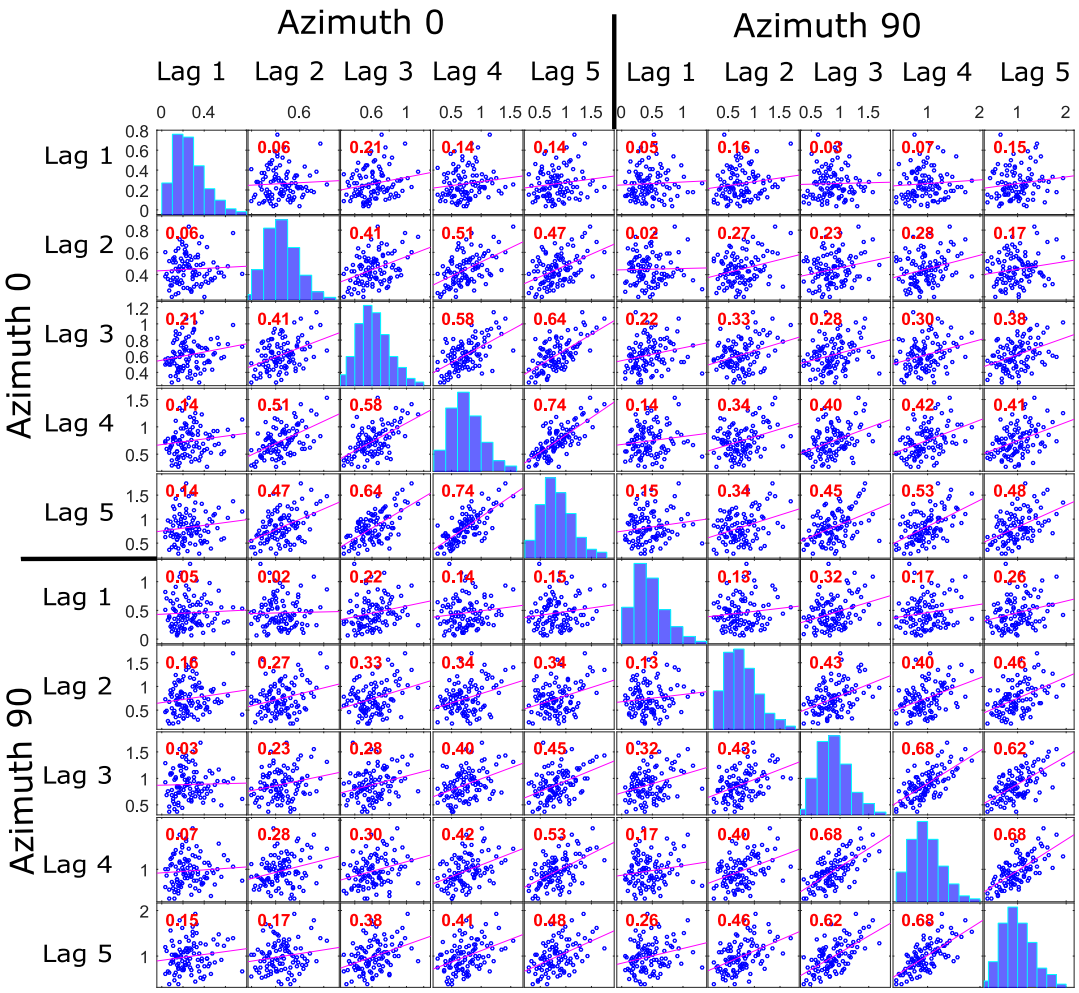

Figure 3.19: Correlation matrix between the sampled variogram realizations for five lag distances of azimuth 0° and azimuth 90°. The number on each cell signifies correlation coefficient between lag distances. The diagonal elements are the marginal Chi-square distribution of each lag distance.

variogram modeling of each realization of major and minor directions. Standardization could be applied on each variogram realization by dividing the variance contributions of each variogram structure by the total sill. Figure 3.20 shows the cross plot between the ranked variance uncertainty by the spatial bootstrap and the ranked total sills of the variogram realizations. There is a very high correlation which signifies that the sill of variogram realizations are associated to the sampled variances. Thus, all variogram realizations can be standardized to sill 1 and associated to a matching SB univariate distribution. Figure 3.21-a and b shows the standardized fitted variogram realizations for azimuths 0° and 90°, respectively. The final 2D variogram realizations are used

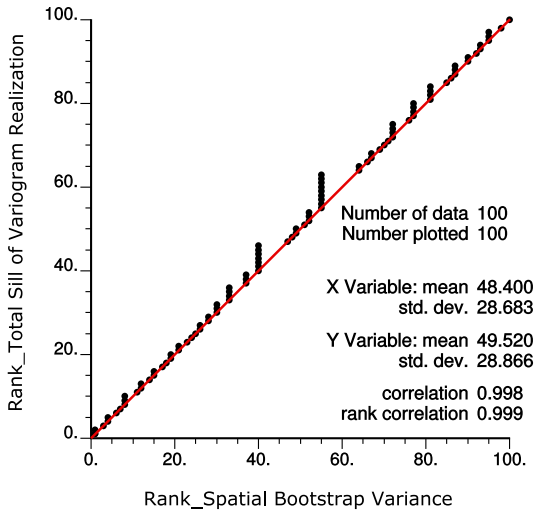

Figure 3.20: Cross plot between the ranked variance uncertainty by the spatial bootstrap and the ranked total sills of the variogram realizations.

in geostatitsical simulation to incorporate variogram uncertainty: one standardized variogram realization (Figure 3.21) is used to simulate one realization of SGS.

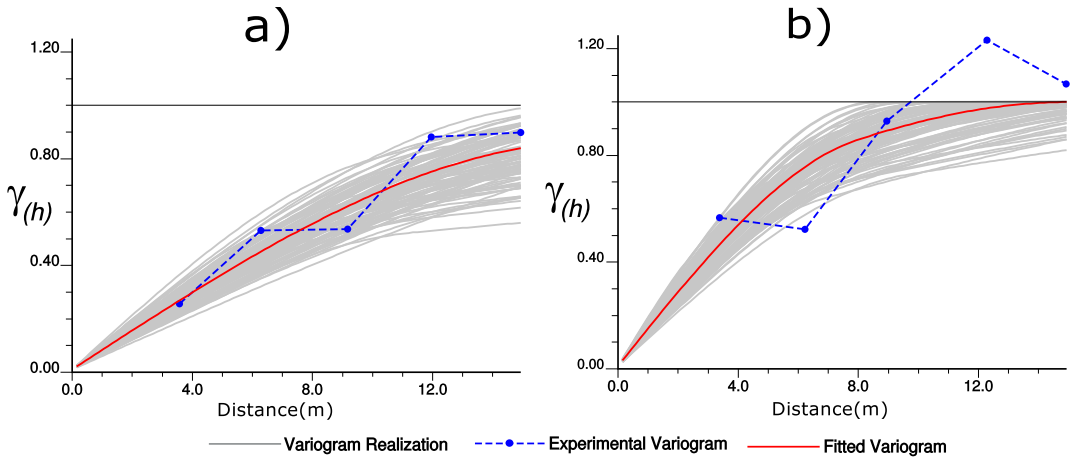

Figure 3.21: Standardized fitted variogram realizations for azimuths 0° (a) and 90° (b).

Figure 3.22 shows posterior variogram uncertainties for azimuths 0° (a) and 90° (b): variograms of SGS realization for nx=50 and ny=50 of data in Figure 3.16-a after incorporating variogram uncertainty. Posterior variogram uncertainties are lower than the prior variograms uncertainty calculated by the DoF approach (see Figure 3.18) because of conditioning data.

In case of a 3D data set, the final horizontal variogram realizations are calculated

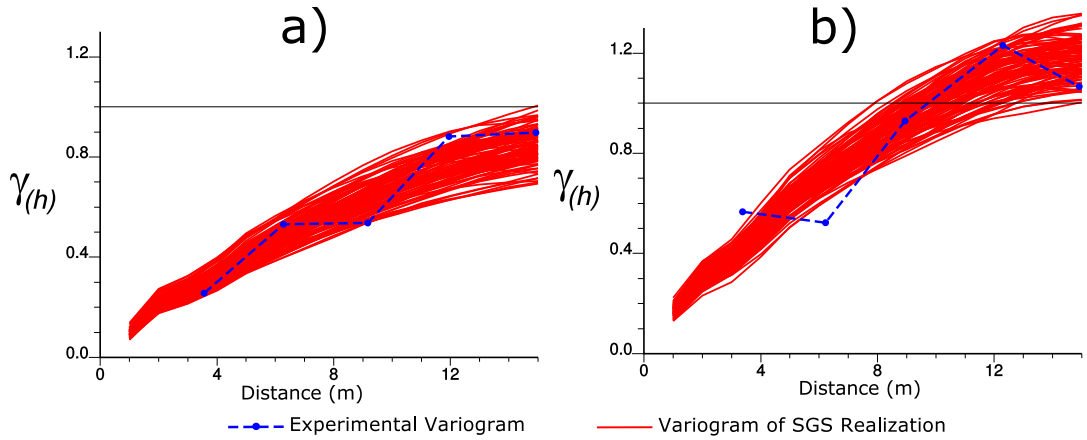

Figure 3.22: Posterior variogram uncertainty (variogram of SGS realization) for azimuths 0° (a) and 90° (b) after incorporating variogram uncertainty.

and then auto variogram modeling is applied with the vertical variogram to build the final 3D variogram realization. The vertical variogram is assumed to have no uncertainty because of many regularly spaced data in the direction of drilling.

3.5 Realistic Example: Amoco Case Study

A real case study of the Amoco 2D data (Chu, Xu, & Journel, 1994) is considered for quantifying variogram uncertainty and transfer through geostatistical modeling. The variable of interest is averaged permeability (in milliDarcies) over the main reservoir layer. Figure 3.23 shows a location map (a), histogram of the permeability (b) and the experimental variograms for azimuths 0° and 90° with the fitted models (c). Variograms are calculated with normal score data. The fitted variogram model has two structures with no nugget effect; the first structure is exponential with variance contribution 0.2 and variogram range of 1000ft for both 0° and 90° directions; the second structure is spherical with ranges of 15000ft and 4500ft for 0° and 90°, respectively. There is a zonal anisotropy for azimuth 0°. Figure 3.24 shows the correlation matrix between lag distances of the two directions by the average fourth order covariance.

Figure 3.25 shows 100 correlated variogram realizations generated by the DoF approach in azimuths 0° and 90°. For incorporating the variogram uncertainty in geostatistical modeling, one standardized variogram realization (Figure 3.25-b) is used to

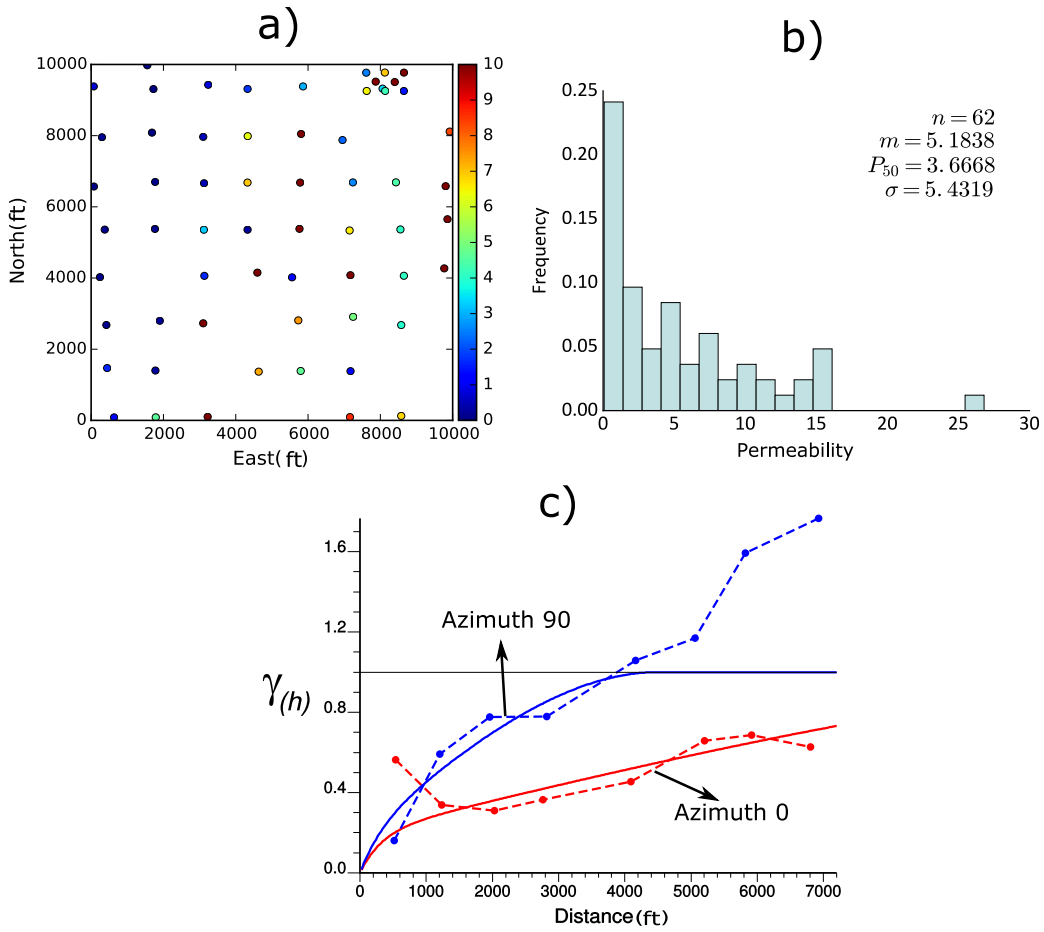

Figure 3.23: Permeability (milliDarcies) variable of Amoco data set: a) Location map. b) Histogram. c) Normal score experimental variograms for azimuths 0° and 90° with the fitted models.

simulate one realization of SGS for nx=100, ny=100 with normal score data per time. Gaussian realizations are transformed back to the original units (permeability). This process is repeated to simulate 100 SGS realizations. Figure 3.26 shows 6 SGS realizations of 100 realizations with variogram uncertainty, and Figure 3.27 shows the variogram realizations used for simulation. Figure 3.28 shows the same SGS realizations using only the reference variogram model of Figure 3.23-c. The difference between realizations in Figures 3.26 to Figure 3.28 is due to different variogram models. The variogram model for realization 1 has the highest variogram range of the second variogram structure for azimuth 0° among all variogram realizations; however, there is not a high continuity for realization 1, azimuth 0° (Figure 3.26) due to the fact that the first variogram structure has a short range and high variance contribution. There are

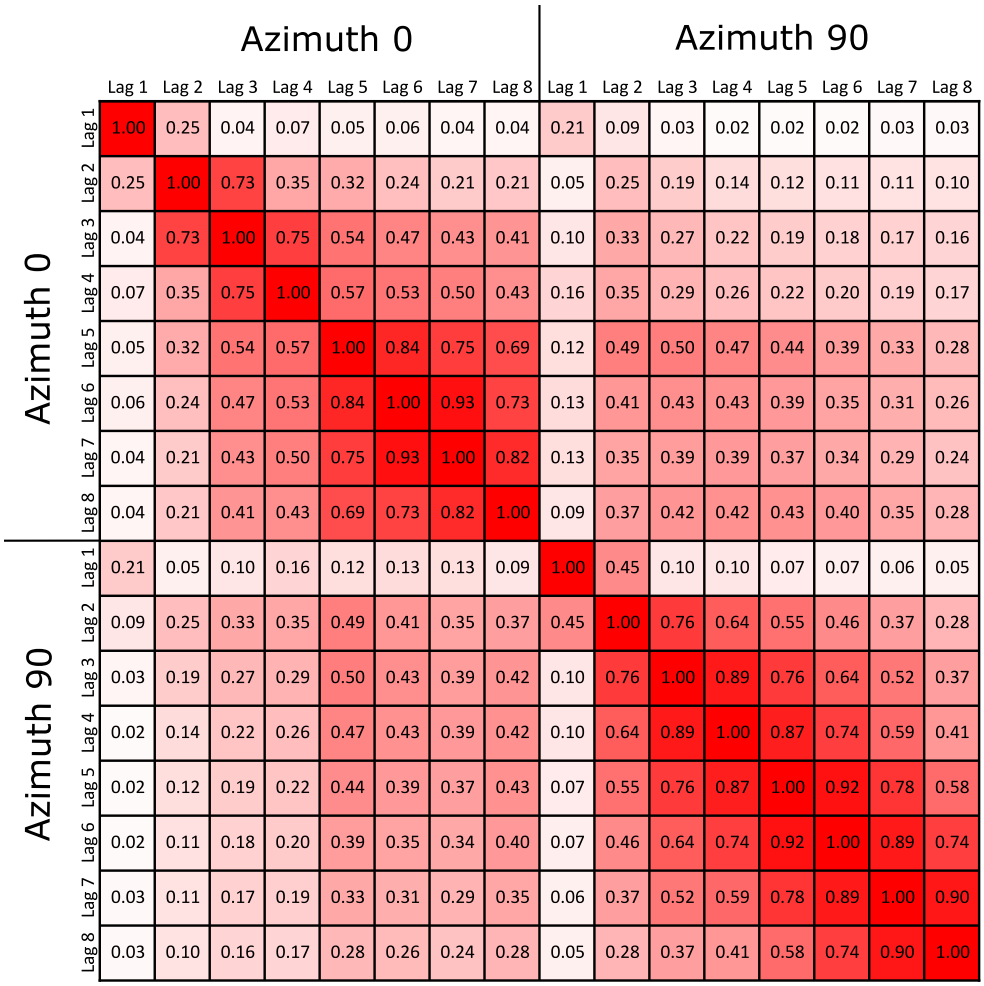

Figure 3.24: Correlation matrix between the sampled variogram realizations for eight lag distances of azimuth 0° and eight lag distances of azimuth 90°.

high continuities for realizations 2 and 5 because of high variogram ranges for azimuth 0° for both first and second structures. The conditioning of data does not allow to reproduce the large variogram range. This leads to lower uncertainty in the posterior variogram. Conditioning of data does not affect reproduction of a very low variogram range. For example, a pure nugget effect variogram can be reproduced by SGS realizations. Figures 3.29 shows the posterior variogram uncertainty (variogram of SGS realizations) in case of using variogram uncertainty (red dashed line) and without using variogram uncertainty (blue line). The posterior variogram uncertainty in case of using variogram uncertainty is just a little higher than without using variogram uncertainty. This is due to the conditioning of the data.

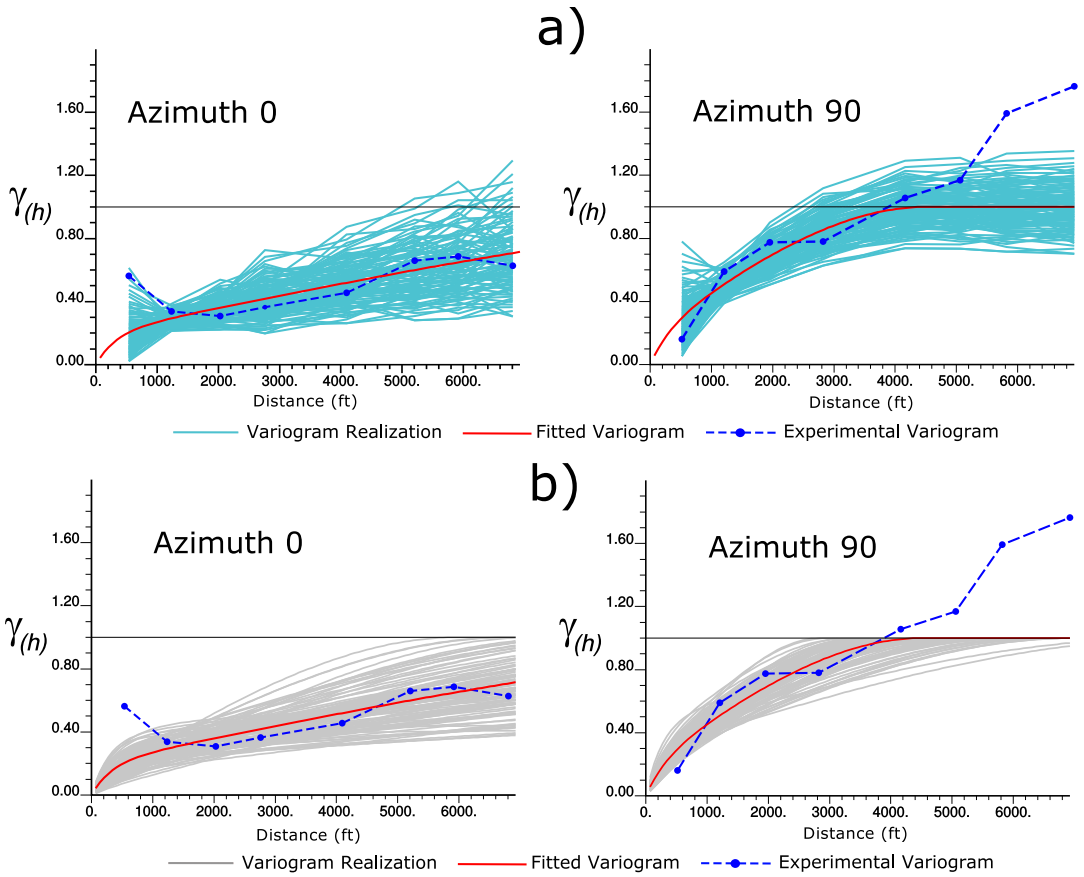

Figure 3.25: 100 correlated variogram realizations (a), fitted and standardized (b) for azimuths 0° and 90°.

3.5.1 Impact of Variogram Uncertainty

In reservoir modeling, variogram uncertainty could be applied for resource modeling or flow simulation (Figure 1.2 of Chapter 1); the impact of variogram uncertainty on the uncertainty of resource estimation such as hydrocarbon initially in place (HIIP), and uncertainty on reservoir performance (flow simulation) are considered in this Section.

For resource estimation, the posterior histogram uncertainty (histograms of realizations after geostatistical modeling) is very important because it is related to the uncertainty of the mean of the variable of interest for resource estimation. Figure 3.30 shows the posterior histogram uncertainty using variogram uncertainty and without using variogram uncertainty. The posterior histogram uncertainties are approximately equal: using variogram uncertainty leads to a little higher standard deviation of the mean and variance in comparison with not using variogram uncertainty: 0.4171 versus

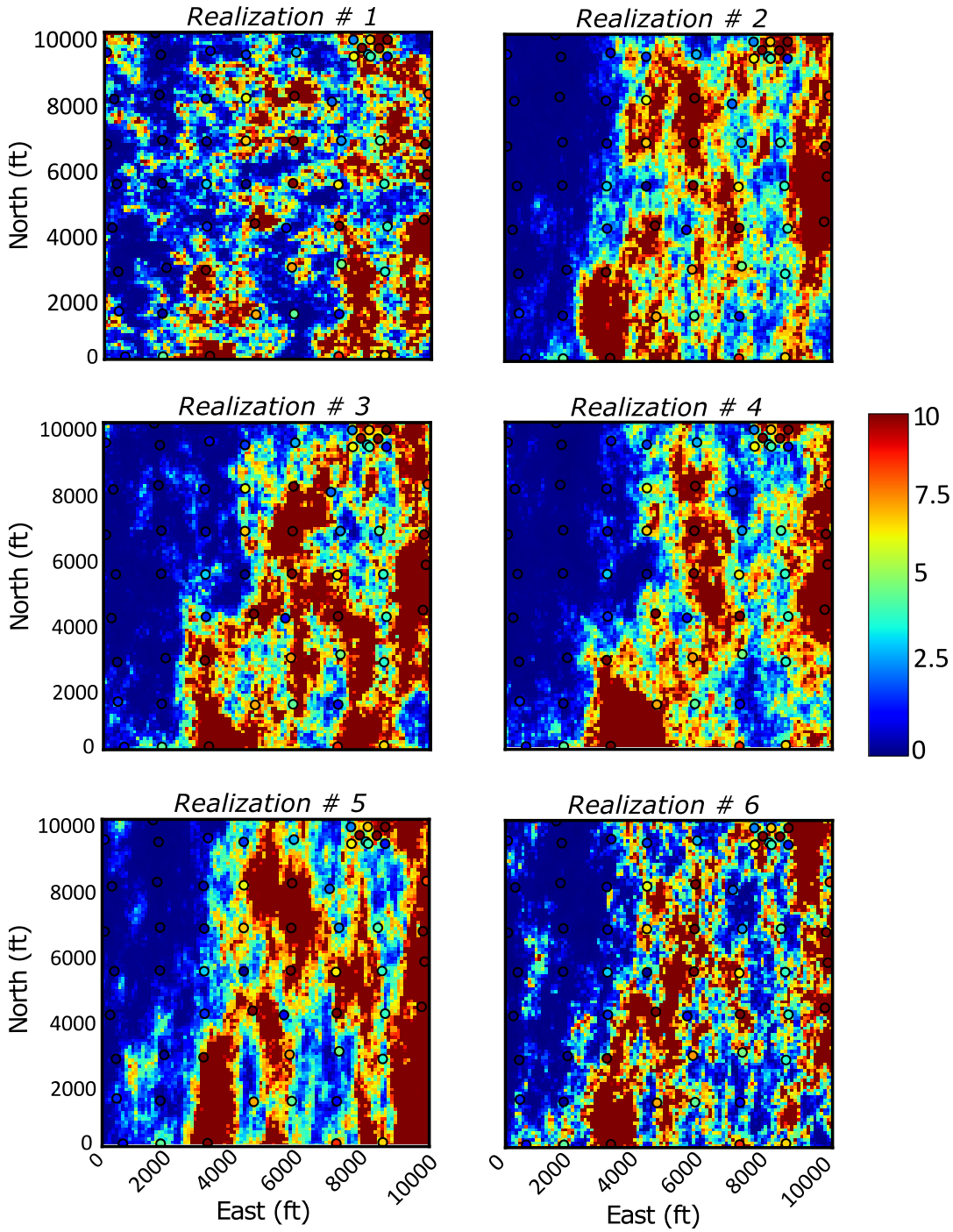

Figure 3.26: 6 SGS realizations with variogram uncertainty.

0.4110 for mean and, 0.3148 versus 0.2972 for variance respectively (see Figure 3.30). The reason for very low impact of variogram uncertainty on the posterior histogram uncertainty is because the high and low values are compensated in geostatistical modeling: the frequency of both high and low values are increased and decreased for high

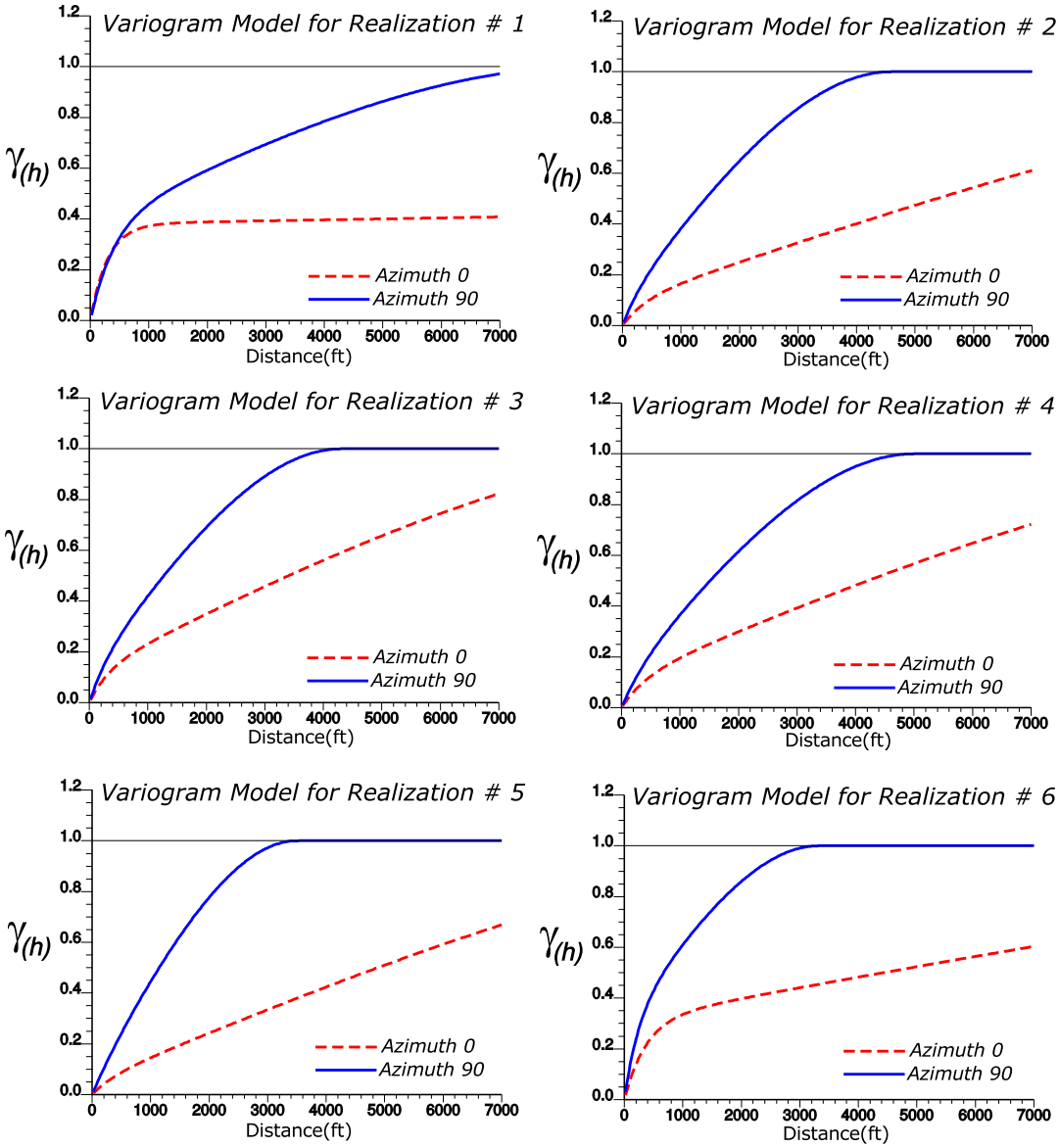

Figure 3.27: Variogram realizations used for SGS in Figure 3.26.

and low variogram ranges (see Figure 3.26). This leads to a little higher uncertainty in HIIP in case of using variogram uncertainty because the posterior histogram uncertainty of all variables associated with the HIIP calculation such as PHIE (effective porosity), NTG (Net To Gross), S_w (water saturation) and S_o (oil saturation) are not changing considerably after incorporating variogram uncertainty (see Chapter 7).

Variogram uncertainty is more influential in flow simulation because it affects the connectivity of rock properties. According to Meddaugh et al. (2011), the variogram could have a tremendous impact on waterflooding. The influence of variogram uncer-

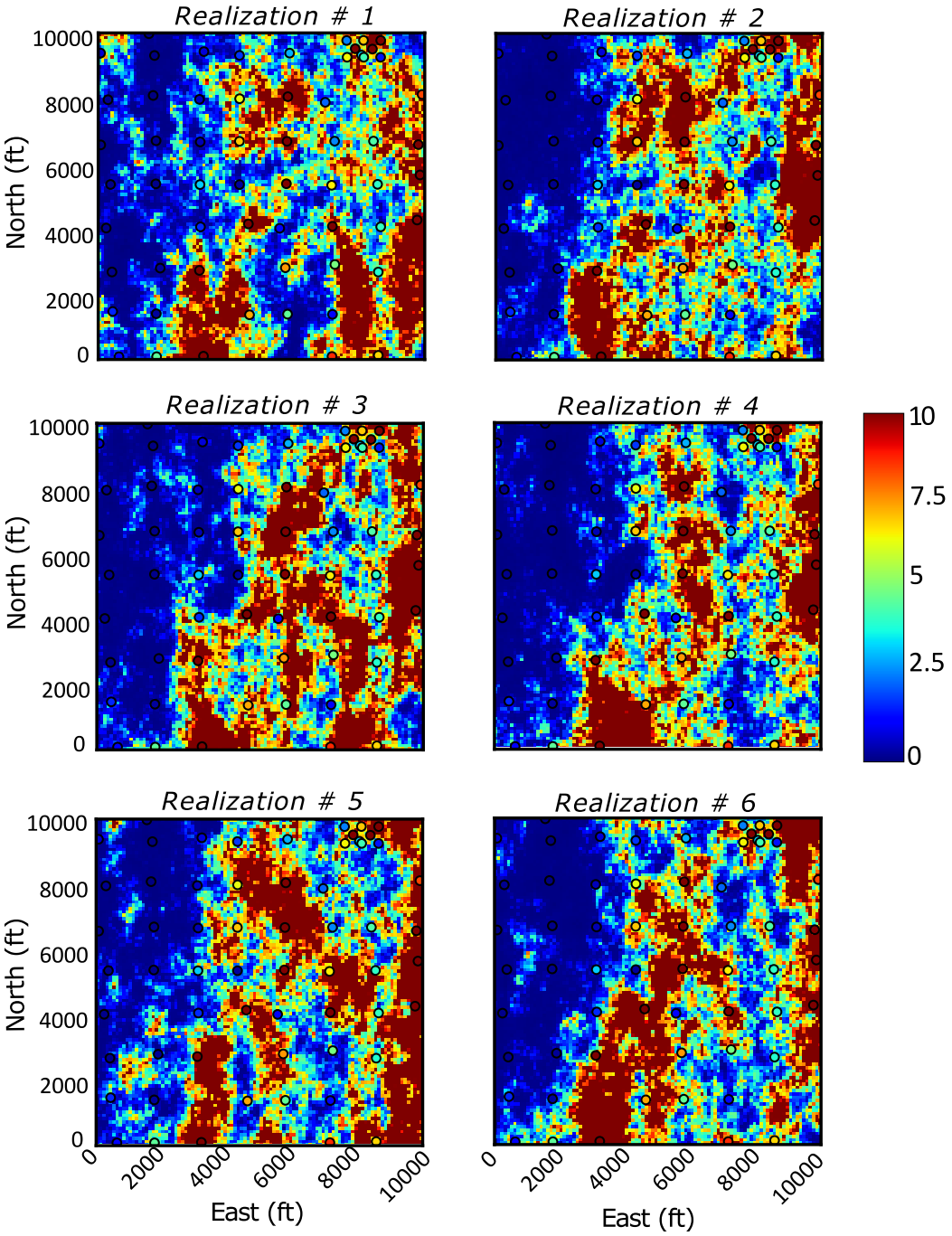

Figure 3.28: 6 SGS realizations without variogram uncertainty.

tainty on flow simulation is considered by the histogram of upscaled permeability of each 100 SGS realizations for the entire simulation area in case of both using variogram uncertainty and without using variogram. Since there is more continuity for azimuth 0°, upscaling of SGS realization is applied on this direction. Steady state

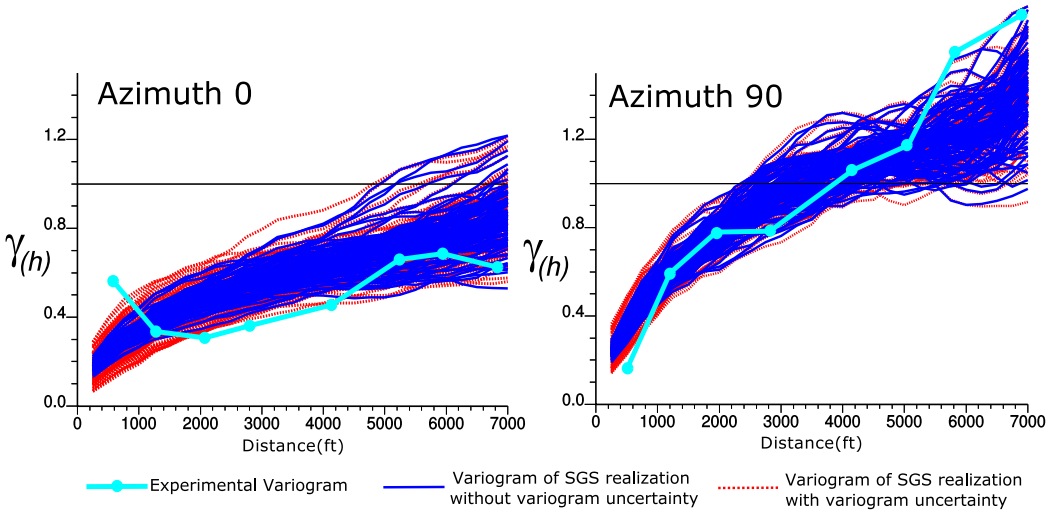

Figure 3.29: Posterior variogram uncertainty for using variogram uncertainty (dashes red line) and without using variogram uncertainty (blue line).

flow simulation is assumed for upscaling. Figure 3.31 shows the histograms of upscaled permeability with variogram uncertainty (a) and without variogram uncertainty (b). Using variogram uncertainty leads to higher uncertainty in the histogram than not using variogram uncertainty ($\sigma = 0.49683$ versus $\sigma = 0.42874$). The higher uncertainty in the histogram of upscaled permeability in presence of variogram uncertainty signifies the impact of variogram uncertainty on flow simulation.

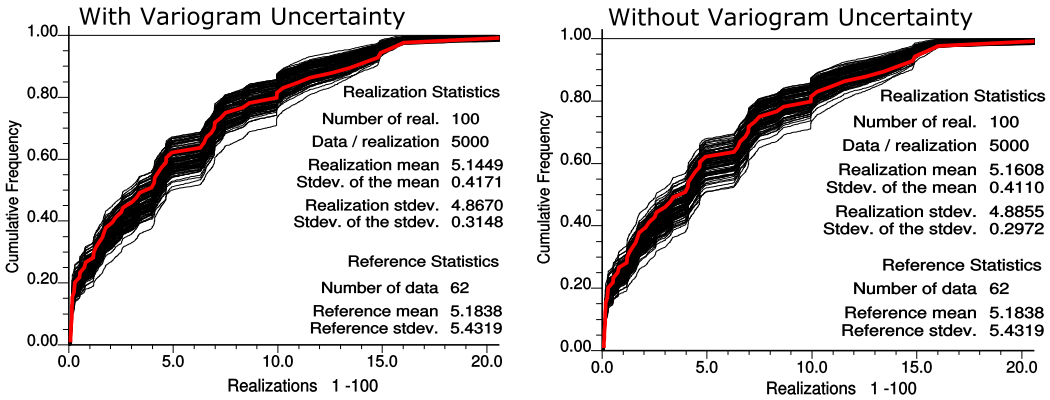

Figure 3.30: Posterior histogram uncertainty using variogram uncertainty and without using variogram uncertainty.

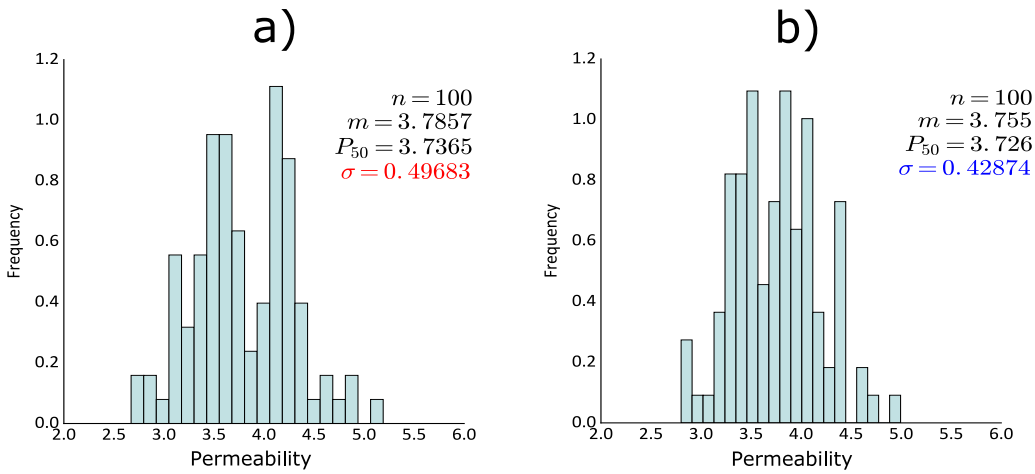

Figure 3.31: Histogram of upscaled permeability with steady state flow simulation using variogram uncertainty (a) and without using variogram uncertainty (b).

3.5.2 Remarks

A methodology for quantifying variogram uncertainty and incorporating this uncertainty in geostatistical modeling is provided in this Chapter. The distribution of each lag distance should be built first and the variogram realizations should be drawn from the distributions. The shape of the variogram distribution is considered to be a Chi-square distribution and it needs only the degrees of freedom (DoF) for each lag distance.

A new approach of quantifying variogram uncertainty (DoF approach) is proposed by direct calculation of degree of freedoms for each lag distance by fourth order covariance since other techniques has drawbacks: FOM (Marchant & Lark, 2004; Ortiz & Deutsch, 2002) has a very high uncertainty (for more details and examples see Ortiz and Deutsch (2002)) because of not considering the degrees of freedom of variogram pairs for calculating variogram uncertainty. Global kriging could also be used for quantifying variogram uncertainty. However, there is a big difference between the calculated variogram uncertainty by OGK and SGK, and global kriging is dependent to the size of domain. The larger the domain, the lower the variogram uncertainty and it leads to higher CPU time because of calculating the average fourth order covariance for the entire domain (right hand side covariance matrix). The new approach of quantifying variogram uncertainty by calculating degrees of freedom of variogram pairs is indepen-

dent from the domain size and it is fast and simple.

Variogram realizations are drawn from the Chi-square distributions to honor the correlation between lag distances. A methodology is presented that applies LU unconditional simulation based on the correlation matrix between lag distances. The correlation matrix is achieved by the average fourth order covariance between lag distances. Each realization is transformed to the marginal Chi-square distribution of each lag distance by quantile-quantile transformation. This ensures the correlation between lag distances is preserved. The variogram realizations are standardized to sill 1 before using in geostatistical simulation.

The calculated uncertainty in the variogram by the DoF approach is high because it is not conditioned to well data. This is similar to the spatial bootstrap for quantifying histogram uncertainty: the greater spatial correlation (variogram range), the higher variogram uncertainty. The calculated variogram uncertainty by the DoF approach is assumed as prior uncertainty in the variogram. The high prior variogram uncertainty is decreased during geostatistical modeling (posterior variogram uncertainty) due to conditioning of data: subsequent conditioning will make the result more expected. In general, conditioning of the data in geostatistical simulation improves unrealistic very high and low variogram ranges if there are enough data locations (say more than 80 wells). This correction is more noticeable for high range variograms; conditioning of data cannot affect a very short variogram range. In case of sparse well data (poor conditioning), high prior variogram uncertainty could be improved by secondary data such as the vertical variogram from well data and the horizontal variogram from seismic data (see Chapters 4 and 5).

The impact of variogram uncertainty for static resource estimation is low due to the fact that the high and low values are compensated in geostatistical modeling for different variogram realizations. Variogram uncertainty is more influential in flow simulation because it has a direct impact on the connectivity of rock properties.

Only 2D data sets are considered in this Chapter for variogram uncertainty. In case of 3D data set, variogram uncertainty is merely calculated for horizontal directions since there is no uncertainty in vertical direction due to many regularly spaced data. 3D

variogram realizations can be achieved by auto variogram modeling of the horizontal variogram realizations with a fixed vertical variogram.

The CPU time is an issue in case of many variogram pairs. It would not be advisable to calculate variogram uncertainty in presence of more than 10000 variogram pairs for each lag distance. In case of 2D data set, 10000 variogram pairs signifies trivial uncertainty in the variogram. Thus, variogram uncertainty is not required (similar to vertical variogram). For solving this problem in case of sparse wells for 3D data set, variogram uncertainty may be calculated after vertical upscaling the data because of reducing variogram pairs.

Chapter 4

Seismic Variogram to Improve Well Variogram

Variogram uncertainty was discussed in Chapter 3. Since the approach of quantifying variogram uncertainty (the DoF approach, Section 3.3) is not conditioned to well data, the calculated variogram uncertainty is high, which is decreased and improved by conditioning data through geostatistical modeling. However, in case of sparse well data, the seismic data (2D lines or 3D block) could be used to reduce the high variogram uncertainty.

The stable seismic variogram could be used in place of the well variogram (Wang & Dou, 2010). However, the processed seismic data are not the same physical attribute as that under consideration and the scale of measurement is much different. Ideally, the information from seismic data could be used to improve the uncertainty in the well variogram. To do this correctly, the spatial cross-correlation between well and seismic data should be taken into consideration to define the constraints the seismic variogram imposes on the well variogram. Seismic data are acquired in time domain that should be transferred to depth. The scale of seismic data is larger than well data so the variogram of the seismic data could be downscaled or the well data could be upscaled to ensure a consistent scale. For this Chapter, it is assumed that the well and seismic data are at the same scale. Variogram downscaling will be discussed in Chapter 5.

This Chapter presents three different methodologies to improve the uncertainty

in the experimental variogram of well data by seismic variogram: 1- Merge variogram distributions. 2- Global cokriging of the variogram pairs. 3- Seismic-derived constraints on the variogram of well data. These techniques use the cross covariance between well and seismic data to determine the relevance of the seismic variogram. Approaches 1 and 2 are early attempts and the drawbacks of these techniques are discussed. Approach 3 is considered the most reliable and can be reasonably applied for improving the uncertainty in the experimental well variogram. Only 2D data sets are shown in this Chapter, but the methodology will be applied for the horizontal variogram in case of 3D data set (Chapter 5).

4.1 Merge Variogram Distributions

The variogram distribution of each lag distance for well and seismic data could be merged to reduce the well variogram uncertainty. A method of combining probability distributions is required. The combined distribution should be convex: the mean of the combined distributions should fall within the mean of the input distributions (well and seismic) and close to the distribution with lower uncertainty. After merging well and seismic variogram distributions for each lag distance, the improved variogram realizations can be achieved by LU simulation with the correlation matrix of lag distances and back-transformed to the correct marginal distributions (see Chapter 3, Section 3.4). This approach needs a method of merging distributions with the mentioned properties (Section 4.1.1). Furthermore, the variogram distributions of each lag distance for well and seismic data should be calculated for this methodology. The uncertainty in the variogram of seismic data is negligible due to the exhaustive sampling. This leads to a combined variogram distribution almost exactly the same as the variogram of seismic data. To avoid this outcome, the variance of the seismic data variogram is calibrated with the cross covariance. The variance of the seismic variogram is increased to give it a reasonable weight.

4.1.1 Combining Independent Information Sources for Variogram Distributions

There are different approaches of combining distributions that could be applied for merging variogram distributions of well and seismic data for each lag distance. Combining different data sources (distributions) is discussed by Winkler (1968) in a Bayesian framework for merging information. Morris (1974, 1977) formally establishes a Bayesian approach. The Bayesian updating technique (Doyen, Den Boer, & Pillet, 1996) is introduced for data integration in geomodeling. In this method, the prior model is updated with the likelihood model built by all secondary data to attain the final posterior or updated model. However, there is no clear prior distribution for combining variogram distributions. Moreover, with Bayesian updating, the mean of the final distribution may not fall within the mean of the input distributions. Blachman (1989); Davis (2007); Roecker (1991) discuss merging multivariate independent Gaussian distributions based on an optimum method of combining error ellipses:

$$\overline{C} = (\sum_{i=1}^{n} C_i^{-1})^{-1} \quad , \quad \overline{\mu} = \overline{C} \sum_{i=1}^{n} (C_i^{-1} \mu_i) \tag{4.1}$$

where μ_i is the mean vector of i^{th} multivariate Gaussian distributions, C_i is the covariance matrix of i^{th} multivariate Gaussian distributions, $\overline{\mu}$ is the vector representing the location of the weighted average of n distributions and $\overline{C}$ is the resultant covariance matrix after merging error ellipses. The one-dimensional derivation of this approach is adapted to merge independent variogram distributions (well and seismic). This approach ensures convexity and the combined mean ($\overline{\mu}$) falls within the mean of the input distributions (μ_i , $i = 1, ..., n$), closest to the distribution with lower uncertainty. It seems reasonable for combining variogram distributions of well and seismic data for each lag distance to achieve an improved variogram if they are assumed to be independent from each other.

The Gaussian distribution of each lag distance is required to apply the error ellipse approach for merging variogram distributions (well and seismic). For simplicity, the

distribution of each lag distance is assumed Gaussian although the Chi-square distribution could be transformed to the Gaussian distribution by an anamorphosis function (see Chapter 5): the assumption of the Gaussian distribution of each lag distance instead of the Chi-square distribution (see Chapter 3) will not significantly change the final result because of the degrees of freedom and closeness of the Chi-squared distributions to Gaussian distributions. The Gaussian variable of each lag is denoted as $z_{i,\gamma(\mathbf{h})}$, $i = 1, ..., n$. Where n is multiple independent observations of the variogram distributions. Multiple independent observations of the mean and variance for a lag distance are denoted as $\overline{z_{i,\gamma(\mathbf{h})}}$ and $Var\left\{z_{i,\gamma(\mathbf{h})}\right\}$, respectively. The estimated mean $\overline{z^*_{\gamma(\mathbf{h})}}$ and variance $Var\left\{z^*_{\gamma(\mathbf{h})}\right\}$ are achieved by merging variogram distributions of independent observations. From the Equation 4.1 for one-dimensional combining independent Gaussian distributions, the estimation variance is written as:

$$Var\left\{z^*_{\gamma(\mathbf{h})}\right\} = \left[\sum_{i=1}^{n}\left(Var\left\{z_{i,\gamma(\mathbf{h})}\right\}\right)^{-1}\right]^{-1} \tag{4.2}$$

optimally weighted linear combination of independent lag distributions can be written as:

$$\overline{z^*_{\gamma(\mathbf{h})}} = Var\left\{z^*_{\gamma(\mathbf{h})}\right\}\sum_{i=1}^{n}\left(Var\left\{z_{i,\gamma(\mathbf{h})}\right\}\right)^{-1}\overline{z_{i,\gamma(\mathbf{h})}} \tag{4.3}$$

a linear aggregation with weights that are inversely proportional to the individual variance contributions is the solution for the optimal weights (for more information of combining independent Gaussian distributions see Davis (2007); Orechovesky (1996); Rezvandehy and Deutsch (2014a); Roecker (1991)). Figure 4.1 shows an example of merging two Gaussian distributions assumed as variogram distributions for a lag distance. Distribution 1 has a mean of 0.6 and variance 0.2, and distribution 2 has a mean 0.1 and variance 0.6. The merged distribution based on Equations 4.2 and 4.3, has a mean of 0.475 and variance 0.15. This approach enforces the mean of the combined distribution to fall within the mean of the input distributions, which is reasonable in this context. Moreover, the variance of the merged distribution is lower than the variance of the distributions 1 and 2 and close to distribution 1 because it has lower

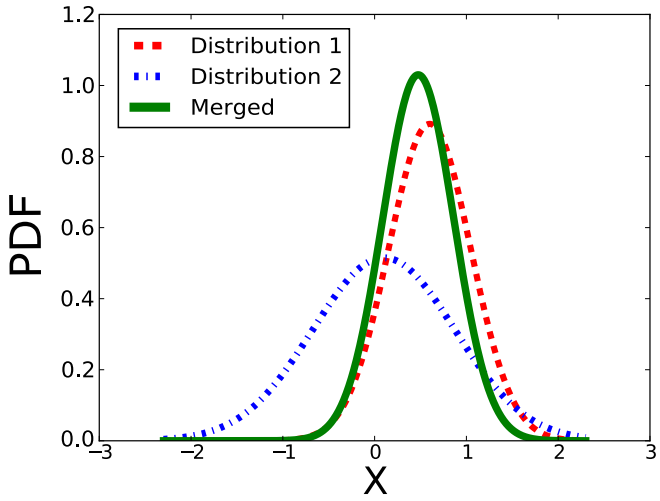

Figure 4.1: Combining distributions 1 and 2 by Equations 4.2 and 4.3.

variance than distribution 2. This could be applied for merging variogram distributions of well and seismic data for each lag distance to improve the high uncertainty in well variogram. The mean of each lag distance could be the fitted variogram model to the experimental variogram and the variance of each lag distance could be calculated by the DoF or FOM approaches (see Chapter 4). Since the DoF approach requires the Chi-square distribution of each lag distance, the FOM approach is used for calculation the variance of each distance in this experiment.

4.1.2 Estimation the Variance of the Seismic Variogram Given Well Variogram

The problem is the variance of the seismic variogram: since there are many variogram pairs for the seismic data, the variance of the seismic variogram for each lag distance will be close to zero. Applying the mentioned methodology (Section 4.1.1) for combining variogram distributions of well and seismic data leads to no weight to the well variogram and the final variogram is equal to the seismic variogram. To solve this problem, the variance of the seismic variogram will be modified based on the cross covariance between well and seismic data: a higher cross covariance makes the seismic variogram more informative. The estimated variance of the seismic variogram should increase as the cross covariance decreases. The relationship between the cross covariance and the

variance of the seismic variogram is fit by third degree polynomial curves. This shape is achieved based on an experiment (Rezvandehy & Deutsch, 2014c) from the positive definite covariance between well and seismic data:

$$Det(\begin{bmatrix} C_Z(\mathbf{h}) = 1 - \gamma_Z(\mathbf{h}) & C_{YZ}(\mathbf{h}) \\ C_{YZ}(\mathbf{h}) & C_Y(\mathbf{h}) = 1 - \gamma_Y(\mathbf{h}) \end{bmatrix}) \geq 0 \quad \Rightarrow \quad \gamma_Y(\mathbf{h}) \leq 1 - \frac{C_{YZ}(\mathbf{h})^2}{1 - \gamma_Z(\mathbf{h})} \tag{4.4}$$

where $C_Z(\mathbf{h})$ and $C_Y(\mathbf{h})$ are covariances of the well and seismic data for lag distance $\mathbf{h}$, $\gamma_Z(\mathbf{h})$ and $\gamma_Y(\mathbf{h})$ are variograms for standardized data, and $C_{YZ}(\mathbf{h})$ is the cross covariance. The variograms of the well and seismic should satisfied in Equation (right) to ensure positive definite covariance. According to this equation, $\gamma_Y(\mathbf{h})$ has the highest variance when $C_{YZ}(\mathbf{h}) = 0$ because it can be any value between 0 and 1. As the $C_{YZ}(\mathbf{h})$ increases, the variance of $\gamma_Y(\mathbf{h})$ drops quickly; it is approximately flatten around the $C_{YZ}(\mathbf{h}) = 0.5$ and it drops sharply again for very high $C_{YZ}(\mathbf{h})$ since a very limited range of $\gamma_Y(\mathbf{h})$ is satisfied in Equation 4.4. Figure 4.2 shows 100 possible curves. The absolute value of the cross covariance is always between 0 and 1 for the standardized or normal score data. The variance of the seismic variogram is within 0 to 5 where 5 denotes a very high variogram uncertainty (practically infinite in Gaussian units).

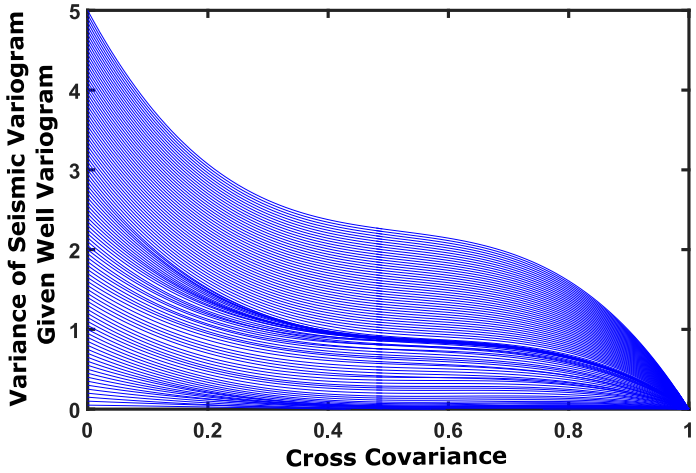

Figure 4.2: 100 third degree polynomial curves for relationship between the cross covariance and the variance of the seismic variogram given well variogram.

As the cross covariance decreases, the variance of the seismic variogram increases. An appropriate curve that reasonably represents the relationship between the cross covari-

ance and the variance of seismic given well data could be achieved by an experimental approach: synthetic well and seismic data are simulated by unconditional sequential Gaussian simulation (SGS). Well and seismic data should have a good correlation ($\rho(0)$ above 0.5). Some samples are drawn from the simulated well data. The experimental variogram of the drawn samples is calculated and a variogram model is fitted. The variance of each lag distance is calculated by the FOM approach and the mean is the fitted variogram model for each lag distance. The seismic variogram and the cross covariance between well and seismic are calculated and fitted. The mean of the seismic variogram distribution is the fitted model but the variance comes from the curves in Figure 4.2 according to the calculated cross covariance for each lag distance. The distribution of each lag distance of the well and seismic data variograms are combined by Equations 4.2 and 4.3 to attain the updated variogram. The mean square error (MSE) between the updated variogram and the real variogram of well data is calculated. This process is repeated for all 100 curves. The curve with minimum MSE is the most accurate curve.

Figure 4.3 shows the methodology. Figure 4.3-a shows the simulated well data by SGS and drawn 36 samples. The variogram uncertainty for the drawn samples is calculated. Figure 4.3-b shows the simulated seismic data by SGS, that has a high correlation with the well data. The variogram of seismic data and the cross covariance between well and seismic are calculated. Since the real variogram of drawn samples is known, the MSE between the update well variogram and the real variogram is calculated (Figure 4.3-c). The updated variogram is calculated by merging the well and seismic variogram distributions (Equations 4.2 and 4.3) using the variance of the seismic variogram from the curves (Figure 4.2). This process is repeated for several synthetic well and seismic data. The curve that makes the average MSE minimum is the most accurate curve for the variance of the seismic variogram given well variogram.

Variogram models are required for the unconditional SGS. Three variogram structures (spherical) with no nugget effect are considered. One variogram structure applies to the well data, and one variogram structure applies to the seismic data. The well and

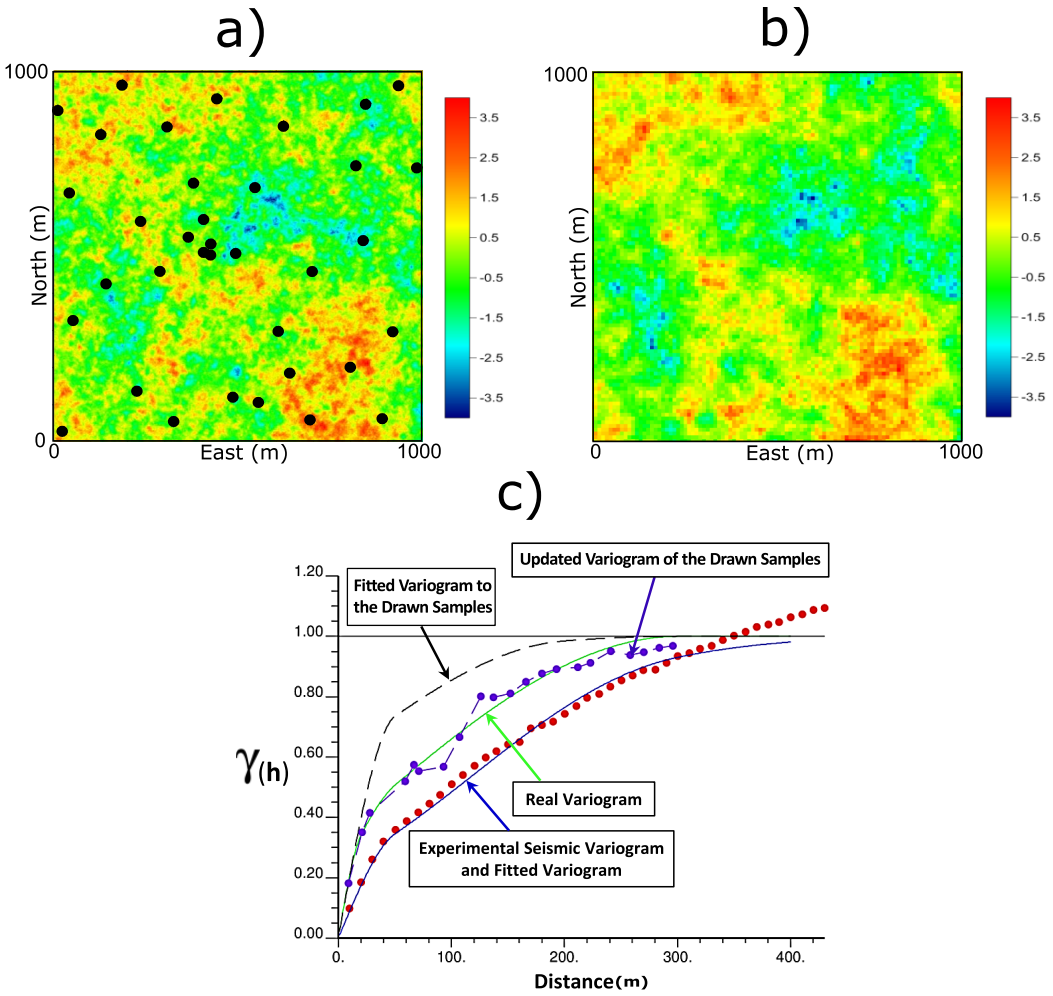

Figure 4.3: a) Simulated well data by SGS and drawn 36 samples. b) Simulated seismic data by SGS that has a high correlation with the well data (a). c) Merging the well and seismic variogram distributions using seismic variance given well variogram from the curves (Figure 4.2) and calculate the MSE between the updated variogram and the real variogram.

seismic data and the variograms are written as:

$$Y_w = a_1.y_1 + a_2.y_2 + a_w.y_w \quad \rightarrow \quad \gamma_w = a_1^2.\Gamma_1(\mathbf{h}) + a_2^2.\Gamma_2(\mathbf{h}) + a_w^2.\Gamma_w(\mathbf{h}) \tag{4.5}$$

where Y_w is well data with 3 geological structures: y_1, y_2 and y_w. y_1 and y_2 are equal with seismic data, and y_w is the unique low range structure for well data. γ_w is the variogram model of well data.

$$Y_s = a_1.y_1 + a_2.y_2 + a_s.y_s \quad \rightarrow \quad \gamma_s = a_1^2.\Gamma_1(\mathbf{h}) + a_2^2.\Gamma_2(\mathbf{h}) + a_s^2.\Gamma_s(\mathbf{h}) \tag{4.6}$$

where Y_s is seismic data with 3 geological structures: y_1, y_2 and y_s. y_1 and y_2 are equal with well data, and y_s is the unique high range structure for seismic data. γ_s is the variogram model of seismic data. According to LMC model (Pyrcz & Deutsch, 2014), there are three constraints for the unknown parameters in Equations 4.5 and 4.6:

$$\begin{cases} a_1^2 + a_2^2 + a_w^2 = 1 \\ b_1^2 + b_2^2 + b_s^2 = 1 \\ a_1.b_1 + a_2.b_2 = \rho(0) \end{cases} \tag{4.7}$$

$\rho(0)$ is the correlation coefficient between well and seismic at zero lag distance. Monte Carlo simulation is applied to define the unknown parameters (a and b) by satisfying Equation 4.7. The same variogram structures between well and seismic data and the unique structures are assumed $\Gamma_1(\mathbf{h}) = 50m$, $\Gamma_2(\mathbf{h}) = 300m$, $\Gamma_w(\mathbf{h}) = 20m$ and $\Gamma_s(\mathbf{h}) = 400m$. According to these parameters, synthetic well and seismic data are simulated with the different correlations for the mentioned methodology. The mean of the calculated MSEs are achieved for each curve (Figure 4.2). The most accurate curve has the minimum average MSE. By simulating 30 synthetic well and seismic data via SGS, the most accurate curve is found. Figure 4.4 shows the mean of MSE for 30 simulated data sets versus the curve number. Based on this experiment, curve 37 is the most accurate curve for demonstrating the relationship between the variance of the seismic variogram given well variogram and the cross covariance (Rezvandehy & Deutsch, 2014c).

4.1.3 Limitation of the Technique

The curve for defining the variance of the seismic variogram given the cross covariance could change from one data set to another because of many variables. The range of the variance of the seismic variogram given well variogram for curves is assumed to be 0 to 5 for this experiment (see Figure 4.2). This range is reasonable for this case, but when the data configuration and number of samples are changed, the variogram uncertainty may decrease or increase. Each case could consider new parameters to simulate synthetic

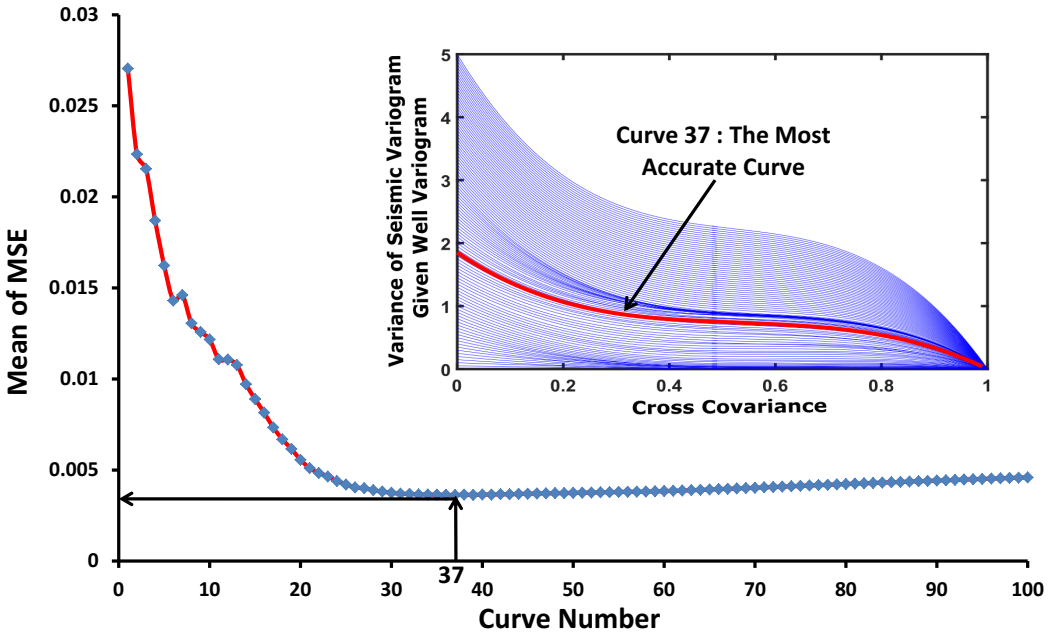

Figure 4.4: Mean of MSE for 30 simulated data set versus the curve number. The most accurate curve is 37 with the lowest MSE.

well and seismic data (see Figure 4.3). The samples from the synthetic well data are drawn based on the data configuration of the real data. According to the calculated variogram uncertainty for the drawn samples, a range of the variance of the seismic variogram given well variogram (from 0 to the highest possible value) for curves are defined. Then, the most accurate curve is achieved for the data set based on the minimum MSE. However, in practice, it is a time consuming process to find the curve for each data set especially in case of 3D data set. Thus, it is not an efficient technique for improving the uncertainty of well variogram by the seismic variogram.

4.2 Global Cokriging of the Variogram Pairs

A global kriging of variogram pairs was discussed in Chapter 2 to decluster the experimental variogram. The variogram pairs of the exhaustive seismic data could be used to estimate the sparse variogram pairs of well data by global cokriging. There are different approaches of cokriging (Goovaerts, 1998): 1- simple cokriging (SCK), 2- ordinary cokriging (OCK), 3- standardized form of the SCK (correlogram notation), 4- standardized form of the SCK with constraint the sum of all weights to 1. These

techniques can estimate a primary variable using secondary data for the entire domain (global cokriging). The simple global cokriging (SGCK) estimator of z for the entire domain A is written:

$$Z^*_{SGCK} - m_1 = \sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1}^{SGCK}[Z_1(\mathbf{u}_{\alpha_1}) - m_1] + \sum_{\alpha_2=1}^{n_2} \lambda_{\alpha_2}^{SGCK}[Z_2(\mathbf{u}_{\alpha_2}) - m_2] \tag{4.8}$$

where m_1 and m_2 are stationary means for primary (well data) and secondary (seismic data) data in the cokriging system. $\lambda_{\alpha_1}^{SGCK}$ is the weight of global mean and variance assigned to the primary datum $z_1(\mathbf{u}_{\alpha_1})$, and $\lambda_{\alpha_2}^{SGCK}$ is the weight of global mean and variance assigned to the secondary datum $z_2(\mathbf{u}_{\alpha_2})$. n_1 and n_2 are the number of primary and secondary data.

The ordinary global cokriging (OGCK) estimator of z for the entire domain A is written:

$$Z^*_{OGCK} = \sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1}^{OGCK}[Z_1(\mathbf{u}_{\alpha_1})] + \sum_{\alpha_2=1}^{n_2} \lambda_{\alpha_2}^{OGCK}[Z_2(\mathbf{u}_{\alpha_2})] \tag{4.9}$$

there is no need for stationary mean of primary and secondary variable in the OGCK system due to the following constraints:

$$\sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1}^{OGCK} = 1 \quad , \quad \sum_{\alpha_2=1}^{n_2} \lambda_{\alpha_2}^{OGCK} = 0 \tag{4.10}$$

the stationary mean of primary (m_1) and secondary (m_2) variables are canceled out in OGCK system. This approach is inefficient: it reduces the influence of the secondary information due to very small or negative weights to satisfy the constraint that the sum of secondary weights be 0.

The standardized form of simple global kriging is achieved by scaling the auto and cross-covariances in SGCK in terms of correlograms. This is useful when the variances of primary and secondary variables differ considerably, which leads to a large difference between the covariances and a risk of numerical instability (Goovaerts, 1998). The standardized form of SGCK is obtained by dividing the terms in SGCK (see Equation

4.8) system by the standard deviations of the variables:

$$\frac{Z^*_{SGCK} - m_1}{\sigma_1} = \sum_{\alpha_1=1}^{n_1} \upsilon_{\alpha_1}^{SGCK}[\frac{Z_1(\mathbf{u}_{\alpha_1}) - m_1}{\sigma_1}] + \sum_{\alpha_2=1}^{n_2} \upsilon_{\alpha_2}^{SGCK}[\frac{Z_2(\mathbf{u}_{\alpha_2}) - m_2}{\sigma_2}] \quad (4.11)$$

where σ_1 is the standard deviation of the primary variable and σ_2 is the standard deviation of the secondary variable. $\upsilon_{\alpha_1}^{SGCK}$ and $\upsilon_{\alpha_2}^{SGCK}$ are the weights of the standardized form of SGCK assigned to the primary datum and secondary datum.

The standardized form of SGCK is more efficient than SGCK for application with the variogram pairs of the well and seismic data due to the fact that the variance of the quadratic differences of variogram pairs for well and seismic data (fourth order covariance of each pair to itself) might differ by several orders of magnitude. Since there is no constraint on the primary and secondary weights, it might lead to unbalanced weights (very low or very high to primary and secondary). A constraint on the standardized form of SGCK is proposed to make the sum of all primary and secondary weights to 1 with the following system of linear equations:

$$\begin{cases} \sum_{\beta_1=1}^{n_1} \upsilon_{\beta_1}^{SCGK} C_{11}(\mathbf{u}_{\beta_1} - \mathbf{u}_{\alpha_1})/\sigma_1\sigma_1 + \sum_{\beta_2=1}^{n_2} \upsilon_{\beta_2}^{SGCK} C_{12}(\mathbf{u}_{\beta_2} - \mathbf{u}_{\alpha_1})/\sigma_1\sigma_2 \\ \qquad +\mu = \overline{C_{11}(\mathbf{u}_{\alpha_1} - A)}/\sigma_1\sigma_1 \quad , \quad \alpha_1 = 1....n_1 \\ \\ \sum_{\beta_1=1}^{n_1} \upsilon_{\beta_1}^{SGCK} C_{12}(\mathbf{u}_{\beta_1} - \mathbf{u}_{\alpha_2})/\sigma_1\sigma_2 + \sum_{\beta_2=1}^{n_2} \upsilon_{\beta_2}^{SGCK} C_{22}(\mathbf{u}_{\beta_2} - \mathbf{u}_{\alpha_2})/\sigma_2\sigma_2 \\ \qquad +\mu = \overline{C_{12}(\mathbf{u}_{\alpha_2} - A)}/\sigma_1\sigma_2 \quad , \quad \alpha_2 = 1....n_2 \\ \\ \sum_{\beta_1=1}^{n_1} \upsilon_{\beta_1}^{SCGK} + \sum_{\beta_2=1}^{n_2} \upsilon_{\beta_2}^{SCGK} = 1 \end{cases} \quad (4.12)$$

where n_1 is number of data for primary variable, and n_2 is number of data for secondary variable. $C_{11}(\mathbf{u}_\beta - \mathbf{u}_\alpha)$ is the covariance between locations $\mathbf{u}_\beta$ and $\mathbf{u}_\alpha$ for primary variable; $C_{12}(\mathbf{u}_\beta - \mathbf{u}_\alpha)$ is the cross covariance between locations $\mathbf{u}_\beta$ and $\mathbf{u}_\alpha$; and $C_{22}(\mathbf{u}_\beta - \mathbf{u}_\alpha)$ is the covariance between locations $\mathbf{u}_\beta$ and $\mathbf{u}_\alpha$ for secondary variable. $\overline{C_{11}(\mathbf{u}_\alpha - A)}$ is the average covariance between each data location $\mathbf{u}_\alpha$ and the whole area A for primary variable, and $\overline{C_{12}(\mathbf{u}_\alpha - A)}$ is the average cross covariance between each data location $\mathbf{u}_\alpha$ and the whole area A. μ is Lagrange parameter. The estimation

variance is written as:

$$Var\{Z^*_{SGCK}\} = \overline{C_{11}(A-A)} - \sum_{\alpha_1=1}^{n_1} \upsilon^{SGCK}_{\alpha_1} \overline{C_{11}(\mathbf{u}_{\alpha_1}-A)} - \sum_{\alpha_2=1}^{n_2} \upsilon^{SGCK}_{\alpha_2} \overline{C_{12}(\mathbf{u}_{\alpha_2}-A)} - \mu \tag{4.13}$$

where $\overline{C_{11}(A-A)}$ is the average covariance of the entire area to itself (Goovaerts, 1997) for primary variables.

The standardized form of SGCK for estimating the well variogram pairs with the exhaustive seismic variogram pairs can be written:

$$\frac{X^*_{s,\mathbf{h}} - \gamma_1(\mathbf{h})}{\sigma_1'} = \sum_{\alpha_1=1}^{n_1(\mathbf{h})} \upsilon^{SGCK}_{\alpha_1,\mathbf{h}} \left[\frac{X_{1,\alpha_1,\mathbf{h}}(\mathbf{u}) - \gamma_1(\mathbf{h})}{\sigma_1'}\right] + \sum_{\alpha_2=1}^{n_2(\mathbf{h})} \upsilon^{SGCK}_{\alpha_2,\mathbf{h}} \left[\frac{X_{2,\alpha_2,\mathbf{h}}(\mathbf{u}) - \gamma_2(\mathbf{h})}{\sigma_2'}\right] \tag{4.14}$$

where $\gamma_1(\mathbf{h})$ and $\gamma_2(\mathbf{h})$ are the fitted variogram models to the experimental variogram of well and seismic data assumed as the stationary means for each lag distance. σ_1' and σ_2' are the standard deviations of the half of the quadratic differences for variogram pairs of well and seismic data. $n_1(\mathbf{h})$ and $n_2(\mathbf{h})$ are the number of variogram pairs for well and seismic data. $X^*_{s,\mathbf{h}}$ is the estimated variogram (standardized form of SGCK) by the variogram pairs of well data $X_{1,\mathbf{h}}(\mathbf{u})$, and seismic data $X_{2,\mathbf{h}}(\mathbf{u})$ for each lag distance $\mathbf{h}$:

$$\begin{array}{cc} Z_1(\mathbf{u}+\frac{\mathbf{h}}{2}) & Z_2(\mathbf{u}+\frac{\mathbf{h}}{2}) \\ \uparrow & , \quad \uparrow \\ Z_1(\mathbf{u}-\frac{\mathbf{h}}{2}) & Z_2(\mathbf{u}-\frac{\mathbf{h}}{2}) \end{array} \qquad \begin{aligned} X_{1,\mathbf{h}}(\mathbf{u}) &= \frac{[Z_1(\mathbf{u}-\frac{\mathbf{h}}{2}) - Z_1(\mathbf{u}+\frac{\mathbf{h}}{2})]^2}{2} \\ X_{2,\mathbf{h}}(\mathbf{u}) &= \frac{[Z_2(\mathbf{u}-\frac{\mathbf{h}}{2}) - Z_2(\mathbf{u}+\frac{\mathbf{h}}{2})]^2}{2} \end{aligned} \tag{4.15}$$

where $Z_1(\mathbf{u}-\frac{\mathbf{h}}{2})$ is the tail value, and $Z_1(\mathbf{u}+\frac{\mathbf{h}}{2})$ is the head value of the well variogram pair $\uparrow$ for lag distance $\mathbf{h}$; $Z_2(\mathbf{u}-\frac{\mathbf{h}}{2})$ is the tail value, and $Z_2(\mathbf{u}+\frac{\mathbf{h}}{2})$ is the head value of the seismic variogram pair $\uparrow$ for lag distance $\mathbf{h}$.

The weights of the standardized SGCK for variogram pairs of each lag distance $\mathbf{h}$, $\upsilon^{SGCK}_{\alpha_1,\mathbf{h}}$ for well data and $\upsilon^{SGCK}_{\alpha_2,\mathbf{h}}$ for seismic data, are obtained by solving the following

system of linear equations:

$$\begin{cases} \sum_{\beta_1=1}^{n_1(\mathbf{h})} \upsilon_{\beta_1,\mathbf{h}}^{SCGK} F_{11}(\mathbf{u}_{\beta_1}-\mathbf{u}_{\alpha_1})/\sigma_1'\sigma_1' + \sum_{\beta_2=1}^{n_2(\mathbf{h})} \upsilon_{\beta_2,\mathbf{h}}^{SGCK} F_{12}(\mathbf{u}_{\beta_2}-\mathbf{u}_{\alpha_1})/\sigma_1'\sigma_2' \\ \qquad +\mu = \overline{F_{11}(\mathbf{u}_{\alpha_1}-A)}/\sigma_1'\sigma_1' \quad , \quad \alpha_1 = 1....n_1(\mathbf{h}) \\ \\ \sum_{\beta_1=1}^{n_1(\mathbf{h})} \upsilon_{\beta_1}^{SGCK} F_{12}(\mathbf{u}_{\beta_1}-\mathbf{u}_{\alpha_2})/\sigma_1'\sigma_2' + \sum_{\beta_2=1}^{n_2(\mathbf{h})} \upsilon_{\beta_2}^{SGCK} F_{22}(\mathbf{u}_{\beta_2}-\mathbf{u}_{\alpha_2})/\sigma_1'\sigma_2' \\ \qquad +\mu = \overline{F_{12}(\mathbf{u}_{\alpha_2}-A)}/\sigma_1'\sigma_2' \quad , \quad \alpha_2 = 1....n_2(\mathbf{h}) \\ \\ \sum_{\beta_1=1}^{n_1(\mathbf{h})} \upsilon_{\beta_1,\mathbf{h}}^{SCGK} + \sum_{\beta_2=1}^{n_2(\mathbf{h})} \upsilon_{\beta_2,\mathbf{h}}^{SCGK} = 1 \end{cases} \tag{4.16}$$

where $F_{11}(\mathbf{u}_\beta - \mathbf{u}_\alpha)$ is the fourth order covariance between variogram pairs of $\mathbf{u}_\beta$ and $\mathbf{u}_\alpha$ for well data; $F_{12}(\mathbf{u}_\beta - \mathbf{u}_\alpha)$ is the fourth order cross covariance between variogram pairs of $\mathbf{u}_\beta$ and $\mathbf{u}_\alpha$; and $F_{22}(\mathbf{u}_\beta - \mathbf{u}_\alpha)$ is the fourth order covariance between variogram pairs of $\mathbf{u}_\beta$ and $\mathbf{u}_\alpha$ for seismic data. $\overline{F_{11}(\mathbf{u}_\alpha - A)}$ is the average fourth covariance between each variogram pair $\mathbf{u}_\alpha$ and the whole area A for well data, and $\overline{F_{12}(\mathbf{u}_\alpha - A)}$ is the average fourth order cross covariance between each variogram pair of $\mathbf{u}_\alpha$ and the whole area A for seismic data. The fourth order cross covariance can be calculated by cross second order moments (cross covariance/variogram between well and seismic data). The estimation variance of SGCK for each lag distance is written as:

$$\begin{aligned} Var(2X_{s,\mathbf{h}}^*) = \overline{F_{11}(A-A)} - \sum_{\alpha_1=1}^{n_1(\mathbf{h})} \upsilon_{\alpha_1,\mathbf{h}}^{SGCK} \overline{F_{11}(\mathbf{u}_{\alpha_1}-A)} \\ - \sum_{\alpha_2=1}^{n_2(\mathbf{h})} \upsilon_{\alpha_2,\mathbf{h}}^{SGCK} \overline{F_{12}(\mathbf{u}_{\alpha_2}-A)} - \mu \quad \rightarrow \quad Var\left\{\hat{\gamma}(\mathbf{h})\right\} = \frac{1}{4} \times \left(Var\left\{2X_{s,\mathbf{h}}^*\right\}\right) \end{aligned} \tag{4.17}$$

where $\overline{F_{11}(A-A)}$ is the average fourth order covariance of the entire domain A to itself. The average fourth order covariance was discussed in Chapter 2 (see Figure 2.3). $Var\left\{\hat{\gamma}(\mathbf{h})\right\}$ is the variance of each lag distance by global cokriging of the well and seismic variogram pairs.

Figure 4.5 shows location maps of synthetic well and seismic data to demonstrate how the standardized form of SGCK can be applied on variogram pairs of the well and seismic data for lag distance 4m. There are 9 well data and 25 seismic data. The

estimated variogram pair for lag distance $\mathbf{h} = 4m$ of this example by the standardized form of SGCK is written as:

$$\frac{X^*_{s,\mathbf{h}} - \gamma_1(\mathbf{h})}{\sigma_1^{'}} = \sum_{\alpha_1=1}^{9} \upsilon^{SGCK}_{\alpha_1,\mathbf{h}} [\frac{X_{1,\alpha_1,\mathbf{h}}(\mathbf{u}) - \gamma_1(\mathbf{h})}{\sigma_1^{'}}] + \sum_{\alpha_2=1}^{25} \upsilon^{SGCK}_{\alpha_2,\mathbf{h}} [\frac{X_{2,\alpha_2,\mathbf{h}}(\mathbf{u}) - \gamma_2(\mathbf{h})}{\sigma_2^{'}}] \tag{4.18}$$

where $\gamma_1(\mathbf{h})$ and $\gamma_2(\mathbf{h})$ are the fitted variogram models to well and seismic data. $\sigma_1^{'}$ and $\sigma_2^{'}$ are the standard deviations of the half of the quadratic differences for variogram pairs of well and seismic data.

Equation 4.19 shows the fourth order covariance matrix for the standardized SGCK for 9 variogram pairs of well data and 25 variogram pairs of seismic data:

$$\begin{bmatrix} F_{11}(1-1)/\sigma_1^{'}\sigma_1^{'} & .. & F_{11}(1-9)/\sigma_1^{'}\sigma_1^{'} & F_{12}(1-1)/\sigma_1^{'}\sigma_2^{'} & .. & F_{12}(1-25)/\sigma_1^{'}\sigma_2^{'} & 1 \\ : & & : & : & & : & : \\ F_{11}(9-1)/\sigma_1^{'}\sigma_1^{'} & .. & F_{11}(9-9)/\sigma_1^{'}\sigma_1^{'} & F_{12}(9-1)/\sigma_1^{'}\sigma_2^{'} & .. & F_{12}(9-25)/\sigma_1^{'}\sigma_2^{'} & 1 \\ F_{12}(1-1)/\sigma_1^{'}\sigma_2^{'} & .. & F_{12}(1-9)/\sigma_1^{'}\sigma_2^{'} & F_{22}(1-1)/\sigma_2^{'}\sigma_2^{'} & .. & F_{22}(1-25)/\sigma_2^{'}\sigma_2^{'} & 1 \\ : & & : & : & & : & : \\ F_{12}(25-1)/\sigma_1^{'}\sigma_2^{'} & .. & F_{12}(25-9)/\sigma_1^{'}\sigma_2^{'} & F_{22}(25-1)/\sigma_2^{'}\sigma_2^{'} & .. & F_{22}(25-25)/\sigma_2^{'}\sigma_2^{'} & 1 \\ 1 & .. & 1 & 1 & .. & 1 & 0 \end{bmatrix}$$

$$\times \begin{bmatrix} \upsilon^{SCGK}_{\alpha_1=1,\mathbf{h}} \\ : \\ \upsilon^{SCGK}_{\alpha_1=9,\mathbf{h}} \\ \upsilon^{SCGK}_{\alpha_2=1,\mathbf{h}} \\ : \\ \upsilon^{SCGK}_{\alpha_2=25,\mathbf{h}} \\ \mu \end{bmatrix} = \begin{bmatrix} \overline{F_{11}(\mathbf{u}_{\alpha_1=1} - A)}/\sigma_1^{'}\sigma_1^{'} \\ : \\ \overline{F_{11}(\mathbf{u}_{\alpha_1=9} - A)}/\sigma_1^{'}\sigma_1^{'} \\ \overline{F_{12}(\mathbf{u}_{\alpha_2=1} - A)}/\sigma_1^{'}\sigma_2^{'} \\ : \\ \overline{F_{12}(\mathbf{u}_{\alpha_2=25} - A)}/\sigma_1^{'}\sigma_2^{'} \\ 1 \end{bmatrix} \tag{4.19}$$

This covariance matrix can be calculated by the reference variogram models of the well and seismic data $\gamma_1(\mathbf{h})$ and $\gamma_2(\mathbf{h})$, and the standard deviations of the half of the quadratic differences for variogram pairs of the well and seismic data ($\sigma_1^{'}$ and $\sigma_2^{'}$).

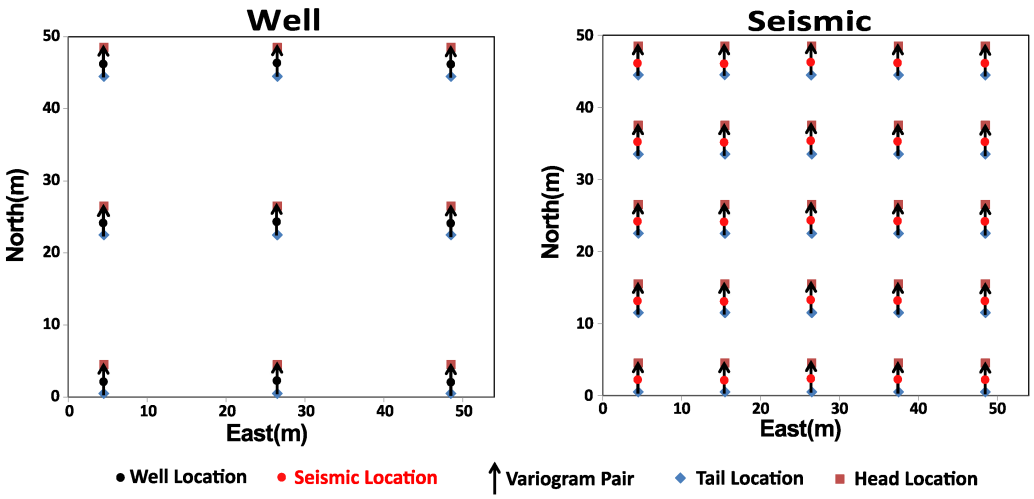

Figure 4.5: Location maps of a synthetic well and seismic data. There are 9 well data and 25 seismic data, which leads to the same number of variogram pairs for lag distance 4m of each data.

These parameters for standardized data are assumed as:

$$\begin{cases} \gamma_1(\mathbf{h}) = 0.20 Sph_{\substack{h_{major}=15m \\ h_{minor}=15m}}(\mathbf{h}) + 0.80 Sph_{\substack{h_{major}=30m \\ h_{minor}=30m}}(\mathbf{h}) \\ \gamma_2(\mathbf{h}) = 0.25 Sph_{\substack{h_{major}=15m \\ h_{minor}=15m}}(\mathbf{h}) + 0.75 Sph_{\substack{h_{major}=30m \\ h_{minor}=30m}}(\mathbf{h}) \\ \sigma_1^{'} = 1.2 \ , \ \sigma_2^{'} = 0.9 \end{cases} \tag{4.20}$$

where Sph is the spherical variogram model. h_{major} and h_{minor} are variogram ranges for major and minor directions of continuity, which are equal for isotropic data. A cross covariance model (cross variogram) between well and seismic data is achieved by an intrinsic model (Markov model) (Pyrcz & Deutsch, 2014). The covariance matrix is solved for two correlations between the well and seismic data: $\rho(0) = 0.8$ and $\rho(0) = 0.4$. Table 4.1 shows sum of the weights for the variogram pairs of the well data $v_{\alpha_1,\mathbf{h}}^{SGCK}$ and seismic data $v_{\alpha_2,\mathbf{h}}^{SGCK}$. The standardized form of SGCK enforces sum of all weights be 1. For high correlation ($\rho(0) = 0.8$), sum of the weights to the seismic pairs (0.81184) is much higher than the sum of weights to the well pairs (0.18816). As correlation decreases to $\rho(0) = 0.4$, sum of the weights to the well pairs increases from 0.18816 to 0.40508, and sum of the weights to the seismic pairs diminishes from 0.81184 to

0.59492. The weights are used in Equation 4.18 to estimate variogram for lag distance 4m using the variogram pairs of seismic data.

Table 4.1: Sum of the weights for variogram pairs of the well and seismic data (Figure 4.5) for two correlations: $\rho(0) = 0.8$ and $\rho(0) = 0.4$.

$\rho(0) = 0.8$		$\rho(0) = 0.4$	
Sum of the Weights to Well Pairs	Sum of the Weights to Seismic Pairs	Sum of the Weights to Well Pairs	Sum of the Weights to Seismic Pairs
0.18816	0.81184	0.40508	0.59492

4.2.1 Synthetic Example

A 2D synthetic example is considered to assess the proposed methodology of global cokriging for estimating the well data variogram. Synthetic well and seismic data are simulated by unconditional sequential Gaussian simulation (SGS). Figure 4.6 shows synthetic well (left Figure) and seismic data (right Figure) which are at the same scale and have a correlation coefficient of approximately 0.6. The grid has $nx = 50 \times ny = 50$. Instead of drawing well locations, variogram pairs (head and tale locations) are drawn from the simulated well and seismic data to estimate well variogram pairs with the seismic variogram pairs by SGCK. This avoids taking many variogram pairs for one lag distance and no pair or very few pairs for another lag distance of well data. Figure 4.7 shows the middle points of the variogram pairs where tale and head locations are drawn from the simulated well and seismic data for azimuth 0°. Sparse well pairs and exhaustive seismic pairs are drawn to apply the global cokriging. The experimental variograms of the well and seismic are calculated and variogram models are fit:

$$\begin{cases} \gamma_1(\mathbf{h}) = 0.05 + 0.45Exp_{\substack{h_{major}=10m \\ h_{minor}=10m}}(\mathbf{h}) + 0.40Exp_{\substack{h_{major}=30m \\ h_{minor}=30m}}(\mathbf{h}) \\ \gamma_2(\mathbf{h}) = 0.20 + 0.40Exp_{\substack{h_{major}=10m \\ h_{minor}=10m}}(\mathbf{h}) + 0.35Exp_{\substack{h_{major}=30m \\ h_{minor}=30m}}(\mathbf{h}) \end{cases} \tag{4.21}$$

where $\gamma_1(\mathbf{h})$ is the fitted model to the experimental well variogram, and $\gamma_2(\mathbf{h})$ is the fitted model to the experimental seismic variogram. The experimental seismic variogram

is well-defined because of the exhaustive sampling. *Exp* is the exponential variogram model. h_{major} and h_{minor} are variogram ranges for major and minor directions of continuity, which are equal for isotropic data. A cross covariance model (cross variogram) between well and seismic data achieved by an intrinsic model (Markov model) (Pyrcz & Deutsch, 2014) is required to calculate the fourth order cross covariance between well and seismic pairs.

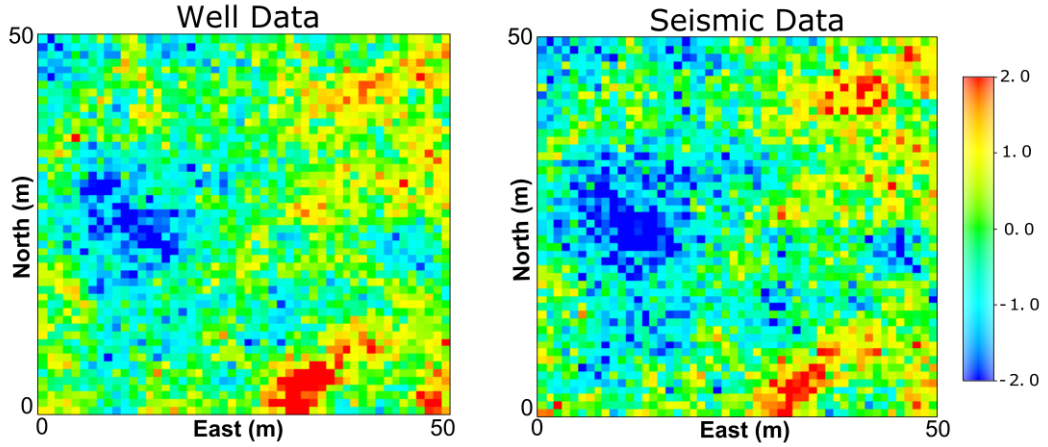

Figure 4.6: Synthetic well (left Figure) and seismic data (right Figure) simulated by unconditional SGS.

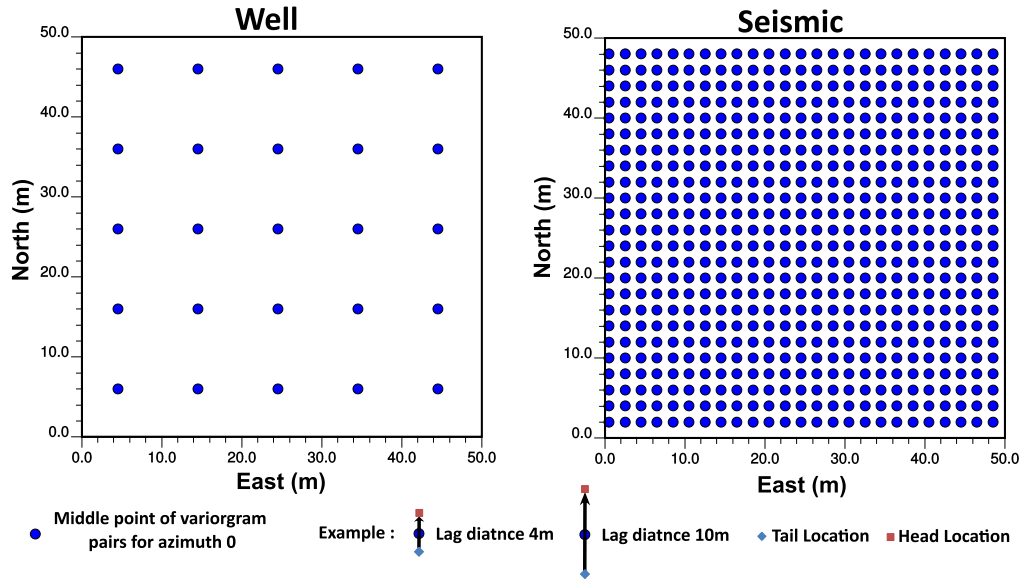

Figure 4.7: Middle points of the drawn variogram pairs (tale and head locations) for azimuth 0° for well (left Figure) and seismic data (right Figure).

Figure 4.8 shows the result of the cokriging for the well and seismic variogram

pairs by the standardized form of SGCK with the constraint that the weights sum to 1 (Equation 4.16). The experimental well variogram is noisy. Although SGCK removes noise and artifacts in the experimental variogram, the updated variogram converges to the fitted variogram model and not the real variogram. The estimated mean of the simple cokriging for variogram pairs converges to the stationary mean of the primary

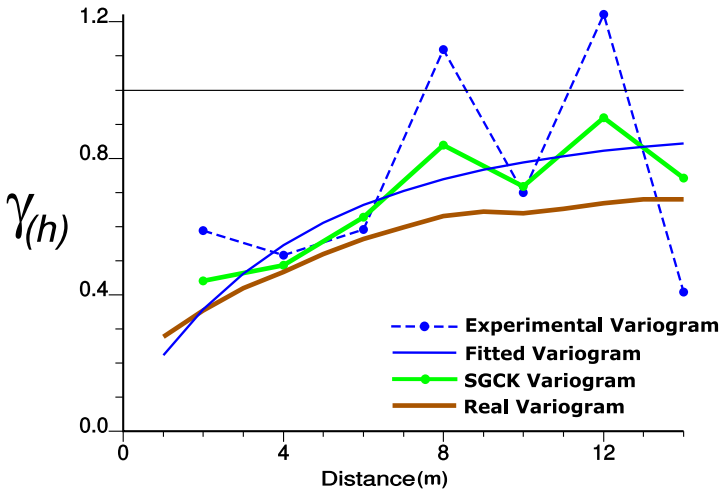

Figure 4.8: Combination of the well and seismic variogram pairs by the standardized form of SGCK with constraint that the weights sum to 1. The SGCK variogram converges to the fitted variogram model not the real variogram of well data.

variable which is the fitted variogram model to the experimental well variogram ($\gamma_1(\mathbf{h})$). All approaches of cokriging needs stationary means for both primary and secondary variables, which leads to convergence to the stationary mean of the primary variable (Goovaerts, 1998), except ordinary cokriging because the sum of the weights to primary variable is 1, and the sum of the weights to secondary variable is zero (see Equation 4.9). Ordinary cokriging is an inefficient technique due to the fact that it reduces the influence of the secondary information to nothing. Therefore, the cokriging approach is also considered inappropriate to improve the variogram of the well data by the seismic variogram.

4.3 Seismic-derived Variogram of the Well Data

The covariance matrix between well and seismic data for a variogram pair at each lag distance provides a constraint to keep the results positive definite (Pyrcz & Deutsch, 2014) for geostatistical modeling. The only unknown covariance in the covariance matrix is the covariance of the well data because the covariance of the seismic data and the cross covariance are well-defined. The acceptable range of the unknown covariance of the well data for each lag distance could be achieved by enforcing this covariance matrix to be positive definite assuming the well and seismic covariances are at same scale. The acceptable range for each lag distance leads to seismic-derived limits to the well data variogram. Variogram realizations of well data should be within these limits.

A totally positive matrix (Gantmacher & Krein, 1950) of well and seismic data is proposed to attain the seismic-derived variogram because it is more stable than a positive semi definite matrix. A square matrix is totally positive if the determinant of any corresponding square submatrix (including both principal and nonprincipal) are positive (Gantmacher & Krein, 1950; Vandebril, Van Barel, & Mastronardi, 2008).

For building this covariance matrix, it is assumed that for each data location, well data is present as well as seismic data. A variogram pair of each lag distance $\mathbf{h}$ in presence of well and seismic data is written:

$$\begin{matrix} Y(\mathbf{u}+\frac{\mathbf{h}}{2}), Z(\mathbf{u}+\frac{\mathbf{h}}{2}) \\ \uparrow \\ Y(\mathbf{u}-\frac{\mathbf{h}}{2}), Z(\mathbf{u}-\frac{\mathbf{h}}{2}) \end{matrix} \tag{4.22}$$

where $Y(\mathbf{u}-\frac{\mathbf{h}}{2})$ and $Z(\mathbf{u}-\frac{\mathbf{h}}{2})$ are the tail locations, and $Y(\mathbf{u}+\frac{\mathbf{h}}{2})$ and $Z(\mathbf{u}+\frac{\mathbf{h}}{2})$ are the head locations for seismic $Y(\mathbf{u})$ and well $Z(\mathbf{u})$ variables of each variogram pair $\uparrow$ and lag distance $\mathbf{h}$, respectively. The covariance matrix between $Y(\mathbf{u}-\frac{\mathbf{h}}{2})$, $Y(\mathbf{u}+\frac{\mathbf{h}}{2})$,

$Z(\mathbf{u}-\frac{\mathbf{h}}{2})$, and $Z(\mathbf{u}+\frac{\mathbf{h}}{2})$ for standardized data (variance 1) is written as:

$$A = \begin{array}{c} \\ Y(\mathbf{u}-\frac{\mathbf{h}}{2}) \\ Z(\mathbf{u}-\frac{\mathbf{h}}{2}) \\ Y(\mathbf{u}+\frac{\mathbf{h}}{2}) \\ Z(\mathbf{u}+\frac{\mathbf{h}}{2}) \end{array} \begin{array}{c} \begin{array}{cccc} Y(\mathbf{u}-\frac{\mathbf{h}}{2}) & Z(\mathbf{u}-\frac{\mathbf{h}}{2}) & Y(\mathbf{u}+\frac{\mathbf{h}}{2}) & Z(\mathbf{u}+\frac{\mathbf{h}}{2}) \end{array} \\ \begin{bmatrix} 1 & C_{YZ}(0) & C_Y(\mathbf{h}) & C_{YZ}(\mathbf{h}) \\ C_{YZ}(0) & 1 & C_{YZ}(\mathbf{h}) & C_Z(\mathbf{h}) =? \\ C_Y(\mathbf{h}) & C_{YZ}(\mathbf{h}) & 1 & C_{YZ}(0) \\ C_{YZ}(\mathbf{h}) & C_Z(\mathbf{h}) =? & C_{YZ}(0) & 1 \end{bmatrix} \end{array} \tag{4.23}$$

where $C_Y(\mathbf{h})$ is the covariance function of the seismic data, which is always well-defined. $C_{YZ}(\mathbf{h})$ is the cross covariance function between well and seismic data. The cross covariance function is fairly-well defined. The diagonal elements of this covariance matrix are the variance of the well and seismic data. These diagonal elements are 1 in case of standardized well and seismic data. The only unknown function of this covariance matrix is the covariance function of well data $C_Z(\mathbf{h})$. This covariance should be between 0 and 1 for standardized data (variance 1). The acceptable covariances of well data based on seismic data make the covariance matrix of Equation 4.23 positive definite: the determinant of matrix A should be positive $Det\{A\} > 0$:

$$\begin{aligned} Det\{A\} = -C_Z^2(\mathbf{h}).(1-C_Y^2(\mathbf{h})) - C_Z(\mathbf{h}).(2.C_Y(\mathbf{h}).C_{YZ}^2(0) + 2.C_Y(\mathbf{h}).C_{YZ}^2(\mathbf{h}) - \\ 4.C_{YZ}(0).C_{YZ}(\mathbf{h})) - C_Y^2(\mathbf{h}) + 4.C_Y(\mathbf{h}).C_{YZ}(0).C_{YZ}(\mathbf{h}) - 2.C_{YZ}^2(0).C_{YZ}^2(\mathbf{h}) + \\ C_{YZ}^4(0) - 2.C_{YZ}^2(0) + C_{YZ}^4(\mathbf{h}) - 2.C_{YZ}^2(\mathbf{h}) + 1 \end{aligned} \tag{4.24}$$

where $Det\{A\}$ based on the unknown parameter $C_Z(\mathbf{h})$, is a quadratic function $-ax^2 \pm bx \pm c$, $0 \le x \le 1$. For each lag distance $\mathbf{h}$, many regular samples of $C_Z(\mathbf{h}) \in [0,1]$ are drawn (say 10000). The range of $C_Z(\mathbf{h})$ that has a positive determinant $Det\{A\} > 0$, are calculated.

Another constraint may be taken from the totally positive matrix A. All order leading principal and nonprincipal submatrices of matrix A should be positive definite. All order leading principal submatrices are always positive. For example, 1^{st} order

leading principal submatrix which include the unknown $C_Z(\mathbf{h})$ is:

$$B = \begin{array}{c} \\ Z(\mathbf{u}-\frac{\mathbf{h}}{2}) \\ Y(\mathbf{u}+\frac{\mathbf{h}}{2}) \\ Z(\mathbf{u}+\frac{\mathbf{h}}{2}) \end{array} \begin{array}{c} \begin{array}{ccc} Z(\mathbf{u}-\frac{\mathbf{h}}{2}) & Y(\mathbf{u}+\frac{\mathbf{h}}{2}) & Z(\mathbf{u}+\frac{\mathbf{h}}{2}) \end{array} \\ \begin{bmatrix} 1 & C_{YZ}(\mathbf{h}) & C_Z(\mathbf{h}) =? \\ C_{YZ}(\mathbf{h}) & 1 & C_{YZ}(0) \\ C_Z(\mathbf{h}) =? & C_{YZ}(0) & 1 \end{bmatrix} \end{array} \tag{4.25}$$

the determinant of the matrix B is always positive for any $C_Z(\mathbf{h}) \in [0,1]$. Hence, another constraint could be taken from the nonprincipal submatrix of matrix A:

$$C = \begin{array}{c} \\ Y(\mathbf{u}-\frac{\mathbf{h}}{2}) \\ Z(\mathbf{u}-\frac{\mathbf{h}}{2}) \end{array} \begin{array}{c} \begin{array}{cc} Y(\mathbf{u}+\frac{\mathbf{h}}{2}) & Z(\mathbf{u}+\frac{\mathbf{h}}{2}) \end{array} \\ \begin{bmatrix} C_Y(\mathbf{h}) & C_{YZ}(\mathbf{h}) \\ C_{YZ}(\mathbf{h}) & C_Z(\mathbf{h}) =? \end{bmatrix} \end{array} \quad , \quad Det\{C\} = C_Z(\mathbf{h}).C_Y(\mathbf{h}) - C^2_{YZ}(\mathbf{h}) \tag{4.26}$$

the determinant of this matrix should be positive ($Det\{C\} > 0$) to make matrix A totally positive. Moreover, this ensures the covariance matrix between well and seismic data for each lag distance to be positive definite. This is the condition of the linear model of coregionalization (LMC) for variogram modeling (Pyrcz & Deutsch, 2014). The function of the determinant matrix C based on the unknown parameter $C_Z(\mathbf{h})$ is a linear function $ax \pm b$. Upper and lower limits of seismic-derived covariance/variogram for each lag distance can be achieved by enforcing $Det\{A\} > 0$ and $Det\{C\} > 0$ for many regular samples of $C_Z(\mathbf{h}) \in [0,1]$ (say 10000). Figure 4.9 shows the general upper and lower limits of seismic-derived covariance based on positive determinant of the matrix A (Equations 4.25) and C (Equation 4.26). There is a minimum acceptable positive determinant δ. This is true for both positive and negative correlations between well and seismic data: the determinants of the matrices A and C are equal for both

positive and negative correlations:

$$A = \begin{bmatrix} 1 & -C_{YZ}(0) & C_Y(\mathbf{h}) & -C_{YZ}(\mathbf{h}) \\ -C_{YZ}(0) & 1 & -C_{YZ}(\mathbf{h}) & C_Z(\mathbf{h}) \\ C_Y(\mathbf{h}) & -C_{YZ}(\mathbf{h}) & 1 & -C_{YZ}(0) \\ -C_{YZ}(\mathbf{h}) & C_Z(\mathbf{h}) & -C_{YZ}(0) & 1 \end{bmatrix} = \begin{bmatrix} 1 & C_{YZ}(0) & C_Y(\mathbf{h}) & C_{YZ}(\mathbf{h}) \\ C_{YZ}(0) & 1 & C_{YZ}(\mathbf{h}) & C_Z(\mathbf{h}) \\ C_Y(\mathbf{h}) & C_{YZ}(\mathbf{h}) & 1 & C_{YZ}(0) \\ C_{YZ}(\mathbf{h}) & C_Z(\mathbf{h}) & C_{YZ}(0) & 1 \end{bmatrix} \tag{4.27}$$

$$C = \begin{bmatrix} C_Y(\mathbf{h}) & -C_{YZ}(\mathbf{h}) \\ -C_{YZ}(\mathbf{h}) & C_Z(\mathbf{h}) \end{bmatrix} = \begin{bmatrix} C_Y(\mathbf{h}) & C_{YZ}(\mathbf{h}) \\ C_{YZ}(\mathbf{h}) & C_Z(\mathbf{h}) \end{bmatrix} \tag{4.28}$$

for example:

$$Det(\begin{bmatrix} 1 & -0.65 & 0.8 & -0.45) \\ -0.65 & 1 & -0.45 & 0.7 \\ 0.8 & -0.45 & 1 & -0.65 \\ -0.45 & 0.7 & -0.65 & 1 \end{bmatrix}) = Det(\begin{bmatrix} 1 & 0.65 & 0.8 & 0.45) \\ 0.65 & 1 & 0.45 & 0.7 \\ 0.8 & 0.45 & 1 & 0.65 \\ 0.45 & 0.7 & 0.65 & 1 \end{bmatrix}) = 0.037 \tag{4.29}$$

$$Det(\begin{bmatrix} 0.8 & -0.45 \\ -0.45 & 0.7) \end{bmatrix}) = Det(\begin{bmatrix} 0.8 & 0.45 \\ 0.45 & 0.7) \end{bmatrix}) = 0.357 \tag{4.30}$$

Since the range of the acceptable covariances are achieved for the standardized data, seismic-derived covariance can be simply converted to seismic-derived variogram upper and lower limits $\gamma(\mathbf{h}) = 1 - C(\mathbf{h})$. The upper and lower limits of the seismic-derived variogram may be taken from the cumulative distribution functions (CDF) of $\gamma(\mathbf{h})$, for example, P_{10} for lower limit and P_{90} for upper limit, this ensures more reasonable upper and lower limits. This process is repeated for all lag distances. Figure 4.10 shows a synthetic example of upper and lower limit seismic derived-variograms of the well data for different cross covariances. The correlation between well and seismic data increases from Figure 4.10-a to d (see the cross covariances). The higher the correlation between well and seismic data, the lower the difference between the upper and lower limits. Figure 4.10-d illustrates when the well and seismic data have very

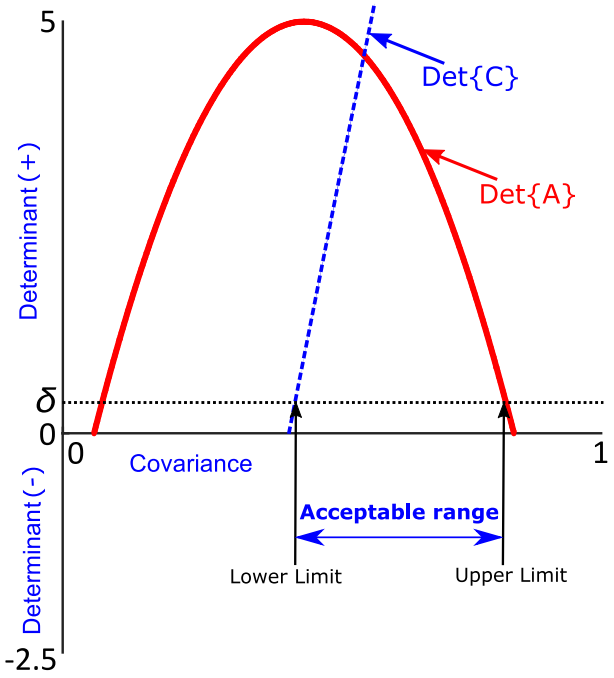

Figure 4.9: Upper and lower limits of seismic-derived covariance based on positive determinant of matrix A (Equation 4.24) and C (Equation 4.26). δ is a minimum acceptable positive determinant.

a high correlation (more than 0.9) with each other, the limits of the seismic-derived variogram are consistent with the seismic variogram. If the correlation is very low (less than 0.2), the variogram of the seismic data does not provide any constraint on the well variogram (Figure 4.10-a) because upper and lower limits of the seismic-derived variogram are very far from each other.

The experimental variogram of the seismic data is very important to calculate the limits of the seismic-derived variogram. Using lag tolerance and azimuth tolerance for variogram calculation leads to huge CPU cost because of the large number of seismic data. A new approach of variogram calculation is proposed by defining a template around the head location of variogram calculation: instead of using one pair, several pairs are assigned based on a template around the head location of variogram calculation. This is repeated for all lag distances, which leads to more reasonable variogram of seismic data and it is so fast.

Figure 4.11 shows a 2D and 3D schematic of using a template in variogram calculation for azimuth 45°. The yellow and red cells are tail and head locations of a

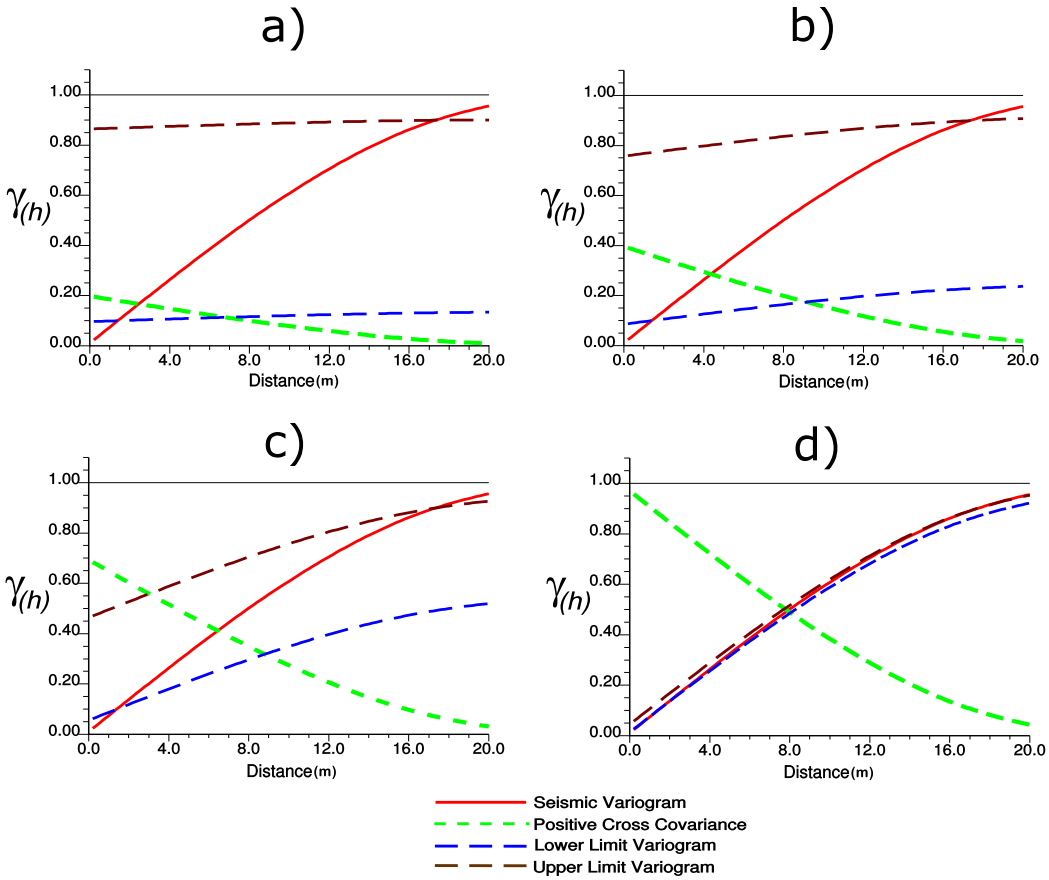

Figure 4.10: Upper and lower limits of the seismic-derived variograms for different cross covariances, increasing from (a) to (d). The higher the cross covariance, the more consistent the seismic-derived variogram and seismic variogram.

pair for variogram calculation. Instead of using one pair (black pair), several pairs are contributed (blue pairs). These pairs are achieved by assigning a template around the head location. The closest pairs to the head location are used for variogram calculation. This approach could be extended for calculating the cross variogram (cross covariance) between the well and seismic data. In this case, the tail of the variogram pairs are well data and the head are seismic data. For more information and details of this methodology see Rezvandehy and Deutsch (2014b). Calculating the cross variogram (cross covariance) at the scale of the well data is a challenge due to the fact that seismic data have higher scale than well data.

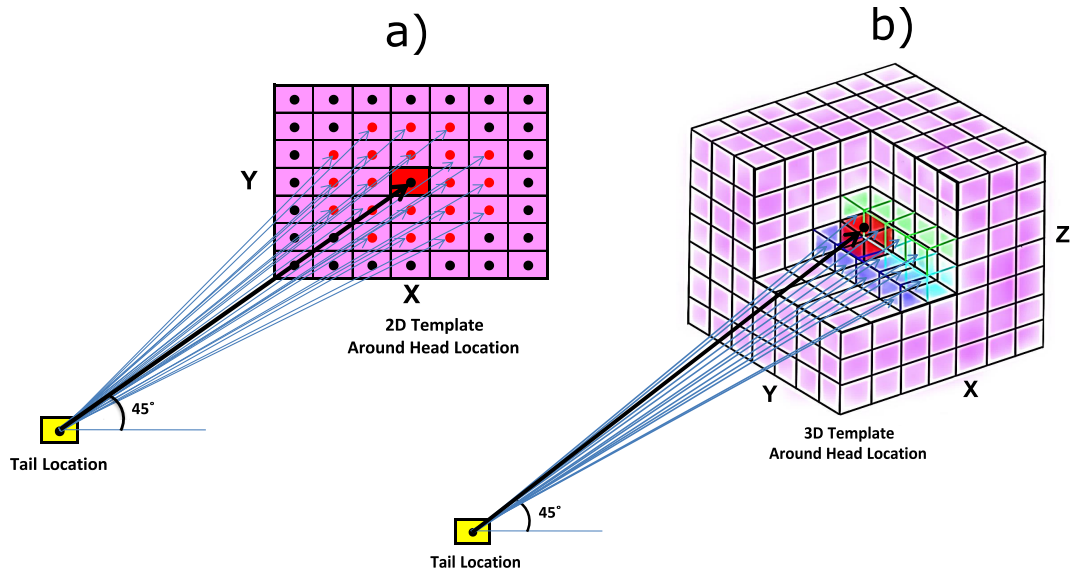

Figure 4.11: a) 2D schematic of using template in variogram calculation for azimuth 45°. b) 3D schematic of using template in variogram calculation for azimuth 45°.

4.3.1 Enforce Upper and Lower Limits of the Seismic-derived Variogram

A rejection sampling is performed to enforce the upper and lower limits of seismic-derived variogram: during simulating variogram realizations by LU simulation with the correlation matrix of the lag distances, those realizations that fail to fall within the upper and lower limits are rejected. This process is repeated until achieving the required number of variogram realizations. The rejection sampling could be applied for all lag distances which leads to the final variogram realizations with lower uncertainty; however, it may not be reasonable to enforce the seismic-derived limits for very low cross covariance (say below 0.2) because it implies no correlation between well data and seismic data. Thus, it is recommended that the rejection sampling is applied up to a minimum correlation, for example $\rho(\mathbf{h}) = 0.2$.

Figure 4.12 shows a synthetic example of applying upper and lower limit seismic-derived variograms on the experimental variogram realizations. Figure 4.12-a shows variogram realizations of the well data with high uncertainty. Figure 4.12-b shows the upper and lower limits of the seismic-derived variogram for correlation coefficient

$\rho(0) = 0.65$. Figure 4.12-c shows the improved variogram realizations by enforcing seismic-derived limits via rejection sampling up to correlation 0.2 ($\rho(\mathbf{h}) = 0.2$) which is lag distance approximately 11.5m (see Figure 4.12-b). The rejection sampling ensures the variogram realizations are within the upper and lower limits. This leads to improve the large variogram uncertainty in Figure 4.12-a (Figure 4.12-c).

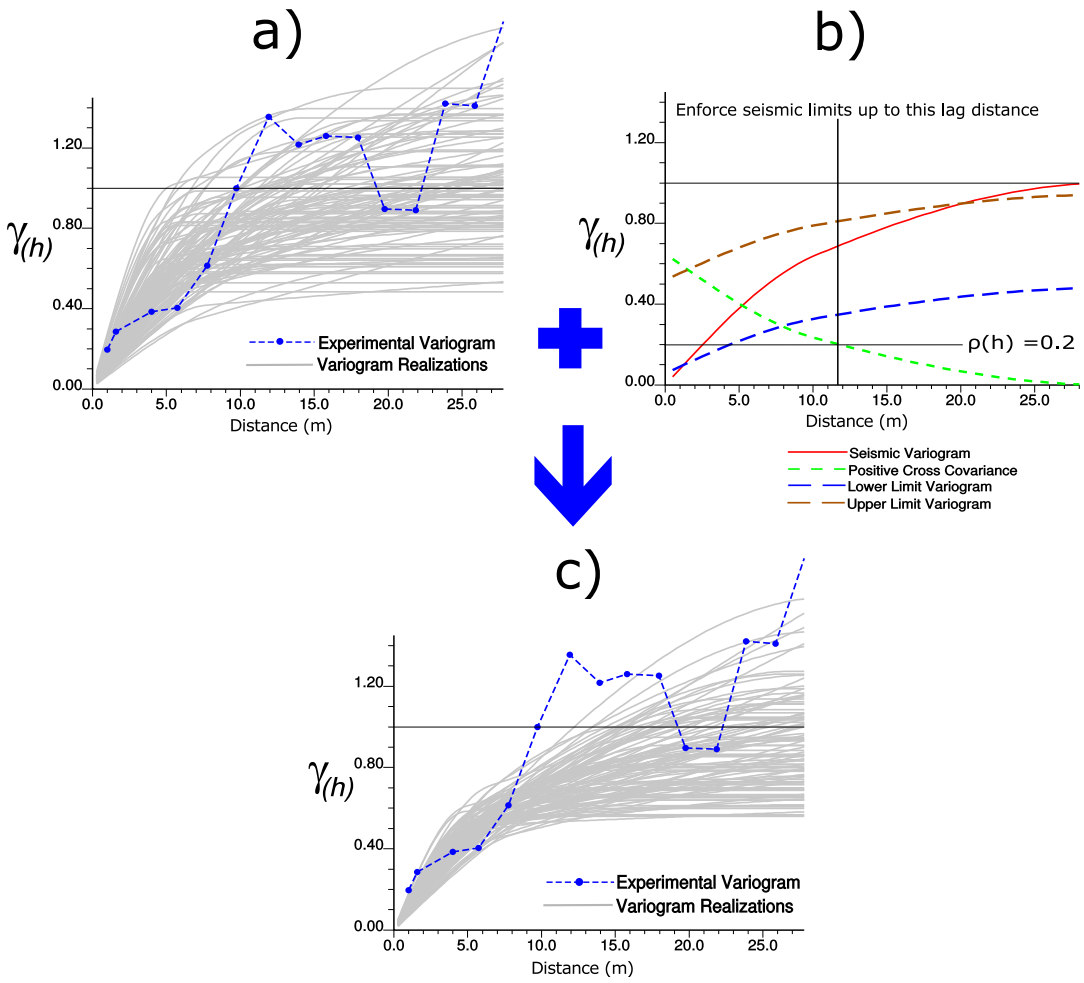

Figure 4.12: Synthetic example of improving well variogram by the seismic-derived variogram (limits). a) High uncertainty in the well variogram. b) Upper and lower limits of the seismic-derived variogram. These limits are enforced on the well variogram realizations (a) up to correlation 0.2 ($\rho(\mathbf{h}) = 0.2$). c) Final variogram realizations by enforcing the seismic-derived variogram.

4.4 Realistic Example: Amoco Case Study

A real case study of the 2D Amoco data (Chu et al., 1994) is considered for improving high variogram uncertainty of well data by seismic data. The variable of interest is averaged porosity (percentage) over the main reservoir layer. This variable is chosen because it has higher correlation with seismic data than permeability. Figure 4.13 shows location map of porosity (a), seismic data (b), the experimental variograms for azimuths 0° and 90° with the fitted models for porosity (c) and seismic data (d). The variograms are calculated after normal score data. There is a zonal anisotropy for azimuth 0°. The variogram realizations based on the methodology mentioned in Chapter 3 by the DoF approach is calculated for azimuths 0° and 90°. Figure 4.14 shows the correlation

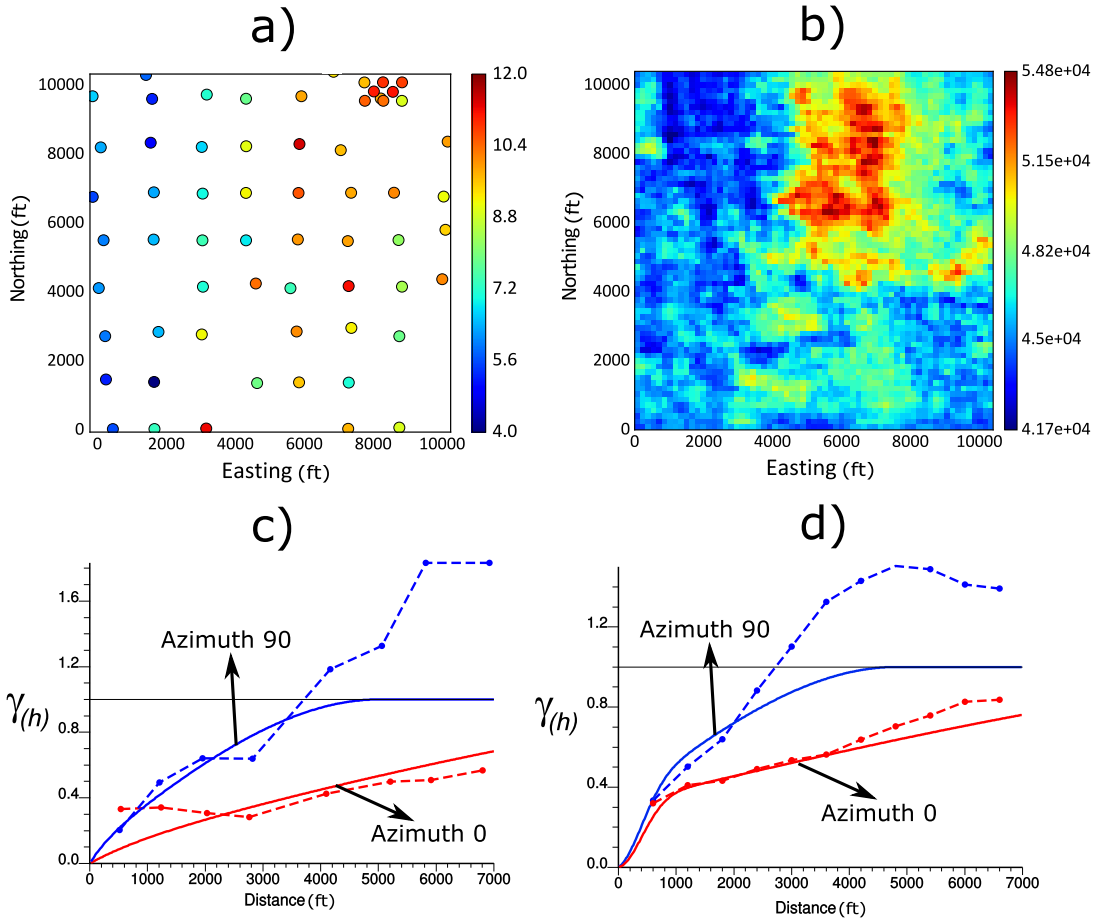

Figure 4.13: Location map of porosity (a), seismic data (b), the experimental variograms for azimuths 0° and 90° with the fitted models for porosity (c) and seismic data (d) for Amoco data set.

matrix between lag distances of two directions. Figure 4.15 shows fitted variogram realizations for azimuths 0° (a) and 90° (b). There is a high variogram uncertainty that could be improved by seismic data if there is a good correlation. Figure 4.16

		Azimuth 0								Azimuth 90							
		Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7	Lag 8	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7	Lag 8
Azimuth 0	Lag 1	1.00	0.19	0.03	0.06	0.04	0.04	0.02	0.02	0.11	0.05	0.02	0.02	0.01	0.02	0.02	0.02
	Lag 2	0.19	1.00	0.72	0.37	0.36	0.30	0.26	0.28	0.03	0.22	0.18	0.14	0.12	0.12	0.11	0.11
	Lag 3	0.03	0.72	1.00	0.79	0.63	0.57	0.53	0.51	0.10	0.30	0.26	0.22	0.20	0.19	0.18	0.16
	Lag 4	0.06	0.37	0.79	1.00	0.67	0.61	0.56	0.50	0.16	0.33	0.28	0.26	0.23	0.21	0.19	0.17
	Lag 5	0.04	0.36	0.63	0.67	1.00	0.89	0.82	0.76	0.12	0.47	0.49	0.47	0.44	0.39	0.34	0.28
	Lag 6	0.04	0.30	0.57	0.61	0.89	1.00	0.95	0.79	0.12	0.37	0.40	0.40	0.37	0.34	0.31	0.26
	Lag 7	0.02	0.26	0.53	0.56	0.82	0.95	1.00	0.86	0.11	0.32	0.36	0.36	0.35	0.32	0.28	0.23
	Lag 8	0.02	0.28	0.51	0.50	0.76	0.79	0.86	1.00	0.08	0.35	0.40	0.40	0.41	0.39	0.34	0.28
Azimuth 90	Lag 1	0.11	0.03	0.10	0.16	0.12	0.12	0.11	0.08	1.00	0.46	0.11	0.11	0.08	0.07	0.06	0.04
	Lag 2	0.05	0.22	0.30	0.33	0.47	0.37	0.32	0.35	0.46	1.00	0.78	0.68	0.59	0.50	0.41	0.30
	Lag 3	0.02	0.18	0.26	0.28	0.49	0.40	0.36	0.40	0.11	0.78	1.00	0.91	0.80	0.69	0.57	0.41
	Lag 4	0.02	0.14	0.22	0.26	0.47	0.40	0.36	0.40	0.11	0.68	0.91	1.00	0.89	0.78	0.63	0.45
	Lag 5	0.01	0.12	0.20	0.23	0.44	0.37	0.35	0.41	0.08	0.59	0.80	0.89	1.00	0.93	0.81	0.63
	Lag 6	0.02	0.12	0.19	0.21	0.39	0.34	0.32	0.39	0.07	0.50	0.69	0.78	0.93	1.00	0.91	0.78
	Lag 7	0.02	0.11	0.18	0.19	0.34	0.31	0.28	0.34	0.06	0.41	0.57	0.63	0.81	0.91	1.00	0.92
	Lag 8	0.02	0.11	0.16	0.17	0.28	0.26	0.23	0.28	0.04	0.30	0.41	0.45	0.63	0.78	0.92	1.00

Figure 4.14: Correlation matrix between eight lag distances of azimuth 0° and eight lag distances of azimuth 90°.

shows cross plot between normal score porosity (y axis) and seismic data (x axis). There is a high correlation. An intrinsic model of cross covariance (Markov model) for azimuths 0° and 90° are calculated for $\rho(0) = 0.65$ and used to attain seismic-derived limits by the mentioned methodology. Figure 4.17 shows the upper and lower limits of seismic-derived variogram for azimuths 0° (a) and 90° (b). The upper and the lower limits of the seismic-derived variogram is calculated based on the assumption the well and the seismic data have the same scale because the well data is averaged for the entire reservoir. Otherwise, the variogram of the seismic data should be downscaled to the

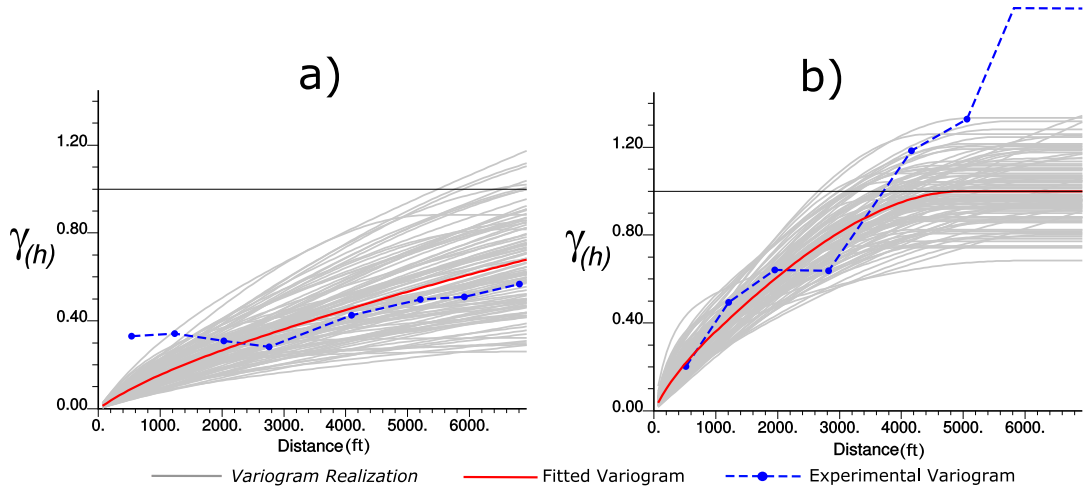

Figure 4.15: Fitted variogram realizations for azimuths 0° (a) and 90° (b).

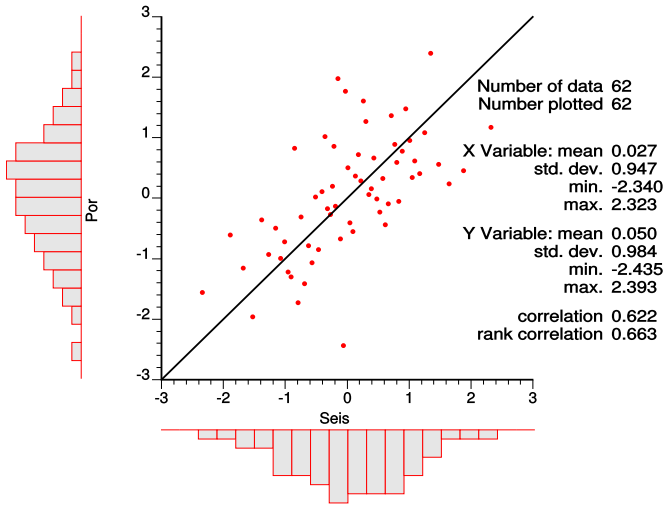

Figure 4.16: Cross plot between normal score porosity (y axis) and seismic data (x axis).

scale of the well data (see Chapter 5 for variogram downscaling). Rejection sampling during LU simulation with the correlation matrix of lag distances (Figure 4.14) is applied to attain the variogram realizations that are within the upper and lower limits up to cross covariance 0.2 ($\rho(\mathbf{h}) = 0.2$). Figure 4.18 shows the improved variogram realizations after applying the upper and lower limits of the seismic-derived variograms for azimuths 0° (a) and 90° (b). There is more improvement for azimuth 0° than 90° because of higher variogram uncertainty for azimuth 0°. The variogram realizations could be standardized to sill 1 and used in geostatistical modeling for incorporating variogram uncertainty.

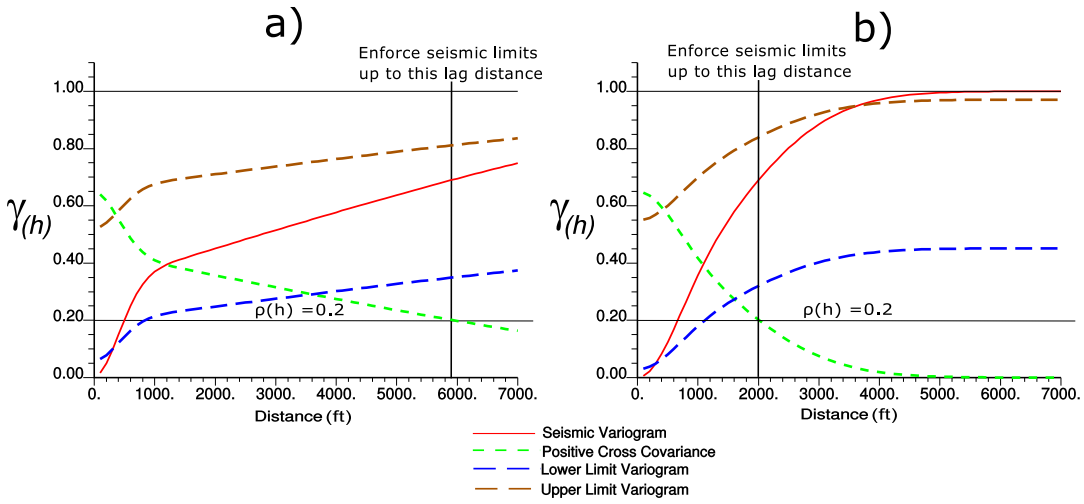

Figure 4.17: The upper and lower limits of seismic-derived variogram for azimuths 0° (a) and 90° (b).

4.5 Remarks

Three different approaches of improving high uncertainty in the well variogram by the seismic variogram is presented. Merging variogram distributions and global cokriging are not efficient. The seismic-derived variogram is an appropriate approach. The summary of each technique is describes.

The variogram distribution of each lag distance for the well and seismic data could be merged to reduce well variogram uncertainty. The curve for defining the variance of the seismic variogram based on the cross covariance could change from one data set to another. The most accurate curve could be achieved for each data set. However, in practice, it is a time consuming process especially in case of 3D data set. Thus, it is not an efficient approach for reducing the uncertainty of the well variogram by the seismic variogram.

The variogram pairs of the exhaustive seismic data could be used to estimate the sparse variogram pairs of the well data by global cokriging approach. The standardized form of simple global cokriging (SGCK) is proposed. Although SGCK removes noise and artifacts in the experimental variogram, the updated variogram converges to the fitted variogram model and not the real variogram according. Therefore, global cokriging approach is also considered inappropriate to improve the variogram of the well

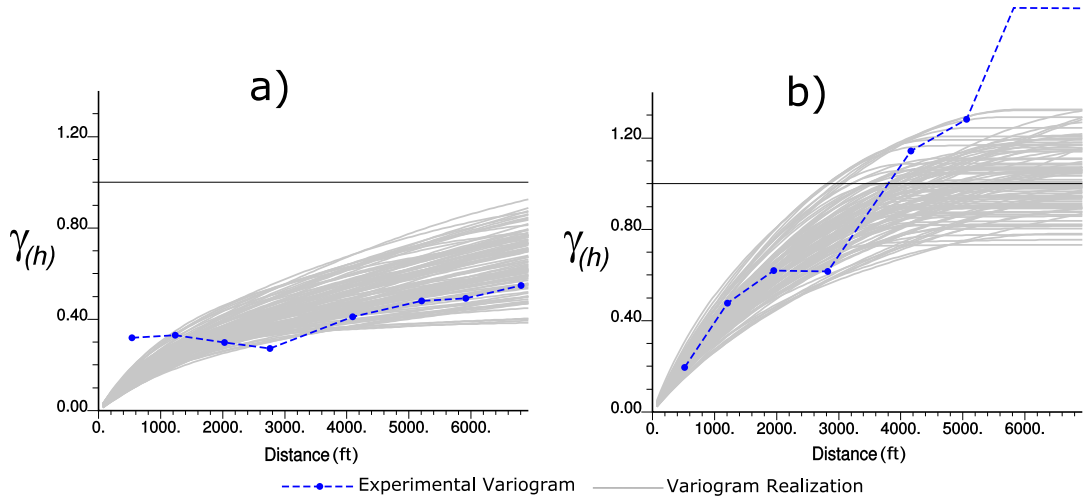

Figure 4.18: Improved variogram realizations after applying the limits of seismic-derived variogram for azimuths 0° (a) and 90° (b).

data by the seismic variogram.

A totally positive covariance matrix between well and seismic data for a variogram pair at each lag distance is proposed to attain the acceptable range of the unknown covariance of the well data. This process is repeated for all lag distances. This leads to the upper and lower limits of the seismic-derived variogram. These limits could be applied on the well variogram uncertainty by a rejection sampling to ensure variogram realizations of the well data fall within the upper and lower limits. The seismic-derived variogram is so efficient and straightforward to use; it does not has the limitations of the approaches of merging variogram distributions and global cokriging and it is computationally so fast. The limits are recommended to apply for a reasonable cross covariance (say above 0.2) to achieve more reliable seismic-derived variogram. The higher the correlation between well and seismic data, the lower the difference between the upper and lower limits. The variogram of the seismic data do not provide any constraint on well variogram if the correlation is very low.

Only 2D data sets are considered to apply the seismic-derived variogram in this Chapter. In case of 3D data set, the seismic-derived variogram is applied on the horizontal variogram of the well data after downscaling the seismic variogram (see Chapter 5).

Chapter 5

Variogram Realizations

This Chapter presents a methodology to improve the horizontal variogram of well data by considering the vertical variogram of well data as well as the horizontal variogram of seismic data because they are often well-defined. Variogram downscaling is also developed because the horizontal variogram of seismic data is at a larger scale than the well data. The vertical variogram is scaled to scenarios of the horizontal variogram and merged with the experimental horizontal variogram and the seismic-derived variogram. Synthetic and realistic examples are presented.

5.1 Methodology

An approach is developed to improve the prediction of the horizontal variogram of properties from well data. The challenge is the horizontal variogram of widely spaced well data; the vertical variogram is often well-defined. The main aim of this Chapter is improving the horizontal variogram uncertainty by the horizontal variogram from seismic data and the vertical variogram from well data.

Figure 5.1 shows a schematic illustrating the methodology for improving a noisy horizontal variogram with all available data (seismic and vertical variogram). The probability distributions characterizing the uncertainty in the individual lag distances for the noisy horizontal variogram along the stratigraphic direction γ_w must be defined (see Figure 5.1-a). In case of preferential sampling, variogram declustering is required to remove artifacts in the experimental variogram. Ordinary global kriging was proposed

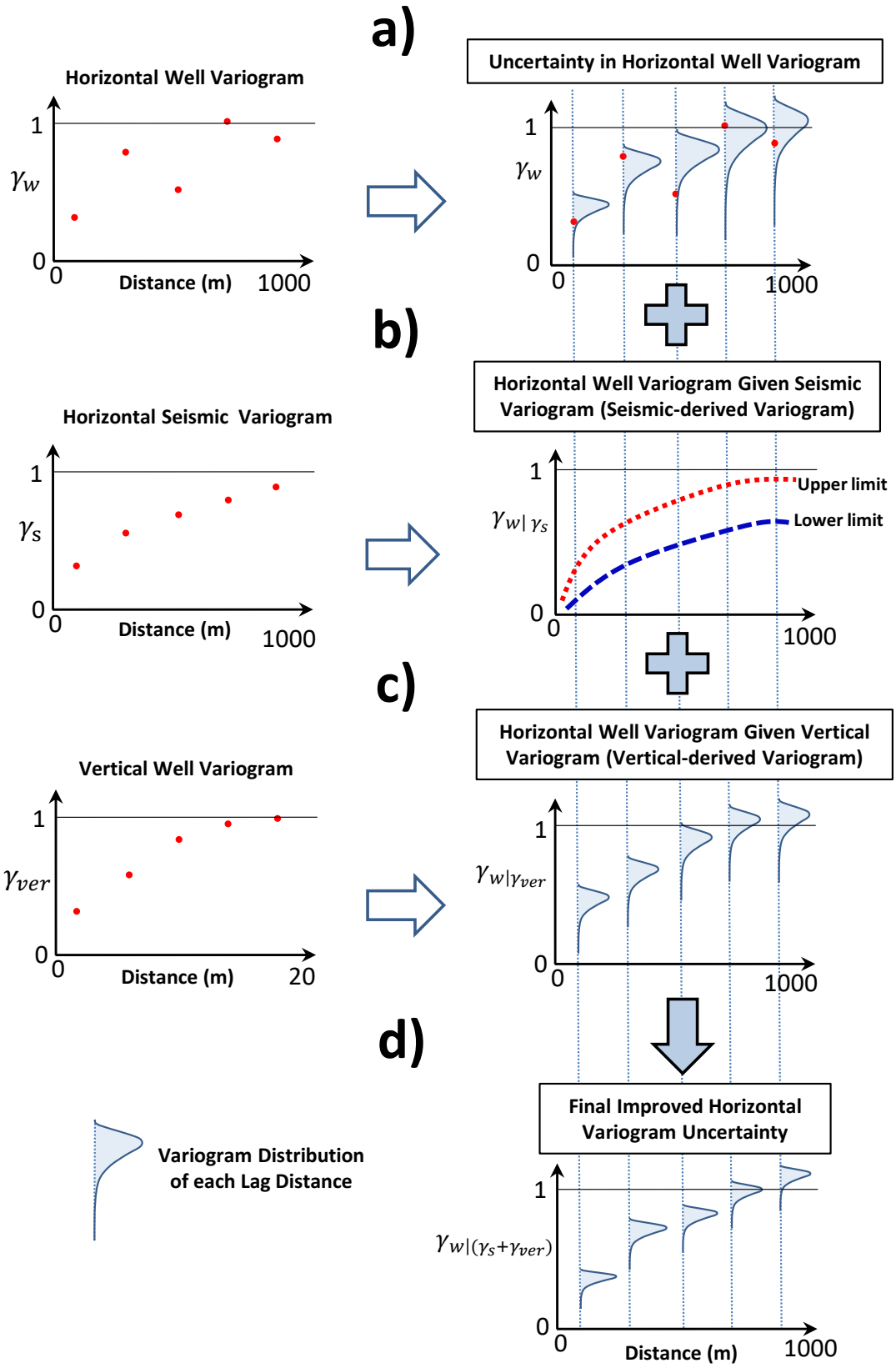

Figure 5.1: Schematic of improving the horizontal variogram of well data (a) by the horizontal variogram of seismic data (b) and the vertical variogram of well data (c).

for declustering the sample variogram in Chapter 2. The variogram model fitted to the declustered sample variogram (horizontal direction) is used to calculate variogram uncertainty. The DoF approach proposed in Chapter 3 is used for calculating variogram uncertainty. The high variogram uncertainty computed by the DoF approach (Figure 5.1-a) is reduced by the horizontal variogram of seismic data (Figure 5.1-b) and the vertical variogram of well data (Figure 5.1-c).

Seismic data are extensively sampled in the horizontal direction although the vertical variogram of seismic data is not particularly helpful due to the resolution of surface seismic. The horizontal variogram of seismic data γ_s and the cross covariance are used to calculate the upper and lower limits of the seismic-derived variogram for horizontal directions by enforcing the covariance matrix between well and seismic to be totally positive. These limits can be applied on the horizontal variogram uncertainty via rejection sampling. This was discussed in Chapter 4, Section 4.3.1. The horizontal variogram of seismic data should be down scaled to the scale of the well data to remove the inconsistency between the scale of the horizontal well and seismic variograms. Figure 5.1-b shows the upper and lower limits of the seismic-derived variogram for horizontal directions after downscaling the horizontal variogram of seismic data.

The vertical variogram of well data γ_{ver} could be scaled to scenarios of the horizontal variogram (see Figure 5.1-c), and used as further data for improving the horizontal variogram uncertainty. This can be done by considering uncertainty in horizontal to vertical anisotropy ratios (H:V) (Kupfersberger & Deutsch, 1999), and uncertainty in the sill (see Section 5.3). The vertical-derived variogram provides a variogram distribution for each lag distance. If seismic data is not available, the variogram distribution of the horizontal variogram (Figure 5.1-a) and the vertical-derived variogram (Figure 5.1-b) could be merged for each lag distance to achieve the final variogram (see Section 5.4). The method of combining error ellipses discussed in Chapter 4, Section 4.1.1 is proposed for merging variogram distributions. If seismic data is available, the limits of the seismic-derived variogram should also be taken into consideration to attain the final improved horizontal variogram uncertainty (Figure 5.1-d).

5.2 Variogram Downscaling

The proposed methodology for improving horizontal well variogram with seismic variogram is applicable when well and seismic data are at the same scale. The scale of seismic data is always bigger than the scale of well data. The horizontal scale of seismic data is the bin size of seismic acquisition, which is usually 20m. The horizontal scale of well data is the diameter of core sample or horizontal influence of well logging tools. The vertical scale of seismic data is approximately 10m versus the vertical scale of well data based on well logging vertical resolution is usually 0.15m. The difference between the scale of well and seismic data leads to inconsistency between the horizontal variograms: the horizontal variogram of seismic data has lower variance contributions and larger ranges than the horizontal variogram of the well data. The difference between the vertical scales of well and seismic data is the most important. This is due to the greater variability in the vertical direction. This inconsistency could be fixed by downscaling the horizontal variogram of seismic data to the scale of well data. The variogram downscaling might not be required for 2D data because well and seismic data are both vertically averaged for the entire reservoir (Chapter 4). There are some publications on variogram downscaling (Babak & Leuangthong, 2008; Frykman & Deutsch, 2002; Kupfersberger, Deutsch, & Journel, 1998). The small scale variogram $\gamma_v(\mathbf{h})$ and the large scale variogram $\gamma_V(\mathbf{h})$ are written as:

$$\gamma_v(\mathbf{h}) = C_v^0 + \sum_{i=1}^{n} C_v^i \gamma_v^i(\mathbf{h}) \quad \textit{and} \quad \gamma_V(\mathbf{h}) = C_V^0 + \sum_{i=1}^{n} C_V^i \gamma_V^i(\mathbf{h}) \tag{5.1}$$

where v is small scale, V is the large scale, C^0 is nugget effect, C^i and $\gamma^i(\mathbf{h})$ are variance contributions and variogram models for variogram structures $i = 1, ..., n$. Variogram downscaling is established under the assumption that the variables average linearly and the variogram shapes are unchanged after downscaling (Kupfersberger et al., 1998). The scaling law for variogram upscaling/downscaling retrieved from Babak and Leuangthong (2008); Frykman and Deutsch (2002); Kupfersberger et al. (1998) is written as:

1. The range of variogram increases when the sampling volume size goes up. This increase depends on the volume difference ($|V|-|v|$). $|V|$ and $|v|$ are volume sizes for large scale and small scale. For example, if V is a large scale $100 \times 100 \times 1$ m^3, then $|V|$ is between $1m$ and $141.42m$. The scaling law for variogram range is written as:

$$a_v^i = a_V^i - (|V| - |v|) \tag{5.2}$$

where a_v^i and a_V^i are the small and the large scale variogram ranges for variogram models $\gamma_v^i(\mathbf{h})$ and $\gamma_V^i(\mathbf{h})$, $i = 1, ..., n$.

2. The scaling law for nugget effect C^0, and variance contribution of each nested structure C^i is written as:

$$C_v^0 = C_V^0 \frac{|V|}{|v|} \quad , \quad C_v^i = C_V^i \frac{1 - \overline{\gamma_{(v,v)}^i}}{1 - \overline{\gamma_{(V,V)}^i}} \tag{5.3}$$

where $\overline{\gamma_{(v,v)}^i}$ and $\overline{\gamma_{(V,V)}^i}$ are the average variograms or the *gamma-bar* (Kupfersberger et al., 1998) for each nested structure i. As the volume size increases, the nugget effect and variance contributions of the large scale variogram decrease because by increasing the volume, the *gamma-bar* goes up. The impact of the vertical scale on variogram downscaling in reservoir is much higher than the horizontal scale due to the fact that the length or the width of reservoir is always higher than reservoir thickness and higher horizontal variogram range than the vertical variogram range. This leads to approximately equal *gamma-bar* for the small and large horizontal scales ($\overline{\gamma_{(v,v)}^i}$, $\overline{\gamma_{(V,V)}^i}$) if the vertical scale is constant.

The small scale variogram $\gamma_v(\mathbf{h})$ and large scale variogram $\gamma_V(\mathbf{h})$ are required for variogram downscaling. $\gamma_V(\mathbf{h})$ is the variogram of the original seismic data at large scale. The small scale variogram of seismic data $\gamma_v(\mathbf{h})$ could be assumed as the variogram of the directly calculated acoustic impedance or synthetic seismogram from well data; acoustic impedance at well scale can be calculated by density log ρ and velocity log ν ($\rho.\nu$); synthetic seismogram can be built by calculating the reflection coefficients $R = \frac{\rho_2.\nu_2 - \rho_1.\nu_1}{\rho_2.\nu_2 + \rho_1.\nu_1}$ and convolve with a wavelet. The downscaled horizontal

seismic variogram should be standardized to sill 1 before calculating the seismic-derived variogram.

The horizontal cross covariance between well and seismic data at the scale of well data is required to calculate the seismic-derived variogram (Chapter 4, Section 4.3). Down scaling the horizontal cross covariance may not be possible since the small scale cross covariance between well and the calculated acoustic impedance leads to artificially high cross covariance due to the fact that the acoustic impedance is achieved from well data not the real seismic data. Thus, an intrinsic cross covariance model (Markov model) (Pyrcz & Deutsch, 2014) is proposed: the covariance model of the normal score seismic data is scaled by the correlation coefficient between the normal score well and seismic data. The correlation coefficient between well and seismic data at the scale of well data is required for this approach: as the scale goes down, the correlation between well and seismic decreases. The correlation coefficient at the small scale could be calculated by fitting an equation to the different upscaled volumes (x axis) and the observed correlations from the data set (y axis). The resulting equation may give an approximate correlation between well and seismic data at scales less than the vertical resolution of seismic data (see Chapter 7).

5.3 The Vertical-derived Variogram $\gamma_{w|\gamma_{ver}}$

The vertical variogram of well data γ_{ver} could be scaled to scenarios of the horizontal variogram. The vertical variogram is often well-defined due to regular sampling in direction of drilling, and it can be scaled to the horizontal variogram by the horizontal-to-vertical anisotropy ratio (H:V) (Kupfersberger & Deutsch, 1999) based on the geological conceptual model or analogue information. The probability distribution of each lag distance can be achieved by considering uncertainty in H:V and uncertainty in the sill of the vertical-derived variogram. A triangular distribution is assumed to represent the uncertainty in H:V. Minimum (Min), Mode and Maximum (Max) values are required. The Gaussian distribution is assumed to represent the uncertainty in the sill of the vertical-derived variogram. The Gaussian distribution is constructed by mean=1,

which is the sill value for the standardized data, and standard deviation σ:

$$\Delta_v = Max((1-Z),(T-1)) \quad , \quad \sigma = \frac{\Delta_v}{P_{90}} \tag{5.4}$$

where Z is zonal anisotropy, and T is trend for the horizontal variogram: Δ_v is the practical maximum deviation of the sill of the horizontal variogram from the sill 1. There may be a zonal anisotropy or trend in the horizontal variogram that can be inferred from the conceptual geological model or seismic data. Accounting for uncertainty in the sill avoids unreasonably low variogram uncertainty in the vertical-derived variogram. The standard deviation of the distribution of the sill σ can be calculated for P_{90} (or P_{10}) of the normal standard distribution. Figure 5.2-a shows the triangular distribution for H:V, and Figure 5.2-b shows the Gaussian distribution of the sill. Simulating many H:V and sill values from the corresponding triangular and Gaussian distributions provides the vertical-derived variogram. This process is repeated to calculate variogram realizations of the vertical-derived variogram.

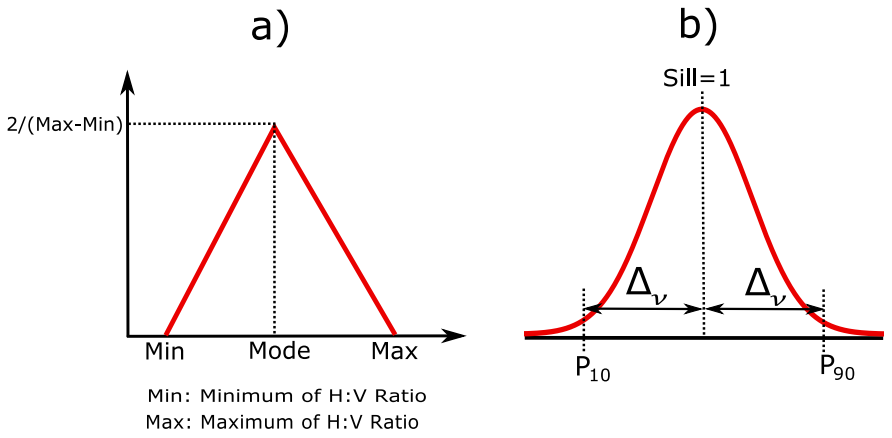

Figure 5.2: a) The triangular distribution of H:V. b) The Gaussian distribution of the sill.

Figure 5.3 shows a synthetic example of scaling the vertical variogram to scenarios of the horizontal variogram by the mentioned approach. Min, $Mode$ and Max of H:V for building the triangular distribution are 80, 100 and 120 respectively. Figure 5.3-a shows the triangular distribution of the H:V. Figure 5.3-b shows the Gaussian distribution of the sill for $\Delta_v = 0.05$. Figure 5.3-c shows a synthetic vertical variogram

with two spherical structures, variance contributions of 0.4 and 0.6, and ranges of 2m and 10m, respectively. Figure 5.3-d shows 100 realizations of the vertical-derived variogram. If the horizontal variogram cannot be calculated because of the limited well data (less than 4 wells), the vertical-derived variogram could be replaced with the horizontal variogram and used in geostatistical modeling.

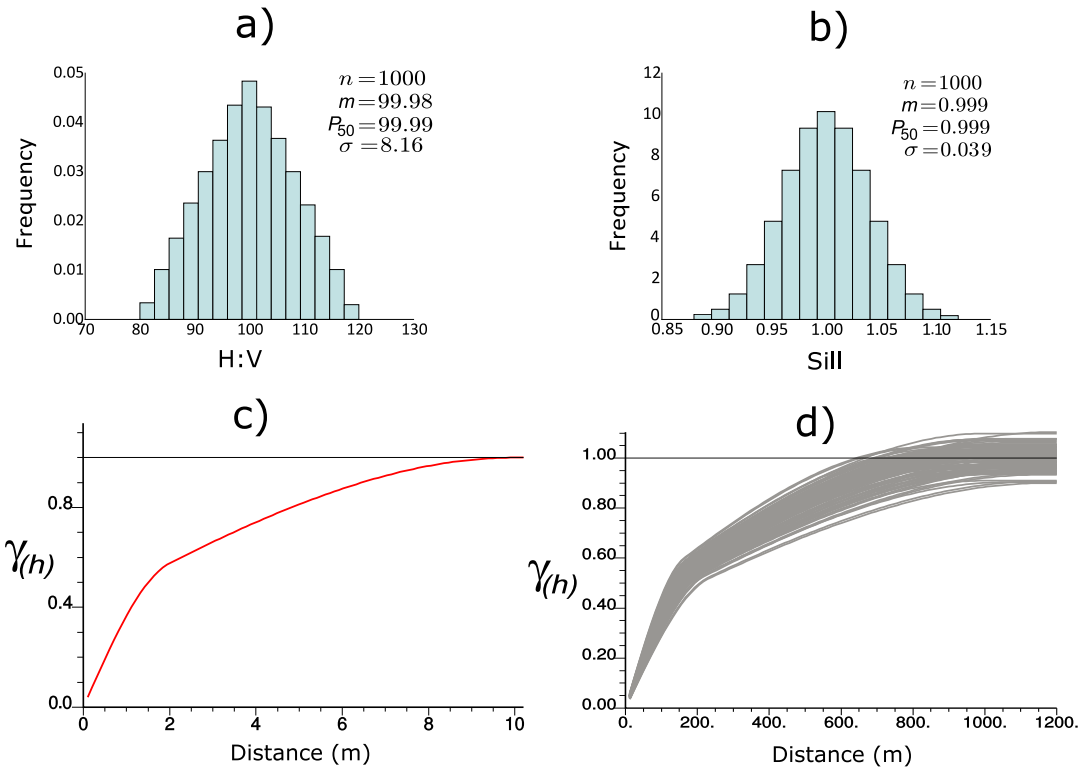

Figure 5.3: a) The triangular distribution of H:V with $Min = 80$, $Mode = 100$, $Max = 120$. b) The Gaussian distribution of the sill for $\Delta_v = 0.05$. c) Synthetic vertical variogram. d) 100 variogram realizations of the vertical-derived variogram.

5.4 Merge the Horizontal and the Vertical-Derived Variograms

The distributions of the horizontal variogram γ_w and the vertical-derived variogram $\gamma_{w|\gamma_{ver}}$ are merged to achieve an updated and more accurate variogram distribution for each lag distance. The error ellipses approach from Chapter 4, Section 4.1.1 is used for merging the variogram distributions. A transformation table is required to transform

the Chi-square distribution of each lag distance to a Gaussian distribution, and back transform the results to the Chi-square distribution. This transformation could be done by quantile matching lookup analogous to the normal score transformation approach (Deutsch & Journel, 1998). That is:

$$y = G^{-1}(F(z)) \quad , \quad z = F^{-1}(G(y)) \tag{5.5}$$

where y is the Gaussian transformed random variable corresponding to the original data z. $F(z)$ is cumulative distribution function (CDF) of original data, $G(y)$ is the standard CDF. An anamorphosis function is constructed from Equation 5.5 based on matching quantiles between a normal score distribution function $G(y)$ and the standard Chi-square distribution function with 1 degree of freedom $F(z)$. This anamorphosis function is denoted:

$$G(y) = \varphi(Chi(z)) \quad , \quad Chi(z) = \varphi^{-1}(G(y)) \tag{5.6}$$

where φ is the anamorphosis function that transform any Chi-square distribution function $Chi(z)$ to the Gaussian distribution function $G(y)$. The merged Gaussian distribution function can be back transformed to the merged Chi-square distribution by inverse of the anamorphosis function φ^{-1}. Figure 5.4 shows that the Chi-square distributions 1, 2 are transformed to the Gaussian distributions, merged in Gaussian space and back transformed to the Chi-square distributions (distribution 3) by the anamorphosis function.

Figure 5.5 shows a synthetic example of merging the Chi-square distributions assumed as the variogram distributions of γ_w (Distribution 1) and $\gamma_{w|\gamma_{ver}}$ (Distribution 2) for a lag distance by the mentioned methodology. Figure 5.5-a shows the Chi-square distributions: the left hand side with the mean of 0.2 and standard deviation of 0.141, and the right hand side with the mean of 0.9 and standard deviation of 0.635. Each Chi-square distribution should be transformed to the Gaussian distribution by the anamorphosis function (see Equation 5.6 and Figure 5.4). Figure 5.5-b shows the

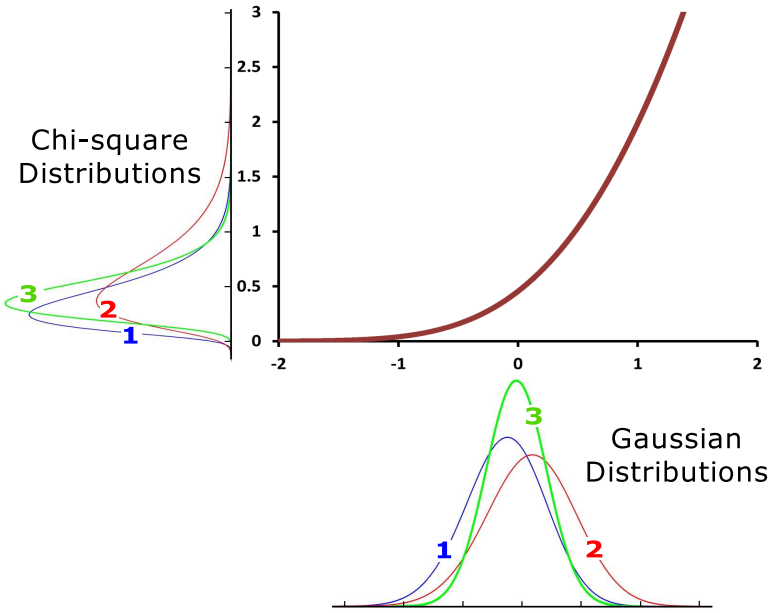

Figure 5.4: The Chi-square distributions 1, 2 are transformed to the Gaussian distributions by the anamorphosis function, merged in the Gaussian space (distribution 3) by error ellipse approach and back transformed to the Chi-square distribution by inverse of the anamorphosis function φ^{-1}. The vertical axis is the standard Chi-square distribution function and the horizontal axis is the standard normal distribution function.

Gaussian transformed distributions of each Chi-square distribution (Figure 5.5-a). Figure 5.5-c shows the merged distribution in Gaussian space in left, and the Chi-square distribution in right after back transformation the merged Gaussian distribution by inverse of the anamorphosis function φ^{-1}. The merged Chi-square distribution is convex: the updated mean is 0.317 which is in the middle of the mean of two Chi-square distributions (0.2, 0.9) and it is close to the distribution with lower uncertainty (Distribution 1). The process of merging variogram distributions of γ_w and $\gamma_{w|\gamma_{ver}}$ should be repeated to achieve the updated variogram distributions for all lag distances.

5.4.1 Variogram Realizations after Merging Variogram Distributions

After merging variograms γ_w and $\gamma_{w|\gamma_{ver}}$, variogram realizations could be drawn from the updated (merged) Chi-square distributions of the variogram lags. Variogram realizations should respect the correlation between lags by LU simulation with the corre-

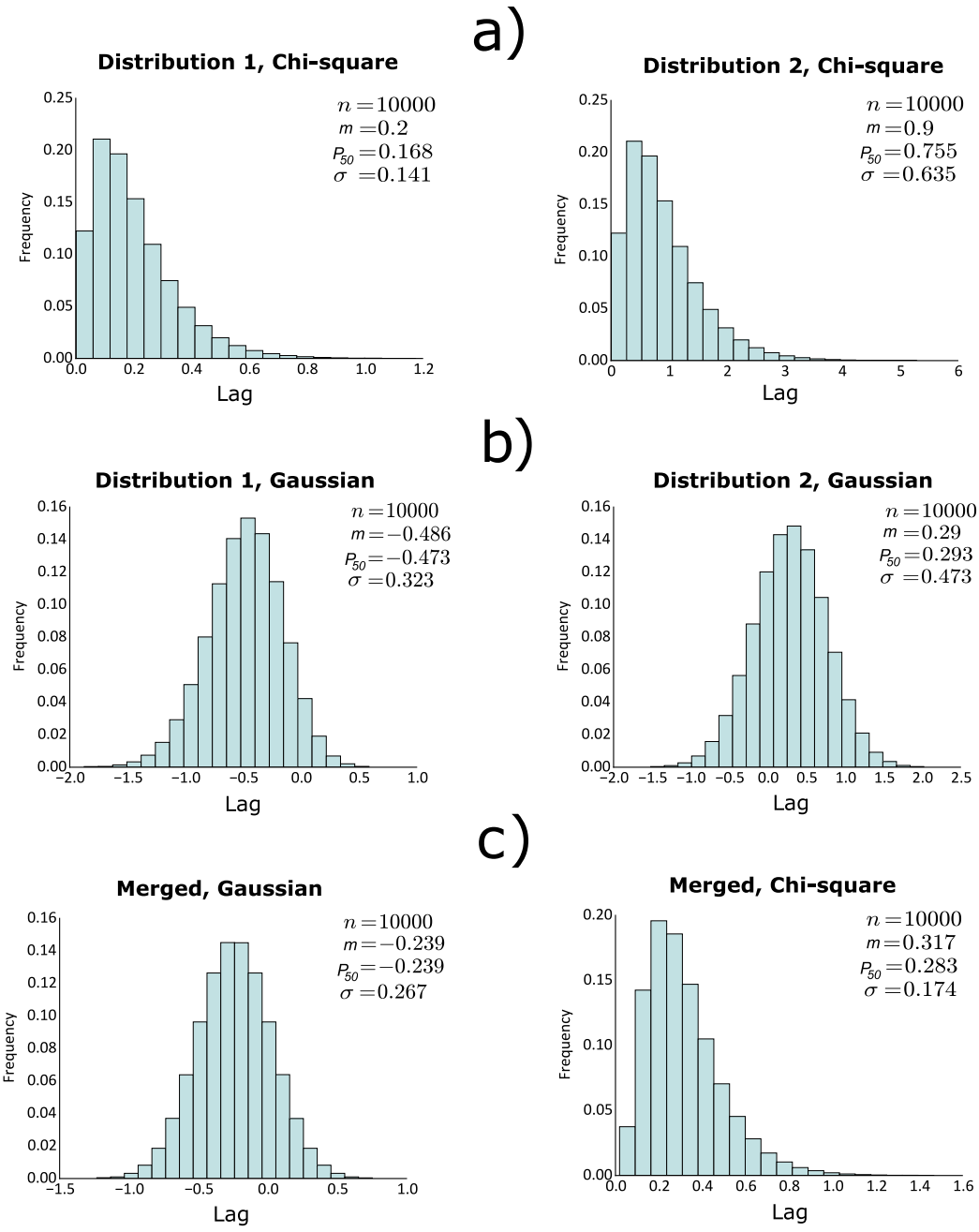

Figure 5.5: a) Two Chi-square distributions assumed to be the distributions of the same lag distance for different variograms (γ_w, $\gamma_{w|\gamma_{ver}}$). b) The Gaussian transformed distribution of each Chi-square distribution. c) Merged distributions in Gaussian space in left Figure and Chi-square space in right Figure.

lation matrix of the variogram lag distances. LU simulation for calculating correlated variogram realizations was discussed in Chapter 3, Section 3.4.

Figure 5.6 shows a synthetic example of merging the uncertainty in the horizon-

tal variogram and the vertical-derived variogram, and sample realizations from the

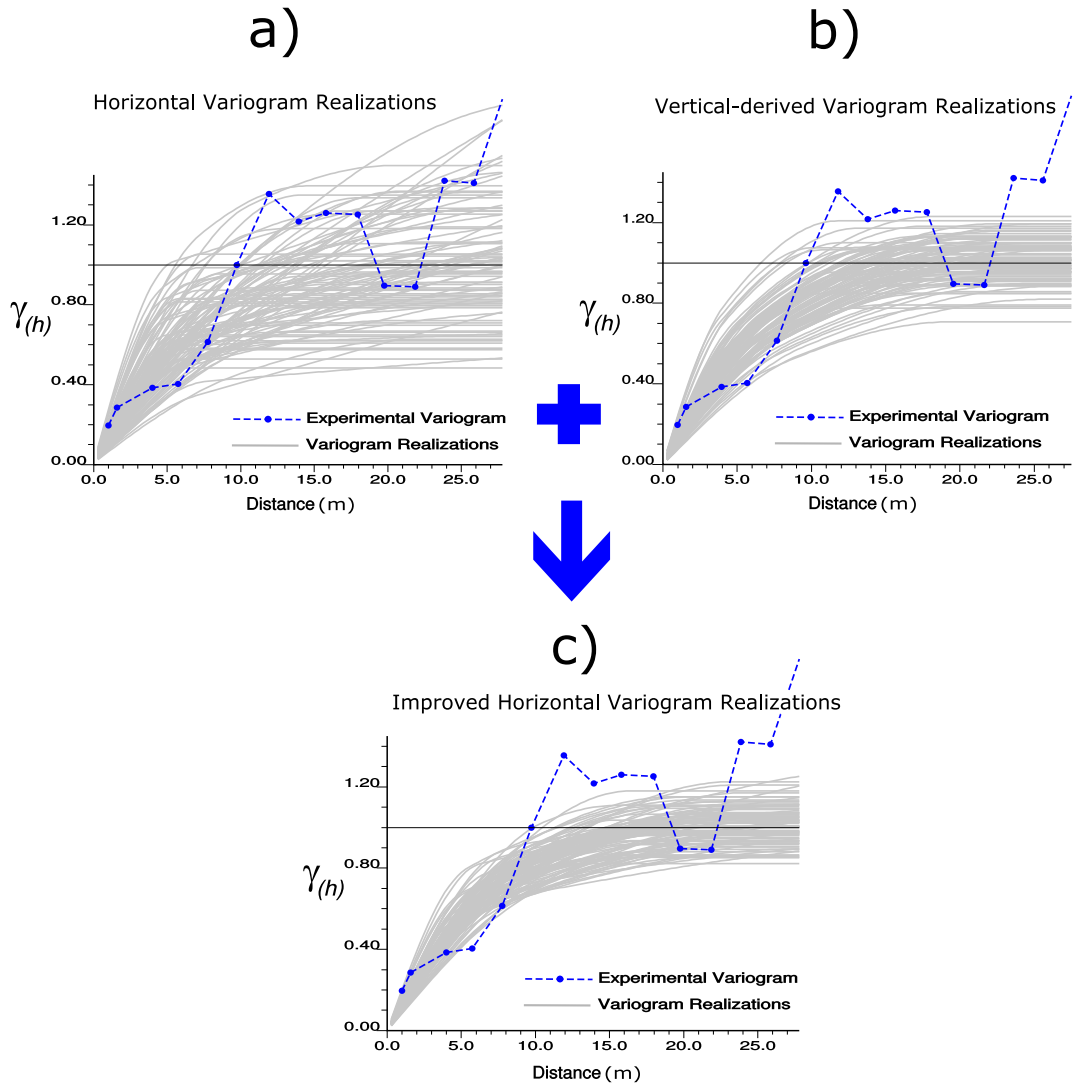

Figure 5.6: Synthetic example of merging the horizontal variogram realizations (a) and the vertical-derived variogram realizations (b). The improved horizontal variogram (c) has lower uncertainty.

merged distributions. The experimental horizontal variogram is shown on each variogram (dashed pointed line). Figure 5.6-a shows the horizontal variogram realizations with high uncertainty. Figure 5.6-b is the vertical-derived variogram realizations. The distribution of each lag distance of the horizontal variogram and the vertical-derived variogram are merged with the error ellipse approach after transforming from the Chi-square distribution to the Gaussian distribution, then back transformed to the Chi-square distribution after merging (see the mentioned methodology above). Variogram

realizations are drawn from the updated distributions of each lag distances. Figure 5.6-c shows the improved horizontal variogram realizations with lower uncertainty. If exhaustive seismic data is not available, merging the horizontal variogram realizations and the vertical-derived variogram is applied.

5.5 Improve Horizontal Variogram by the Vertical-derived and Seismic-derived Variograms

The challenge addressed now is when seismic data is also available with a vertical-derived horizontal variogram. This could be done by enforcing the seismic-derived variogram on both γ_w and $\gamma_{w|\gamma_{ver}}$ and then variogram distributions are merged. A simulation study is presented to confirm the efficiency of this approach against other possible techniques of applying the seismic-derived variogram such as enforcing the seismic-derived variogram on only 1- γ_w, 2- $\gamma_{w|\gamma_{ver}}$, 3- Merged variogram. Synthetic well and seismic data with a high correlation are simulated. Sparse wells are drawn. The horizontal variograms from the well and seismic data and the vertical variogram from well data are calculated. The horizontal variogram uncertainty, the vertical-derived and the seismic-derived variograms are computed. All possible approaches of using the seismic-derived variogram are applied to improve the horizontal variogram uncertainty. The approach that makes the horizontal variogram closest to the real horizontal variogram of well data will be recommended.

Synthetic Gaussian well data are simulated by sequential Gaussian simulation (SGS) with grid numbers of $nx = 100 \times ny = 100 \times nz = 50$. The horizontal grid size (x and y) is 10m and the vertical grid size is 1m; the area is $x = 1000m \times y = 1000m \times z = 50m$. The reference variogram model has two spherical structures with variance contributions of 0.5 (no nugget effect). The horizontal and vertical variogram ranges for the first structure are 140m and 6m, and the horizontal and the vertical variogram ranges for the second structure are 380m and 12m. Well data are drawn from the simulated data set. Figure 5.7 shows the simulated well data and the locations of the ten drawn wells.

Synthetic Gaussian seismic data (acoustic impedance) are simulated with the same

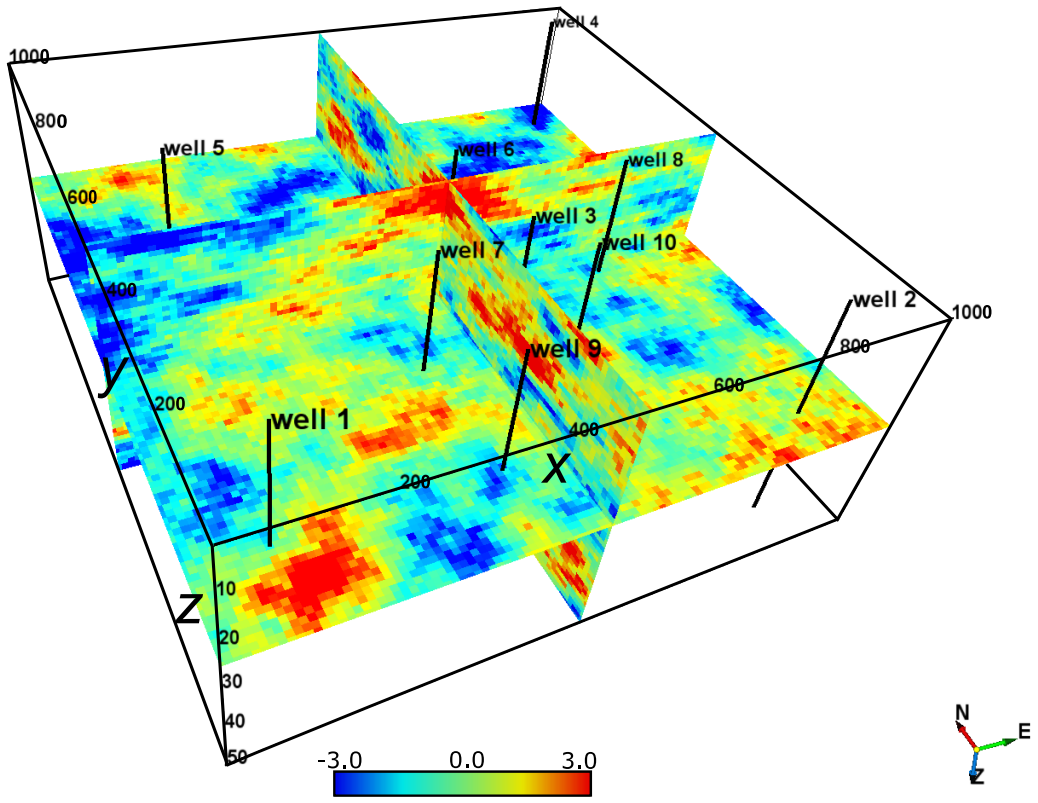

Figure 5.7: Simulated Gaussian well data set with grid numbers of $nx = 100 \times ny = 100 \times nz = 50$ (volume of $x = 1000m \times y = 1000m \times z = 50m$) and locations of the ten drawn wells.

grid numbers and grid sizes of well data enforcing a high negative correlation ($\rho(0) = -0.65$) because in reality, well data (porosity) usually have a negative correlation with the acoustic impedance. The variogram of the seismic data is isotropic with two spherical structures and variance contributions of 0.4 and 0.6 for the first and second structures (no nugget effect). The horizontal and vertical variogram ranges for the first structure are 160m and 6m, and the horizontal and vertical variogram ranges for the second structure are 350m and 12m. It is assumed that well data and seismic data have the same scale. Figure 5.8 shows that the simulated seismic data. Figure 5.9 shows the cross plot between the well and seismic data at ten drawn well locations ($\rho(0) \approx -0.65$). Figure 5.10 shows the horizontal variogram (Figure 5.10-a) for azimuth $0^\circ \pm 30^\circ$ and the vertical variogram (Figure 5.10-b) of the ten drawn wells and fitted variogram models. The horizontal variogram is noisy because of sparse well data (see Figure 5.7); however, the vertical variogram is well-defined. For calculating the horizontal

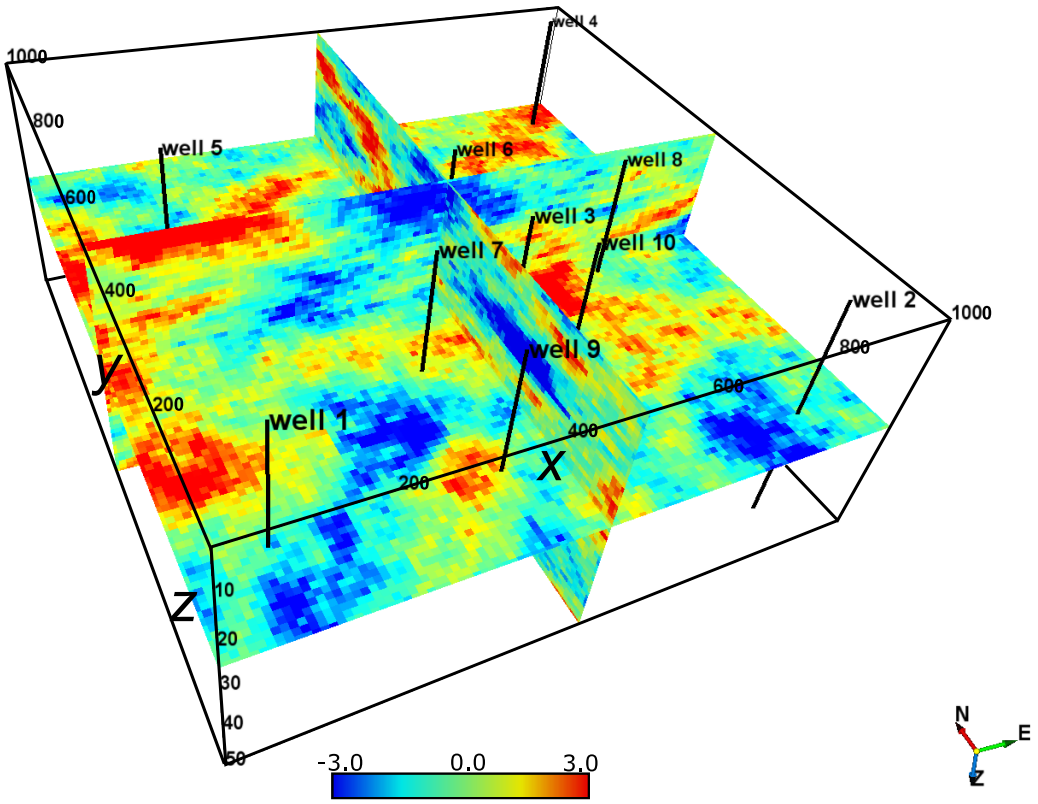

Figure 5.8: Simulated seismic data set having high correlation with the Figure 5.7. Grid numbers are $nx = 100 \times ny = 100 \times nz = 50$ and the volume is $x = 1000m \times y = 1000m \times z = 50m$.

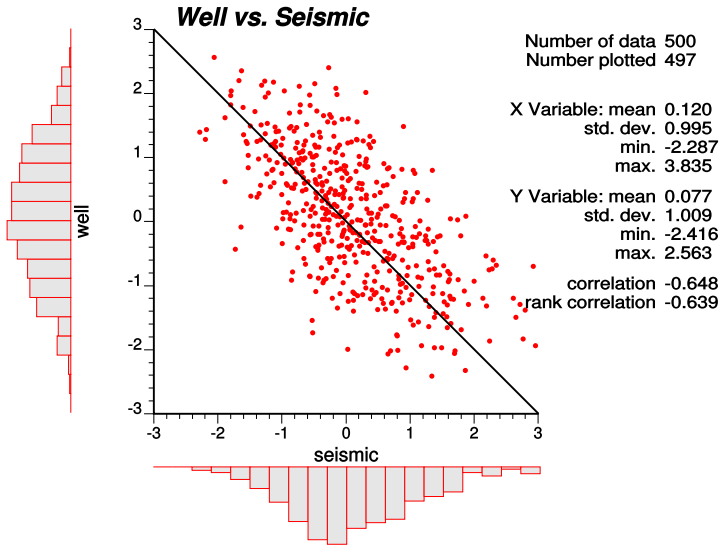

Figure 5.9: Cross plot between well and seismic data for the ten drawn wells.

variogram realizations, the correlation matrix between lag distances are calculated in Figure 5.11-a, and the distribution of uncertainty of each lag distance is calculated by

the DoF approach (Chapter 3). Since the wells are sparse and the variogram range is not large, there is low correlation between lag distances. Variogram realizations are drawn from the uncertainty interval of lag distances: LU simulation is applied with

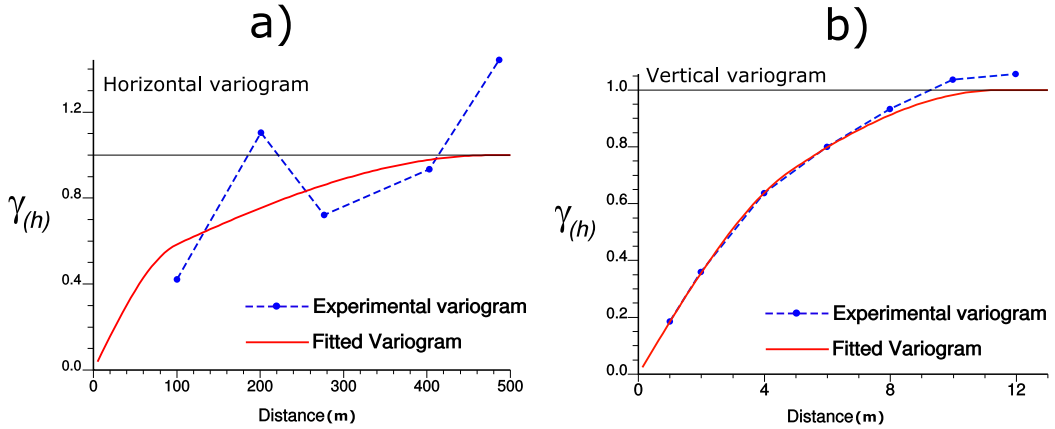

Figure 5.10: The horizontal variogram for azimuth $0° \pm 30°$ (a) and the vertical variogram of the ten wells drawn from the Figure 5.7 (b).

the correlation matrix between lag distances and back transformed to the marginal Chi-square distributions. Figure 5.11-b shows the correlation between lag distances for 100 variogram realizations after LU simulation. There is a slight difference between the correlation matrix (Figure 5.11-a) and correlations between variogram realizations (Figure 5.11-b). Figure 5.12-a shows the horizontal variogram realizations and Figure 5.12-b shows the fitted variogram realizations. The uncertainty is high and can be improved by the vertical-derived variogram and the upper and lower limits of the horizontal seismic-derived variogram. Figure 5.13 shows the vertical-derived variogram $\gamma_{w|\gamma_{ver}}$; uncertainty in H:V is calculated by the triangular distribution with $Min = 15$, $Mode = 21.5$ and $Max = 28$, and $\Delta_v = 0.1$ for calculating the uncertainty in the sill (see Figure 5.2). $\gamma_{w|\gamma_{ver}}$ is achieved by scaling the vertical variogram in Figure 5.10-b and accounting for the uncertainty in the sill. Figure 5.14 shows the upper and lower limits of the horizontal seismic-derived variogram $\gamma_{w|\gamma_s}$. These limits are enforced on variogram realizations up to correlation 0.2 ($\rho(\mathbf{h}) = 0.2$) between the well and seismic data. Figure 5.15 shows the improved horizontal variogram realizations by four approaches: a) Enforce the seismic-derived variogram on γ_w, and merged with $\gamma_{w|\gamma_{ver}}$. b) Enforce the seismic-derived variogram on $\gamma_{w|\gamma_{ver}}$, and merged with γ_w. c) Merge

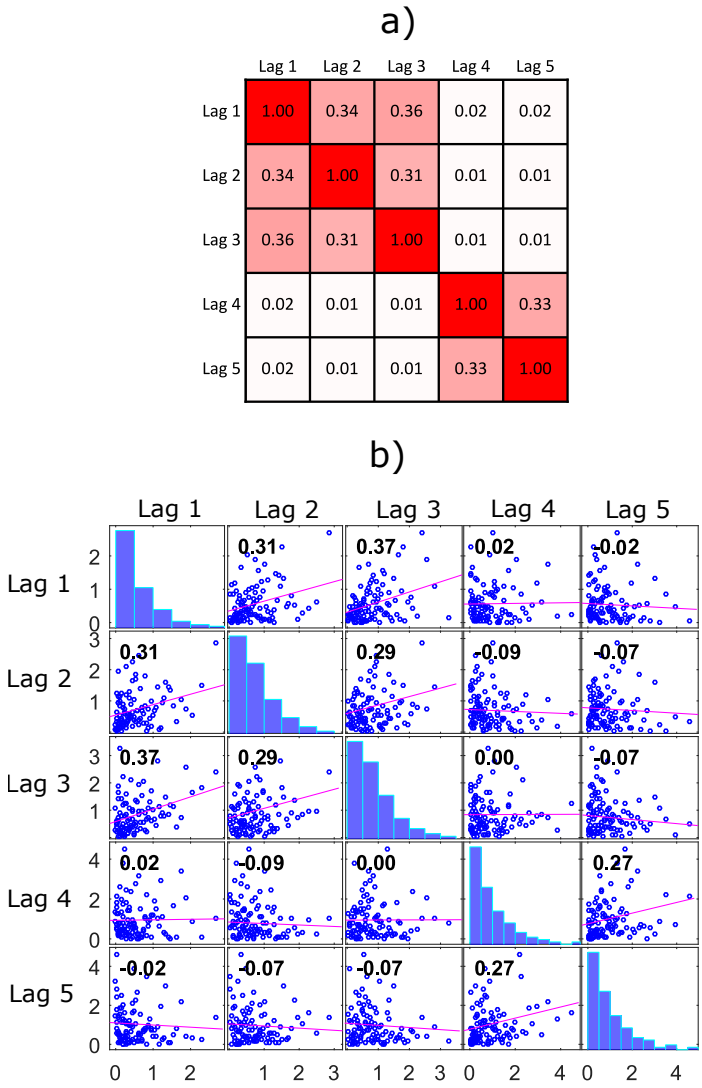

Figure 5.11: a) Correlation matrix between lag distances for the horizontal variogram of ten wells drawn from the Figure 5.7. b) Correlation matrix between lag distances for 100 variogram realizations after LU simulation and back transformation to the marginal Chi-square distributions.

γ_w and $\gamma_{w|\gamma_{ver}}$, and apply the seismic-derived variogram on the merged variogram. d) Enforce the seismic-derived variogram on γ_w and $\gamma_{w|\gamma_{ver}}$ and merge variograms. Although they look similar because of using the same merging approach (error ellipse technique), the difference between approaches could be used to choose the most reliable technique: since the real variogram of the ten drawn wells is available, the mean square error (MSE) between all variogram realizations and the real variogram for all lag distances is calculated. The approach that leads to a minimum MSE may signify

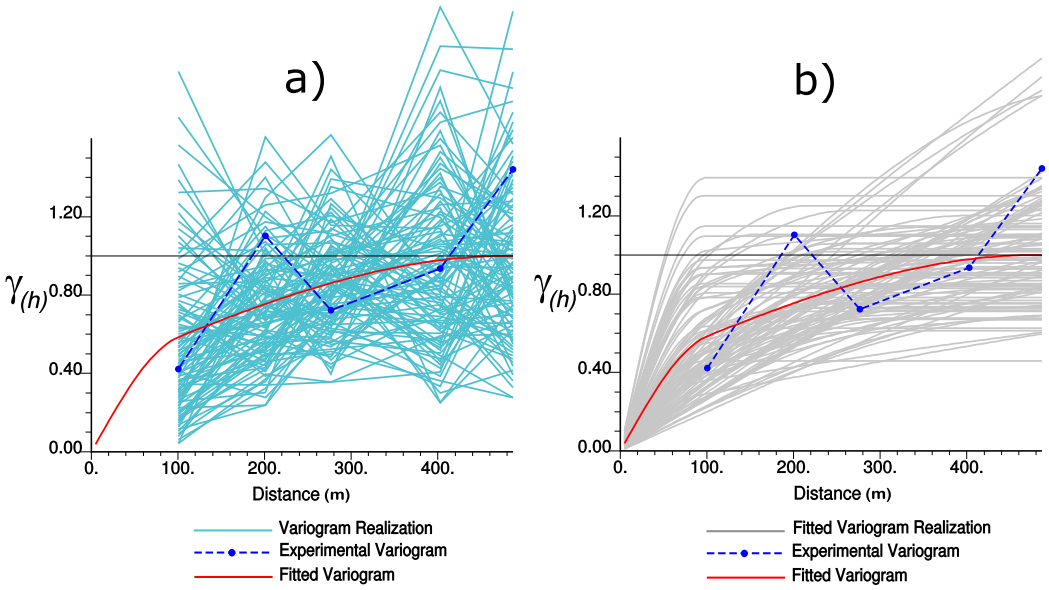

Figure 5.12: The horizontal variogram realizations (a) the fitted variogram realizations of ten drawn wells (b).

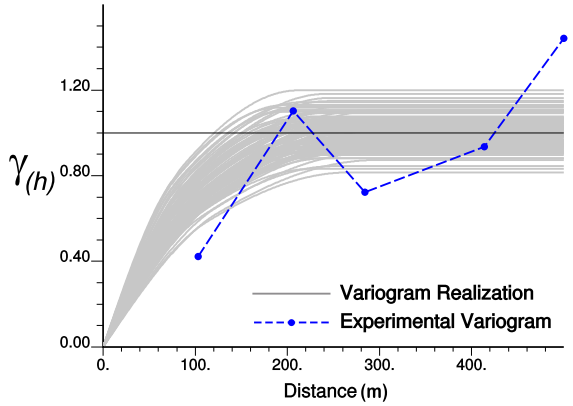

Figure 5.13: The vertical-derived variogram by scaling the vertical variogram in Figure 5.10-b via considering uncertainty in the H:V and sill: the triangular distribution with $Min = 15, Max = 28$ and $Mode = 21.5$ for H:V, and $\Delta_v = 0.1$ for building the Gaussian distribution of the sill.

the most reliable technique. The lowest MSE, which is 0.024284, achieved in case of enforcing the seismic-derived variogram on both γ_w and $\gamma_{w|\gamma_{ver}}$ and then merge variograms (see Figure 5.15-d). This ensures that the merged variogram is situated within the seismic-derived upper and lower limits.

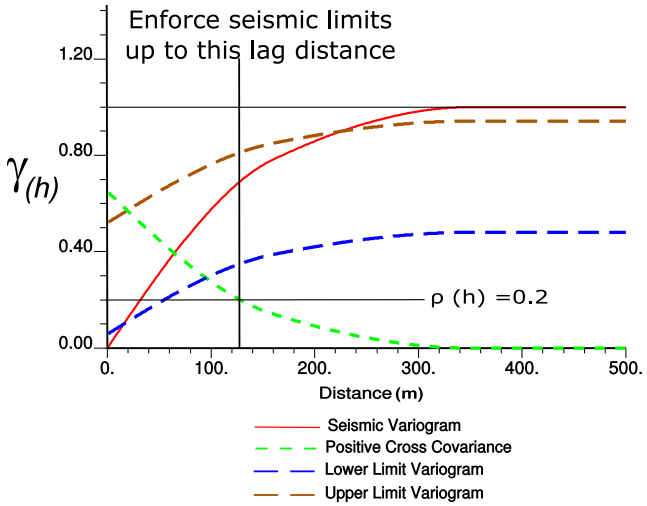

Figure 5.14: The horizontal seismic-derived variogram. The upper and lower limits are enforced up to correlation 0.2 ($\rho(\mathbf{h}) = 0.2$).

5.6 Multivariate Variogram Realizations

In case of multivariate geostatistical modeling, the variogram realizations should consider the correlation between variables: if variables are correlated, the variogram realizations between variables are correlated. For example, if two variables are highly correlated, a variogram realization with high range of one variable cannot be used with a variogram realization having low range of another variable in geostatistical modeling: both variogram realizations should have high range or low range. LU simulation is proposed to respect the correlation between variables for variogram realizations. Realizations of LU simulation (Deutsch & Journel, 1998) preserves the correlation between the elements in a correlation matrix. By LU simulation, multivariate Gaussian distribution of each variable is constructed and ranked. The variogram realizations for a specific lag distance are calculated for each variable and ranked; the lag distance should be reasonably chosen to differentiate variogram realizations. The ranked variogram realizations of the variables are ordered based on the ranked realizations of the LU simulations. The Step by Step procedure is:

1. Construct the positive correlation matrix of variables $\boldsymbol{\rho}$. If there is a negative correlation between variables, that should be converted to positive correlation be-

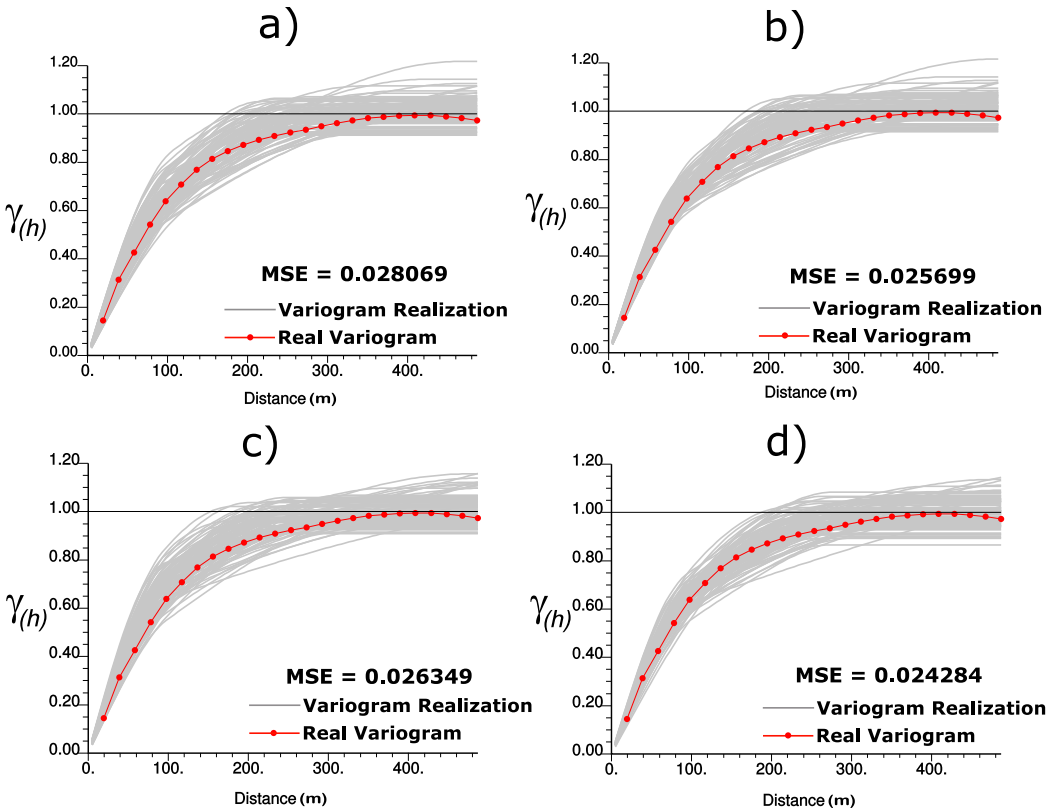

Figure 5.15: Improved horizontal variogram realizations by four different approaches: a) Enforce the seismic-derived variogram on γ_w, and merge with $\gamma_{w|\gamma_{ver}}$. b) Enforce the seismic-derived variogram on the $\gamma_{w|\gamma_{ver}}$, and merge with γ_w. c) Merge γ_w and $\gamma_{w|\gamma_{ver}}$, and apply the seismic-derived variogram on the merged variogram. d) Enforce the seismic-derived variogram on γ_w and $\gamma_{w|\gamma_{ver}}$ and merge variograms. Approach d leads to a minimum mean square error (MSE).

cause even though variables are negatively correlated, the variogram realizations are positively correlated.

2. Compute the Cholesky decomposition of the correlation matrix as $\boldsymbol{\rho} = \mathbf{L}\mathbf{L}^T$
3. Simulate a vector of uncorrelated standard normal deviate $\mathbf{w}$
4. Generate a vector of correlated Gaussian realizations $Y = \mathbf{L}\mathbf{w}$
5. Rank correlated Gaussian realizations for each variable $\mathbf{Y}$
6. Calculate variogram values for a specific lag distance for all variogram realizations of variables Z
7. Rank the variogram values of each variable $\mathbf{Z}$

8. Order **Z** based on **Y** for all variables

this ensures the positive correlations between the variables are reproduced for variogram realizations. Then, the horizontal variogram realizations of each variable can be autofitted with the vertical variogram to achieve 3D variogram realizations of each variable.

5.7 Realistic Example: Hekla Case Study

The Hekla data is presented for improving the horizontal variogram of well data. The Hekla reservoir is a part of North Sea fluvial deposit (offshore) located in Norway. This data set is suitable for the proposed methodology of improving horizontal variogram by the vertical-derived and the seismic-derived variograms. 3D seismic data of this data set is not available, there is only an average of acoustic impedance for the entire thickness of the reservoir. The total thickness of the reservoir is approximately 60m (H1+H2, H1: Horizon 1, H2: Horizon 2). The cell size for X and Y directions of seismic data are 50m; hereby, the scale of seismic data is $X = 50m \times Y = 50m \times Z = 60m$. Figure 5.16 shows the average acoustic impedance (H1+H2 Impedance) and location map of the well data for this data set. There are some missing acoustic impedance values. The input data includes twenty wells. Three wells 8, 15, and 16 are eliminated for this study because there is no acoustic impedance: the variogram of the calculated acoustic impedance from well data is used for downscaling the variogram of the original seismic data to the scale of the well data (see Section 5.2). The variable of well data for improving the horizontal variogram for this study is Log Porosity. The vertical scale of Log Porosity is 1m after upscaling. Figure 5.17 shows Log Porosity of seventeen wells for this data set.

Directional horizontal and vertical variograms of Log Porosity are calculated. Since major and minor directions of continuity cannot be seen in the well data, these directions are obtained from seismic data: according to Figure 5.16, the approximate major and minor directions of continuity are azimuths $25°$ and $-65°$. The horizontal variograms of Log Porosity are calculated for azimuths $25° \pm 30°$ and $-65° \pm 30°$. Both the

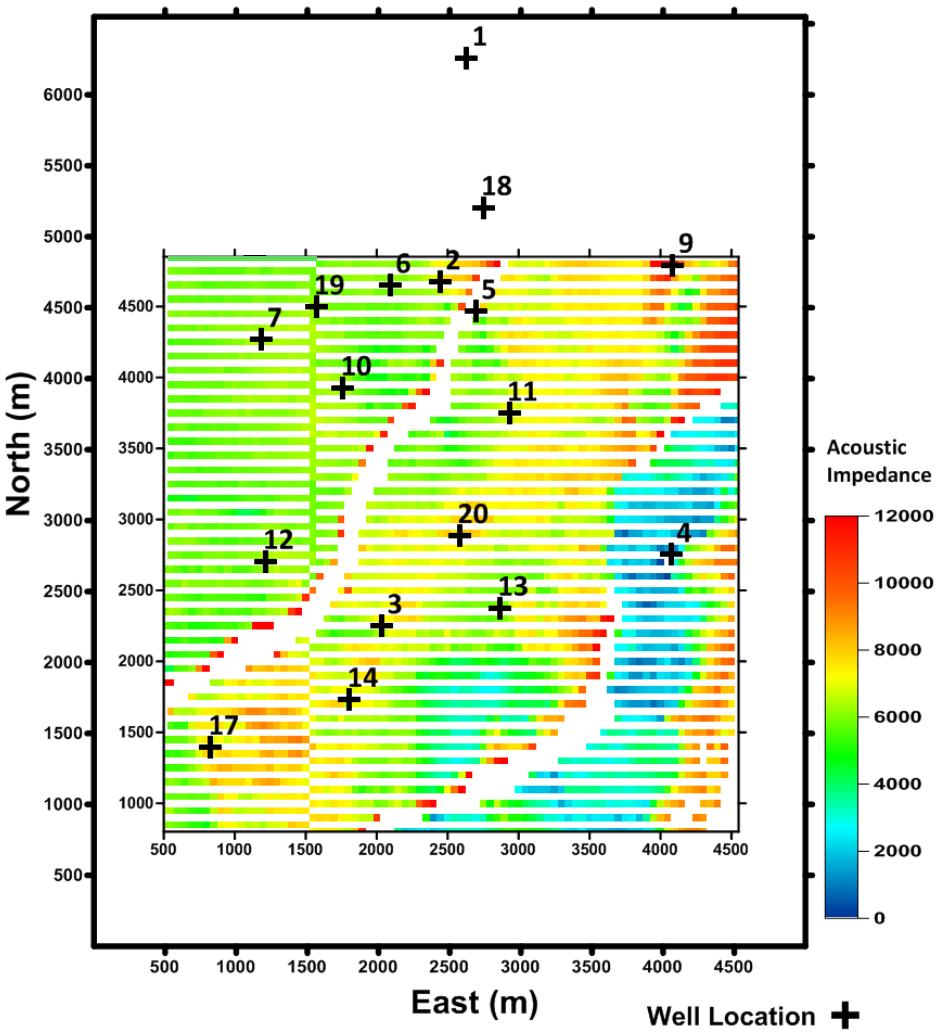

Figure 5.16: Average of the acoustic impedance for the entire reservoir (H1+H2 Impedance) and locations of seventeen wells for Hekla data set.

well and seismic data are normal scored. Figure 5.18 shows the normal score horizontal and vertical variograms of Log Porosity and fitted models. Due to sparse well data, the horizontal variograms are noisy and unreliable, and do not show the spatial correlation. The vertical variogram is well-defined because of regular sampling in the vertical direction. For calculating the horizontal variogram realizations, correlation matrix between the lag distances of azimuths $25°$ and $-65°$ is shown in Figure 5.19. Because of sparse well data, the correlations are relatively low. Variogram uncertainties are calculated by the DoF approach (Chapter 3); LU simulation is applied with the correlation matrix of lag distances and back transformed to the marginal Chi-square distribution. Figures

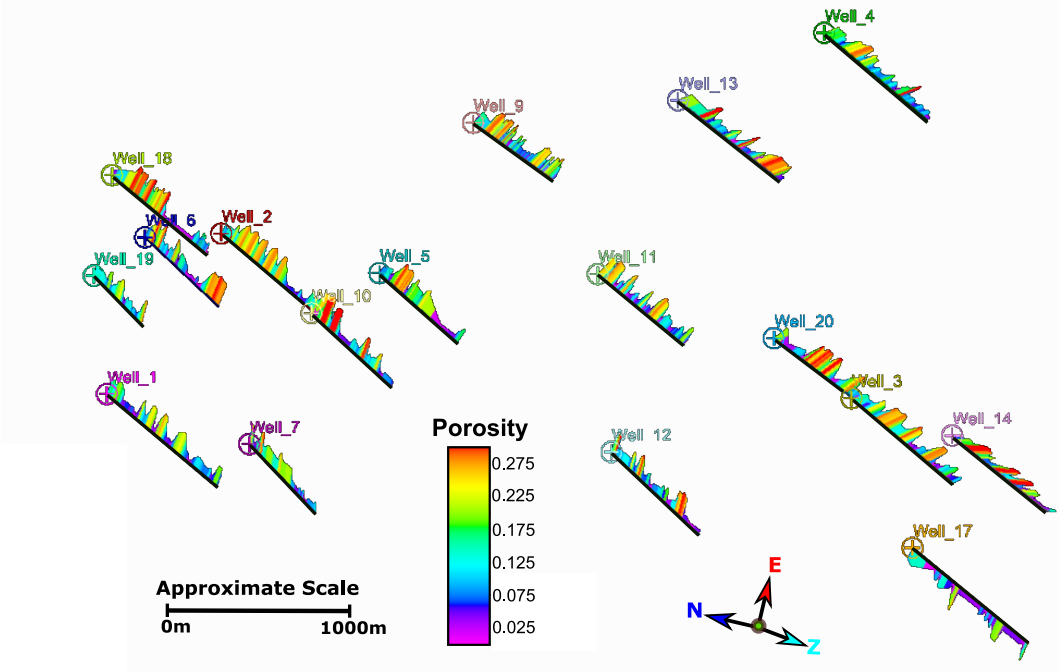

Figure 5.17: 3D Log Porosity of seventeen wells for Hekla data set (see Figure 5.16). The wells are sparse. The approximate scale 1000m is a horizontal scale.

5.20-a and 5.21-a show 100 horizontal variogram realizations for azimuths $25° \pm 30°$ and $-65° \pm 30°$, respectively. Figures 5.20-b and 5.21-b show the fitted variogram realizations. There is a high variogram uncertainty for both directions. The objective is to reduce the horizontal variogram uncertainty using the vertical-derived and seismic-derived variograms.

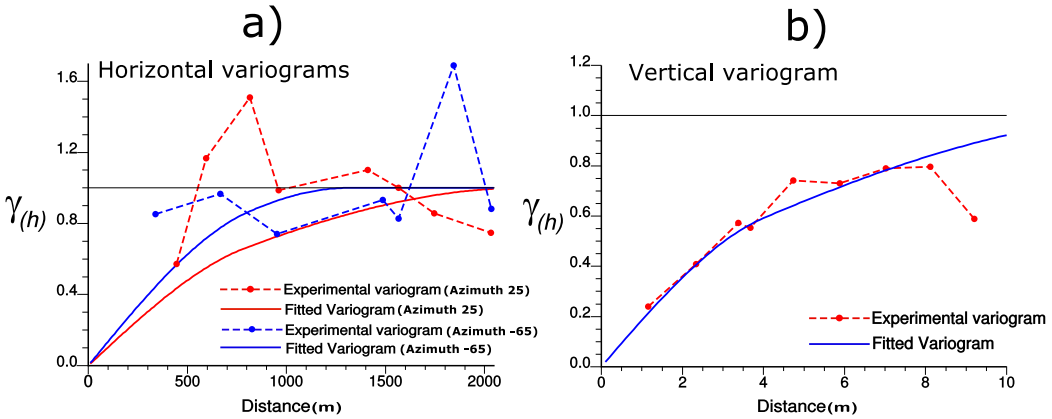

Figure 5.18: a) The horizontal variograms of Log Porosity and fitted models for azimuths $25° \pm 30°$ and $-65° \pm 30°$ (normal score). b) The vertical variogram of Log Porosity and fitted model (normal score).

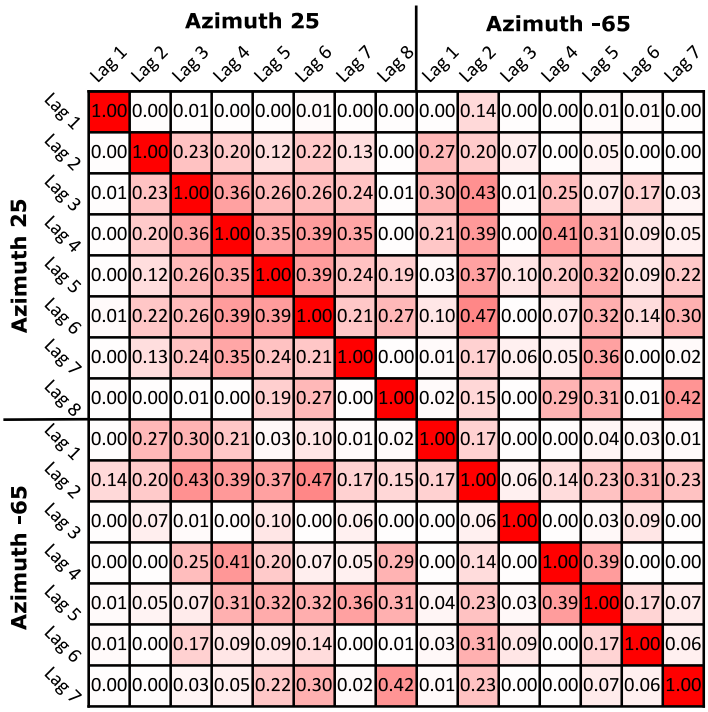

Figure 5.19: Correlation matrix between eight lag distances of azimuth 25° and seven lag distances of azimuth −65°.

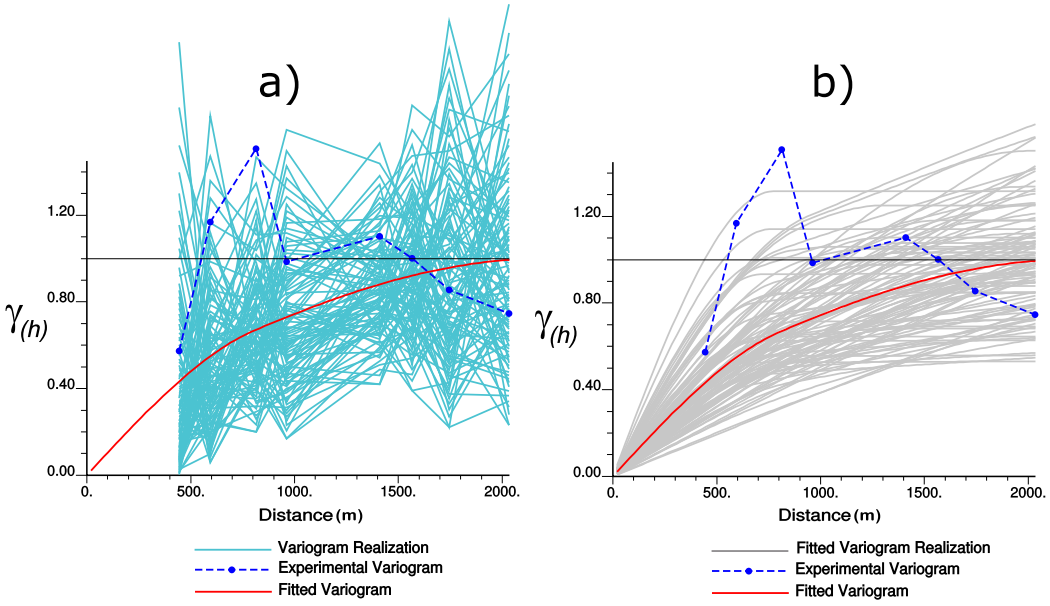

Figure 5.20: The horizontal variogram realizations of Log Porosity for azimuths 25°± 30° (a) and the fitted variogram realizations (b).

The vertical-derived variogram $\gamma_{w|\gamma_{ver}}$ could be calculated for both major and minor directions of continuity if H:V is available for these directions from the conceptual

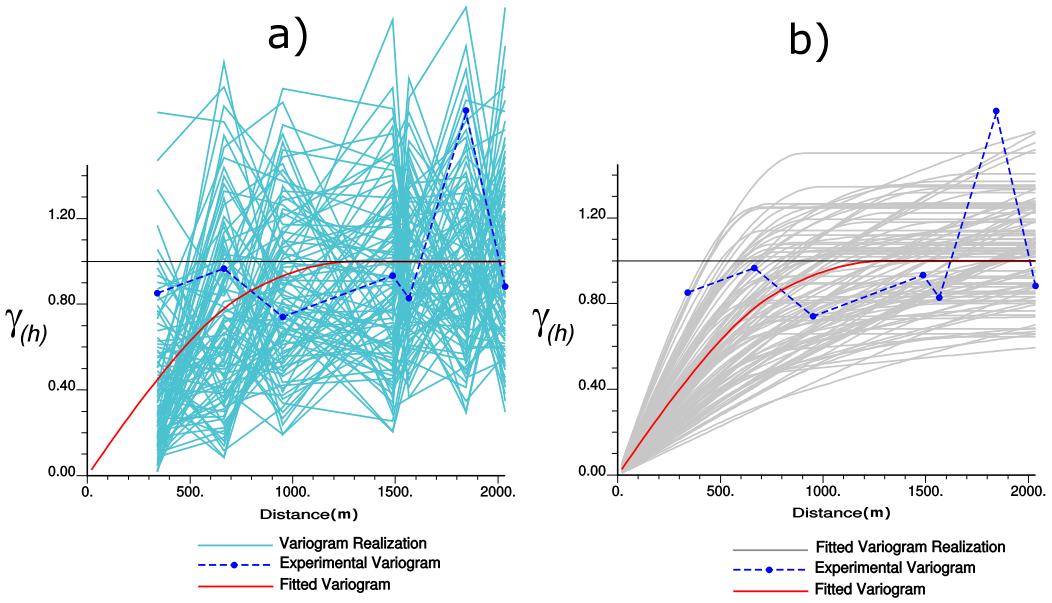

Figure 5.21: The horizontal variogram realizations of Log Porosity for azimuths $-65° \pm 30°$ (a) and the fitted variogram realizations (b).

geological models. However, there is usually one calculated H:V for reservoir. Figure 5.22 shows the vertical-derived variogram $\gamma_{w|\gamma_{ver}}$; uncertainty in H:V is calculated by a triangular distribution with $Min = 80$, $Mode = 100$ and $Max = 120$, and uncertainty in the sill by the Gaussian distribution with $\Delta_v = 0.2$. Simulation is used to draw samples of H:V and sill. $\gamma_{w|\gamma_{ver}}$ is achieved by scaling the vertical variogram (Figure 5.18-b) via the drawn H:V value and assigning the drawn sill value for each scaled-vertical variogram.

The seismic variogram should be down scaled before calculating the upper and lower limits of the seismic-derived variogram. For this data set, there is only a 2D average acoustic impedance over the entire reservoir instead of 3D seismic data. According to the variogram scaling law, both 3D large scale and small scale variograms are required (see Equation 5.3). The large scale variogram is the variogram of seismic data at the scale of $X = 50m \times Y = 50m \times Z = 60m$. The small scale variogram of seismic data at the scale of well data is unknown. The variogram of the calculated acoustic impedance at the scale of well data is available for this data set. Figure 5.23 shows the experimental horizontal and vertical variograms of the calculated acoustic impedance and fitted models. This leads to a 3D variogram model of the calculated acoustic

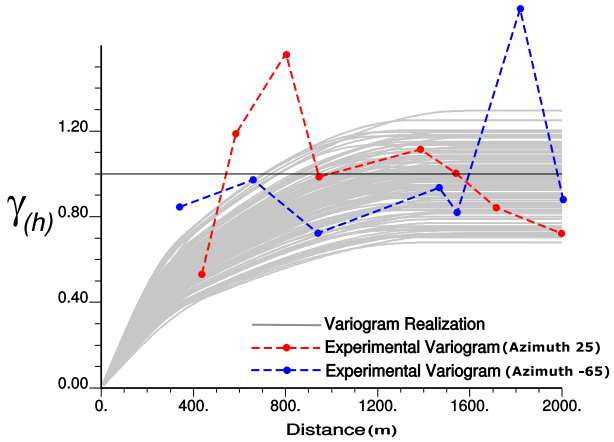

Figure 5.22: The vertical-derived variogram realizations by scaling the vertical variogram in Figure 5.18-b; uncertainty in the horizontal to vertical anisotropy ratio (H:V) is calculated by the triangular distribution with $Min = 80$, $Mode = 100$ and $Max = 120$ and the Gaussian distribution of the sill by $\Delta_v = 0.2$. The horizontal variograms for azimuths 25° and −65° are shown in this Figure.

impedance at the scale of well data used for downscaling the real seismic variogram to the scale of the well data. The vertical scale of well data is 1m and the horizontal scale

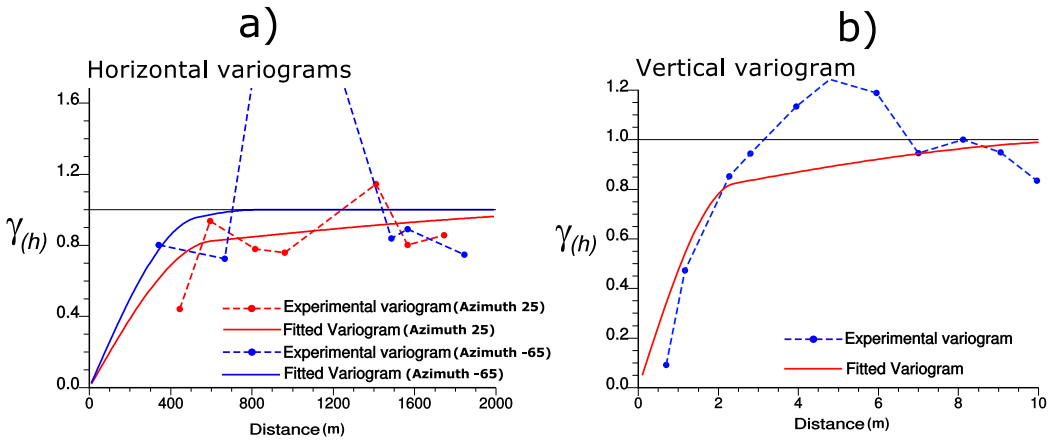

Figure 5.23: a) The horizontal variograms of the calculated acoustic impedance for wells and the fitted models for azimuths 25°±30° and −65°±30° (normal score). b) The vertical variogram of the calculated acoustic and fitted model (normal score).

is assumed to be 0.25m. The impact of the vertical scale on downscaling is much higher than the horizontal scale: the average variogram for the small and large horizontal scale are approximately equal (see Equation 5.3). The scaling law for variogram range has a little impact on the down scaled variogram (see Equation 5.2). Figure 5.24 shows the

experimental variograms, fitted models and the standardized down scaled variograms of the seismic data for azimuths 25° and −65°. The down scaled seismic variogram is at the scale of the horizontal well variogram. The correlation between well and seismic at the scale of well data is required; as the scale of the seismic data decreases, the

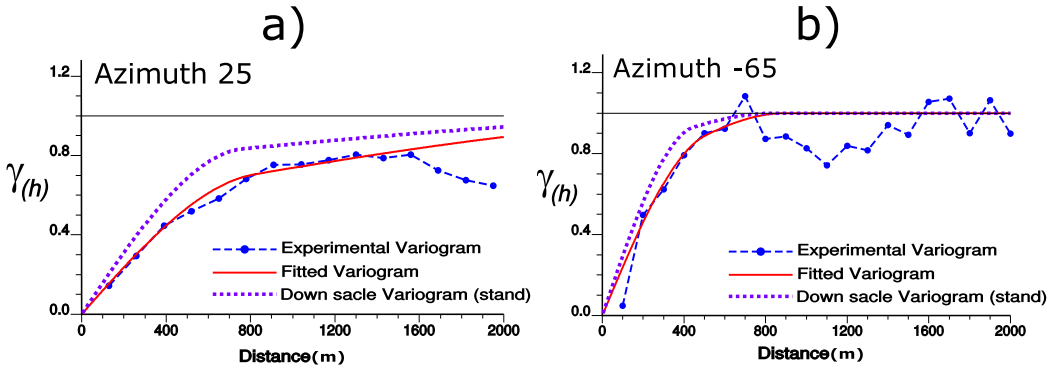

Figure 5.24: The experimental variograms, fitted models and standardized down scaled variograms of the seismic data (see Figure 5.16) for azimuths 25° (a) and −65° (b) (normal score).

correlation between well and seismic goes down. This correlation could be roughly estimated by fitting an equation to the different upscaled volumes and the observed correlations between well and seismic data. Since there is no 3D resolution in this data set for finding this equation, the correlation between the well and seismic data (acoustic impedance) at the scale of well data is assumed to be 0.4. The cross covariance is calculated by the intrinsic model. Figure 5.25 shows the upper and lower limits of the horizontal seismic-derived variograms $\gamma_{w|\gamma_s}$ for azimuths 25° and −65°. These limits are enforced on the horizontal variogram realization for azimuths 25° (Figure 5.20-b) and −65° (Figure 5.21-b) and the vertical-derived variogram $\gamma_{w|\gamma_{ver}}$ (Figure 5.22) up to correlation 0.2 ($\rho(\mathbf{h}) = 0.2$) by the rejection sampling approach with the correlation matrix of the lag distances (Figure 5.19). The final step is merging the horizontal variograms with the vertical-derived variogram for each direction.

Figure 5.26-a shows enforcing the seismic-derived variogram on only the horizontal variogram realizations without merging with the vertical-derived variogram for azimuths 25° and −65°. Since, the correlation between the well and seismic data is low ($\rho(0) = 0.4$), the impact of the seismic variogram to improve the high uncertainty in

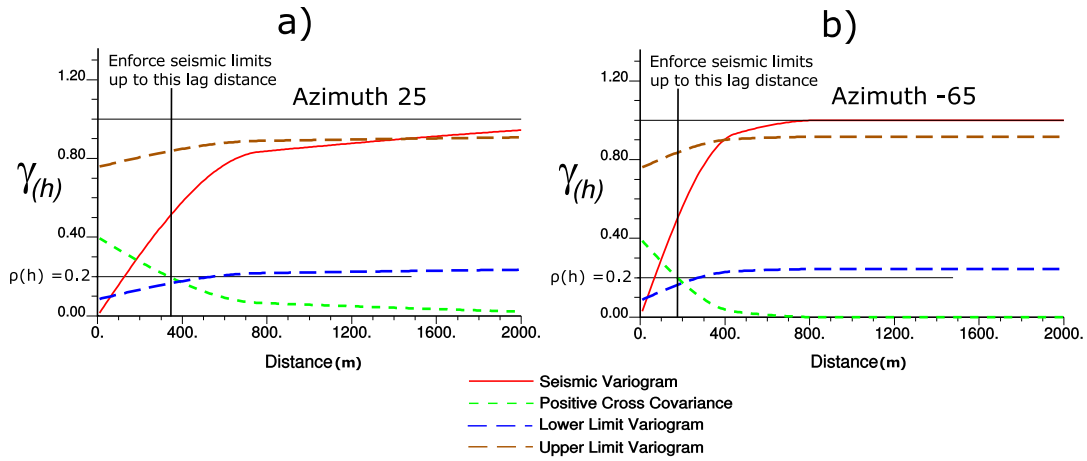

Figure 5.25: The upper and the lower limits of the horizontal seismic-derived variograms for azimuths 25° (a) and −65° (a). These limits are enforced on the horizontal variogram realizations for azimuths 25° and −65° and the vertical-derived variogram up to correlation 0.2 ($\rho(\mathbf{h}) = 0.2$).

the horizontal variogram is not striking. Figure 5.26-b shows merging the horizontal variogram realizations with the vertical-derived variogram without enforcing the seismic-derived variogram for azimuths 25° and −65°. This is the case when seismic data is not available. The uncertainty goes down considerably because the uncertainty in the vertical-derived variogram is lower than the horizontal variogram and merging variogram distributions by error ellipse approach leads to less uncertainty. Figure 5.26-c shows the final improved variogram realizations for azimuths 25° and −65° considering both the vertical-derived and seismic-derived variogram: enforce upper and lower limits on both the horizontal variogram and the vertical-derived variogram realizations, and merge variogram distributions and sample from the merged distribution by LU simulation. The final horizontal variogram is strongly influenced by the vertical-derived variogram. These realizations are fitted with the fixed vertical variogram by auto variogram modeling software to attain the 3D variogram realizations. These variogram realizations are used in geostatistical modeling to incorporate variogram uncertainty.

5.8 Remarks

The horizontal variogram of sparse well data has high uncertainty. A new approach is presented to improve the high uncertainty by the well-defined vertical variogram from

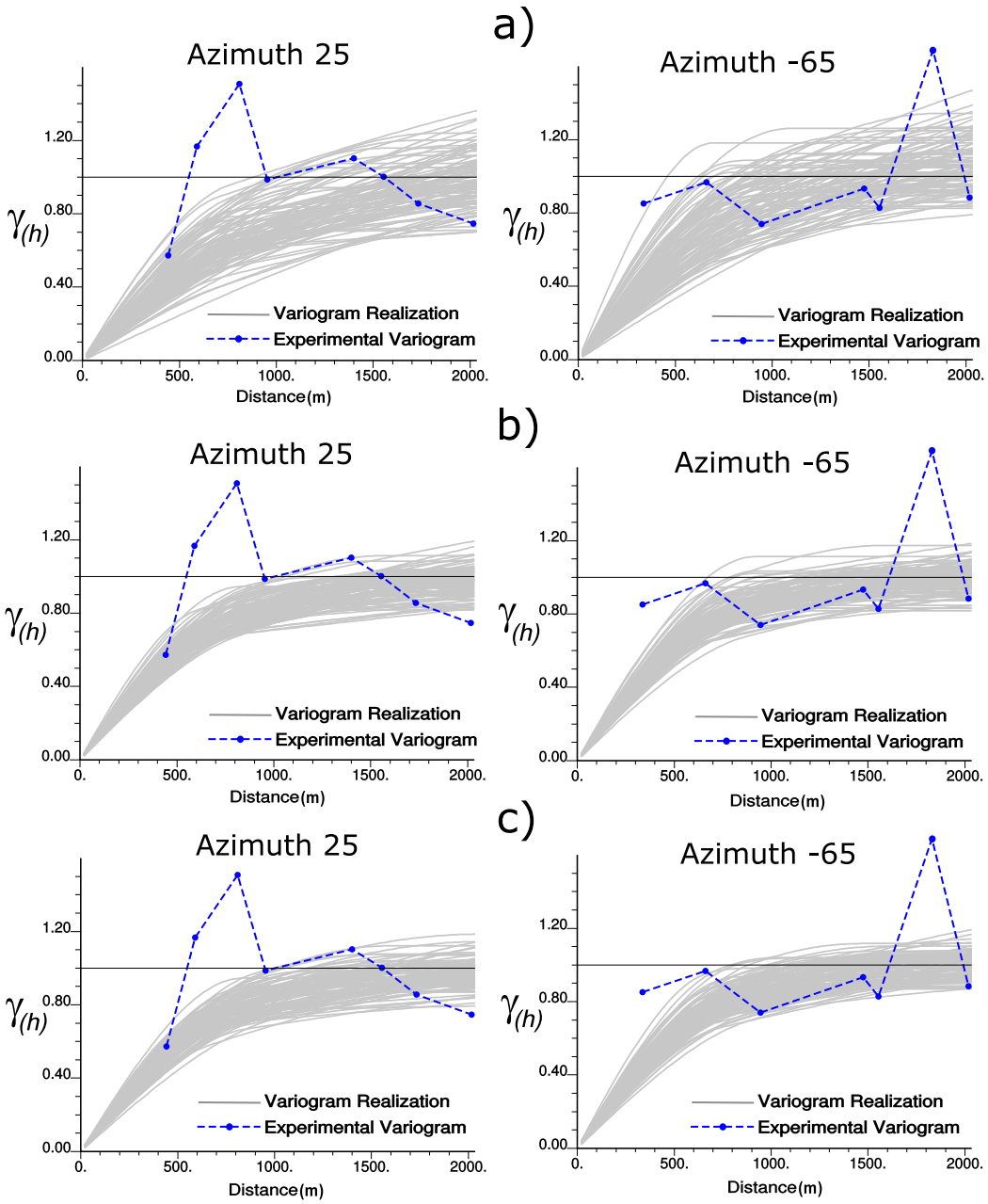

Figure 5.26: a) Enforce the upper and lower of the seismic-derived variogram on the horizontal variogram realizations without merging with the vertical-derived variogram for azimuths 25° and −65°. b) Merge the horizontal variogram realizations with the vertical-derived variogram without enforcing the seismic limits for azimuths 25° and −65°. c) Final improved variogram realizations for azimuths 25° and −65° considering the horizontal variogram of seismic data and the vertical variogram of well data.

well data and the horizontal variogram from seismic data. The uncertainty in the 3D variogram of well data is quantified by the uncertainty in the horizontal variogram.

The vertical variogram is assumed without uncertainty because of regular sampling in the direction of drilling. By improving uncertainty in the horizontal variogram, the 3D variogram realizations will be more reliable. These realizations could be used in geostatistical modeling to transfer variogram uncertainty to the final model. Improved variogram uncertainty leads to more accurate forecasting due to the fact that variogram uncertainty has more impact on flow than resource estimation (see Chapter 3).

Figure 5.27 shows a work flow for achieving the final variogram realizations. First, the horizontal and vertical variograms of well data should be calculated and a theoretical model is fitted for normal score data. In case of preferential sampling, the theoretical model could be used for declustering the experimental variogram to have a new fitted variogram model (see Chapter 2). The new fitted variogram model is used to calculate variogram uncertainty. The proposed methodology works for both 2D and 3D data set. For 2D data set, there is no vertical variogram (see Chapter 4). Since well data are vertically averaged (upscaled) for the entire reservoir as well as seismic data, well and seismic data are approximately at the same scale, hence, downscaling the seismic variogram before calculating the seismic-derived variogram is not required. Variogram realizations for major and minor directions of continuity are computed (see Chapter 3). Final 2D variogram realizations are produced by enforcing the seismic-derived variograms on the variogram realizations by rejection sampling (see Chapter 4, Section 4.3.1).

The methodology for 2D data set can be extended for 3D data set. In this case, the vertical variogram can also be used to improve variogram uncertainty. The horizontal variogram realizations are calculated for major and minor directions of continuity. The vertical variogram is converted to scenarios of the horizontal variogram by considering uncertainty in the horizontal to vertical anisotropy ratio (H:V) and uncertainty in the sill. The horizontal variogram of seismic data are calculated and down scaled to the scale of well data. An intrinsic cross covariance model (Markov model) (Pyrcz & Deutsch, 2014) is calculated by the correlation between well and seismic data at the scale of well data. This correlation could be achieved by fitting an equation to the different upscaled volumes and the observed correlations for data set. The resulting

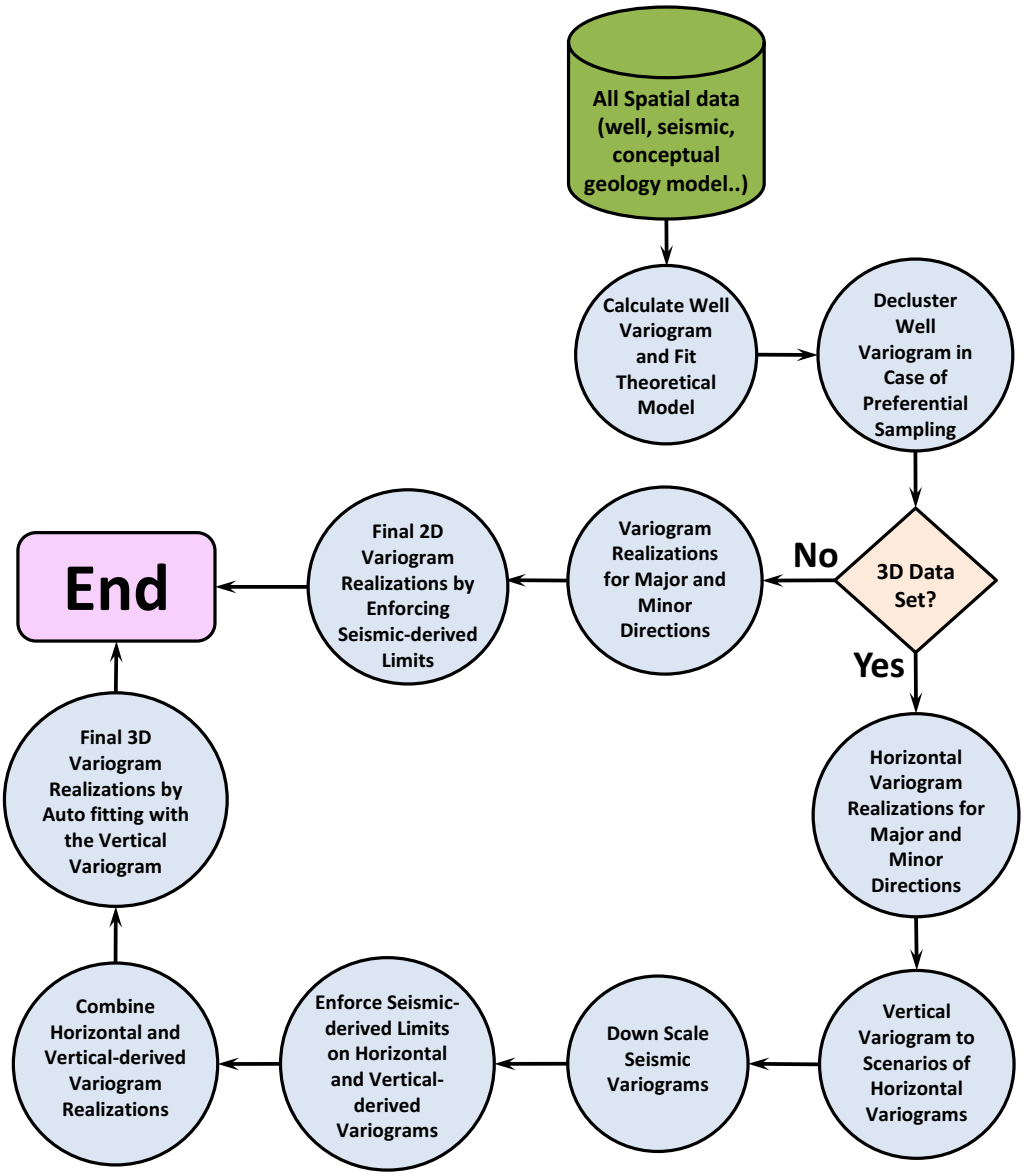

Figure 5.27: Work flow for improving variogram uncertainty.

equation may give an approximate correlation between well and seismic data at any scale. The upper and lower limits of the seismic-derived variogram are calculated and applied on the horizontal variogram realizations and the vertical-derived variogram by rejection sampling. The resulting variogram distributions after applying seismic constrains are merged and sample again by LU simulation with the correlation matrix of lag distances to achieve the improved horizontal variogram realizations.

The horizontal variogram realizations should consider the correlation between vari-

ables for multivariate geostatistical modeling (Section 5.6). LU simulation is proposed with the positive correlation matrix between variables. This ensures that the positive correlations between the variables are reproduced for variogram realizations. Then, the horizontal variogram realizations of each variable can be autofitted with the vertical variogram to obtain the final 3D variogram realizations of each variable.

Chapter 6

Multivariate Distribution Uncertainty

Required statistics such as the mean and the variance of the variables being modeled are often uncertain due to limited data and preferential sampling. Based on geological and geophysical data, the wells are likely located in the areas of higher quality to maximize production. Naive equal weighted statistics will lead to biased resource estimates and inappropriate reservoir development planning. Similar to variogram declustering (see Chapter 2), the univariate distributions of the variables should be declustered to achieve representative distributions.

The conventional declustering algorithms such as cell-declustering may correct the statistics and lead to a representative histogram if there are enough wells in good and poor areas; however, there are unlikely to be enough wells in the stage of development of the reservoir (Pyrcz & Deutsch, 2014). In such case, a representative distribution of the well data for each variable could be obtained with the aid of soft data such as exhaustive seismic data or geological interpretation. Debiasing or soft data declustering techniques (Deutsch et al., 1999, 2005) are used to construct a representative distribution based on conditional distributions inferred from the calibration with the soft secondary.

Similar to variogram uncertainty, there is unavoidable uncertainty in the representative histogram. This uncertainty should be quantified correctly because it affects resource/reserve estimation, investment and development decisions. This Chapter aims at developing a technique to evaluate and improve univariate parameter uncertainty,

that is, uncertainty in the distribution or histogram of each variable. The Chapter begins by introducing a new approach of quantifying histogram uncertainty followed by evaluating all approaches relative to the "true" uncertainty for a single variable. Multivariate parameter uncertainty is developed based on the correct technique. A realistic case study is presented.

6.1 Posterior Histogram Uncertainty

Uncertainty in univariate distribution of regionalized variables (input parameters) can be attained by different approaches. 1- The bootstrap (Efron, 1979) draw values from the data distribution to simulate different possible data sets. Since this approach does not account the spatial correlation between data, it may not lead to the correct uncertainty in the histogram. 2- The spatial bootstrap is an extension of the bootstrap technique that considers the spatial correlation of data (Deutsch, 2004; Journel & Bitanov, 2004) by unconditional LU simulation at data locations. The conditioning data and the area of interest are not considered in this technique. 3- Conditional finite domain (CFD) accounts for the conditioning data and the size of the domain. Generally, CFD is very difficult to operate and leads to very low uncertainty (Babak & Deutsch, 2009). 4- Global kriging can estimate the variance of the mean for the entire domain. It accounts for the conditioning data and the area of interest; the variance will decrease when the domain size increases (Deutsch & Deutsch, 2010).

A simulation-based method for quantifying uncertainty is proposed in this Chapter. Since the spatial bootstrap does not account for the conditioning data, it is likely to have unreliable high histogram uncertainty. The histogram uncertainty quantified by the spatial bootstrap is assumed to be a prior uncertainty. This prior uncertainty could be conditioned to achieve more reliable histogram uncertainty. This is called posterior histogram uncertainty and it is claimed to be more accurate than other techniques (Khan & Deutsch, 2016). The posterior histogram uncertainty is attained by using the spatial bootstrap realization as reference distribution for normal score transformation of data, and back transforming the conditional Gaussian realizations to original units.

This process is repeated for the number of realizations. The simulation area could be clipped to the domain limits for more accurate posterior histogram uncertainty. The procedure is summarized by:

1. Define a stationary covariance function C($\mathbf{h}$).
2. Define the reference distribution *F(z)*.
3. Perform the spatial bootstrap resampling as follows:
   a) Construct the spatial data-to-data covariance matrix $\boldsymbol{C}$.
   b) Compute the Cholesky decomposition of the correlation matrix as $\boldsymbol{C} = \mathbf{L}\mathbf{L}^T$.
   c) Simulate a vector of uncorrelated standard normal deviate $\mathbf{w}$.
   d) Generate a vector of correlated Gaussian values $y = \mathbf{L}\mathbf{w}$.
   e) Transform the unconditional Gaussian values to original units $z(\mathbf{u}_\alpha) = F^{-1}(G(y(\mathbf{u}_\alpha))), \quad \alpha = 1, ..., n.$ Where $\mathbf{u}_\alpha$ is a data location and n is the number of data.
4. Conditional simulation; the spatial bootstrap realization used as a reference distribution for normal score transformation.
5. Back transform the realization to original units with the spatial bootstrap reference distribution.
6. Each realization is limited to the domain limits.

Steps 3-c to 6 are repeated to achieve the posterior histogram uncertainty. This approach accounts for the conditioning data, the domain limits and the spatial correlation between data.

Figure 6.1 shows a 2D synthetic data set with in an area of $1000m \times 1000m$ with nine data locations. The variable is effective porosity. The uncertainty in the mean of the distribution is calculated by the prior and posterior uncertainty. An anisotropic variogram model is assumed with the range of 80m for 0° and 40m for azimuth 90°. Prior uncertainty in the mean is attained by averaging each realization of the spatial bootstrap. The variance of the mean is 2.11×10^{-2}. Posterior uncertainty in the mean

is attained by using each realization of the spatial bootstrap as a reference distribution for normal score transformation of data and back transforming the realizations from the Gaussian space to original units. The mean of each realization is computed. This process is repeated to attain the posterior uncertainty in the mean: the variance of the mean decreases from 2.11×10^{-2} (prior uncertainty) to 1.43×10^{-2} because of the conditioning data and domain limits.

The posterior histogram uncertainty is considered more accurate than the other techniques. The only way to check the approaches of quantifying histogram uncertainty is to design an experimental framework where the true uncertainty in the histogram is known. For this experiment, the spatial bootstrap (prior), the global kriging and posterior approaches are considered since the bootstrap and CFD are not considered viable alternatives for quantifying histogram uncertainty.

6.2 Methodology

The approaches of quantifying parameter uncertainty (histogram uncertainty) are evaluated by comparing to the "true" uncertainty to find the most correct approach; the spatial bootstrap, global kriging and posterior approaches are tested relative to the true parameter uncertainty. Although the experiment is applied for a single variable, the result of this experiment is applicable to multiple variables. The global uncertainty (variance) in the mean of distribution is used to evaluate parameter uncertainty approaches.

In practice, the true variance of the mean could be achieved by a scan-based approach, which looks for similar patterns of a data configuration within a large real image, and calculates the mean of the specified domain as the mean of the values within the domain conditioned to the specified data configuration. The term "data configuration" is a template of data locations with assigned values: the term "template" is only the location of data without value. The proposed scan-based approach for computing the true distribution of the mean is summarized by the following Steps:

1. Select a large grid of real data from remote sensing.

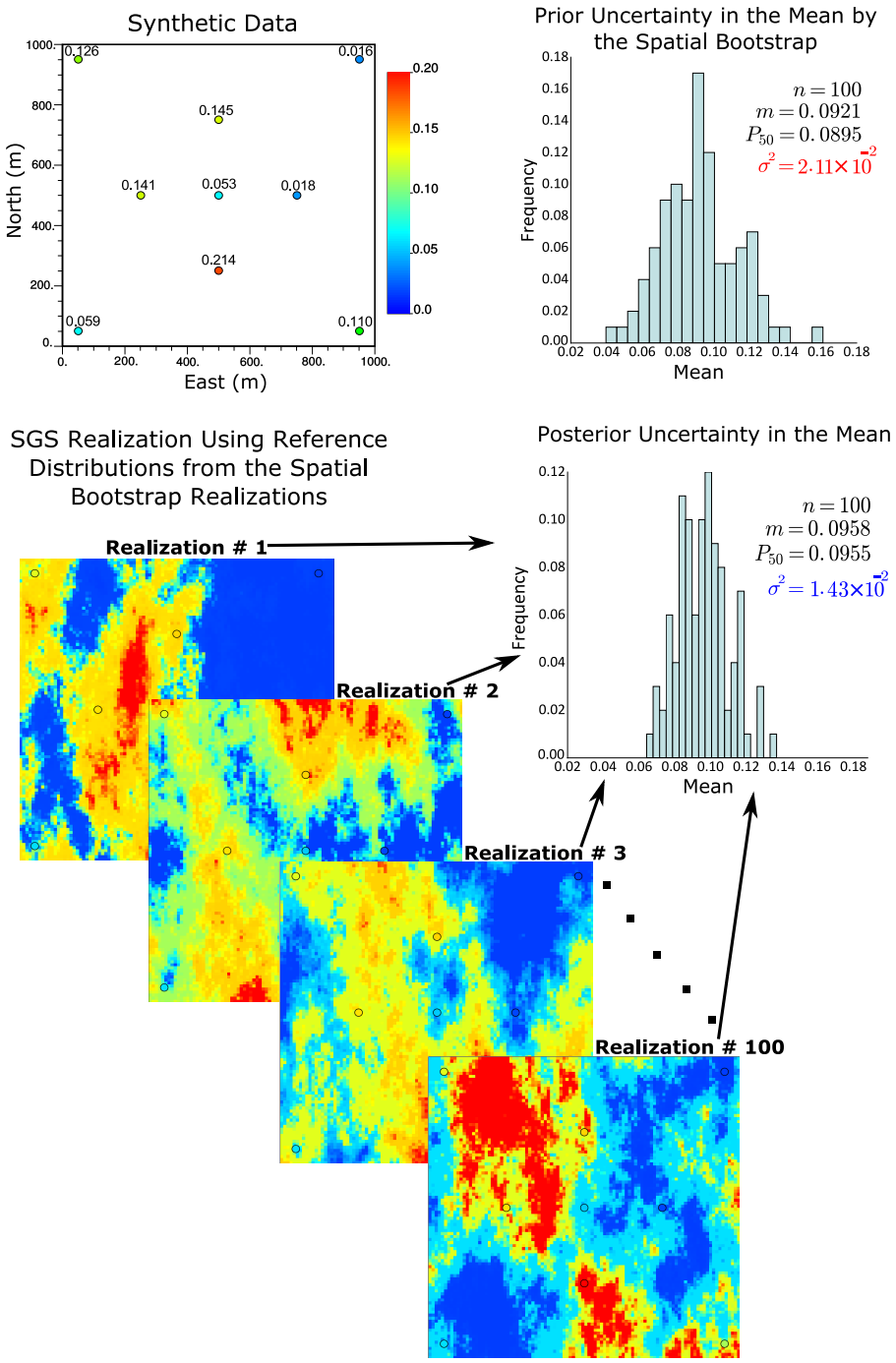

Figure 6.1: A synthetic 2D example of effective porosity. The uncertainty in the mean is calculated by prior and posterior uncertainty. The posterior approach leads to lower uncertainty in the mean of effective porosity: the variance of the mean decreases from 2.11×10^{-2} (prior uncertainty) to 1.43×10^{-2} because of the conditioning data and domain limits.

2. Find stationary area and average variogram by discretization of the image for different patch sizes.

3. Pre-scan the image for a template to find common data configurations.

4. Scan again the data configuration to find exact replicates within the image and average the whole domain near the data configuration to build the true distribution of the mean (see Figure 6.2).

Figure 6.2-a and Figure 6.2-b show a schematic and an example of the scan-based approach. A data configuration with five locations is scanned to find similar patterns in the image. The data configuration should have enough replicates within the image to make a reliable true distribution of the mean. A tolerance between the specified data configuration and the observed pattern in the image should be considered in order to find enough replicates. By scanning many data configurations and several templates over a large image, and calculate the mean of each replicate, the true uncertainty (variance) in the mean is attained. This would support selection of the best approach of quantifying histogram uncertainty.

6.3 Evaluate Parameter Uncertainty Techniques by True Uncertainty

6.3.1 Proper Data Configurations

The true variance of the mean could be attained by scanning an image to find similar replicates of data configuration (see Figure 6.2). The problem with the scan-based approach is finding proper data configurations. The data configurations should have enough replicates. This leads to a reliable distribution of the mean (more accurate variance of the mean). The index method proposed by Deutsch (1992) is applied to find frequent data configurations of a template in image:

$$index = 1 + \sum_{i=1}^{N} [z_i - 1].K^{i-1} \tag{6.1}$$

where K is number of classes which divides the data of image into different classes. N is number of data locations, and z_i is a class number of each data location in image.

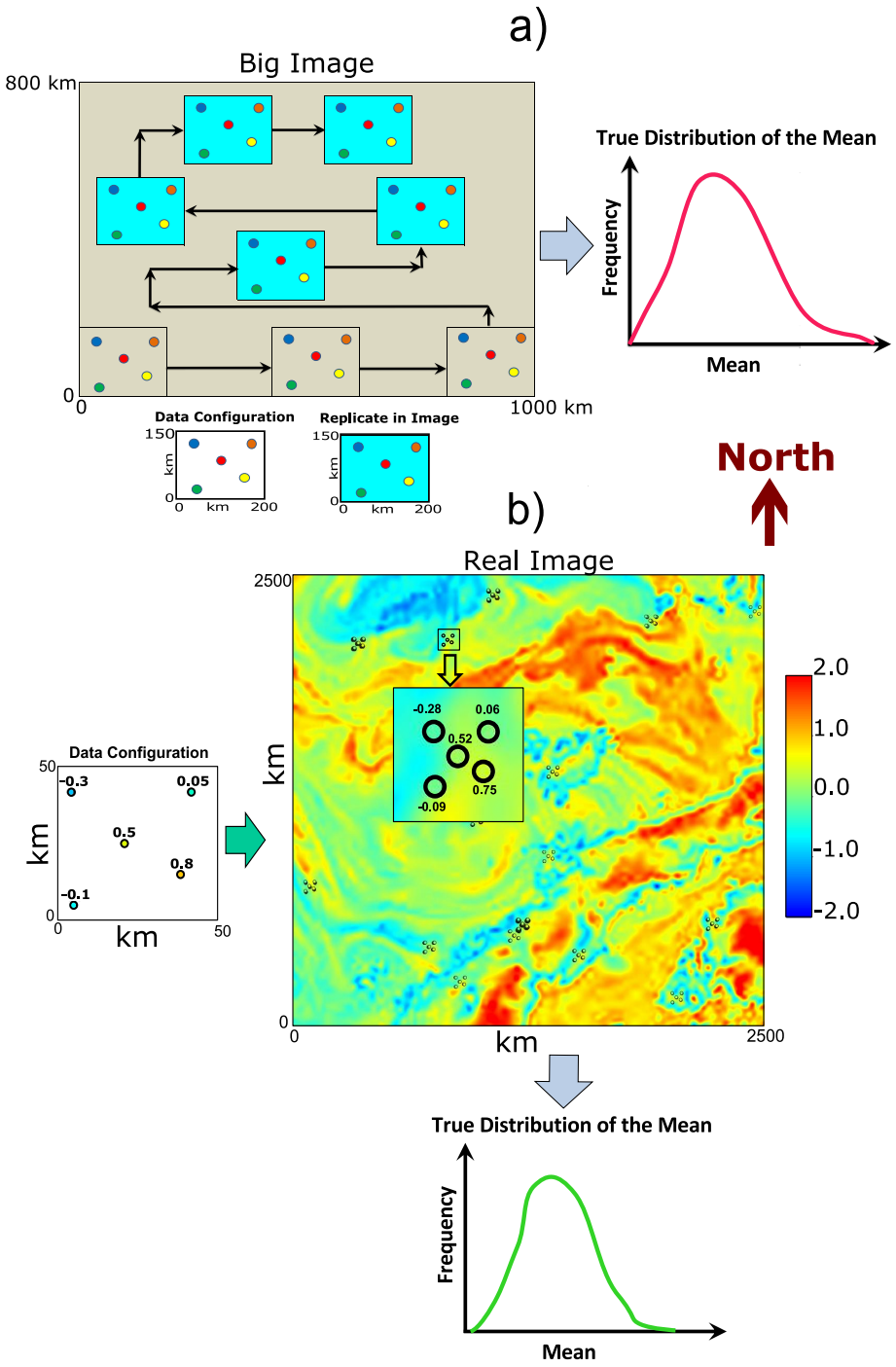

Figure 6.2: a) Schematic of the image-based approach for calculating direct observation of the mean uncertainty. b) An example of the scan-based approach: a data configuration with five locations is scanned within the image and similar patterns of this data configuration are found to build the true distribution of the mean.

By scanning a template (data location without value) in image, frequent indexes are observed. Figure 6.3 shows a schematic of the index approach for finding optimum data configurations that have enough replicates in the image. There is a template with 6 locations ($N = 6$), and the values of the image in this Figure are divided into 10 classes ($K = 10$). For example, class number 2 is between -2.3 to -1.7. This template is scanned within the image to select the index values that have reasonable frequencies (more than average frequency). These indexes are converted to the corresponding data configurations. Figure 6.3 shows how an index value of 251320 is converted to a data configuration. First, the class number of each location is calculated. Then, each class is converted to an appropriate value of image. Since each class number denotes a range of values, Monte Carlo Simulation (MCS) could be used to convert class numbers to appropriate values of the image.

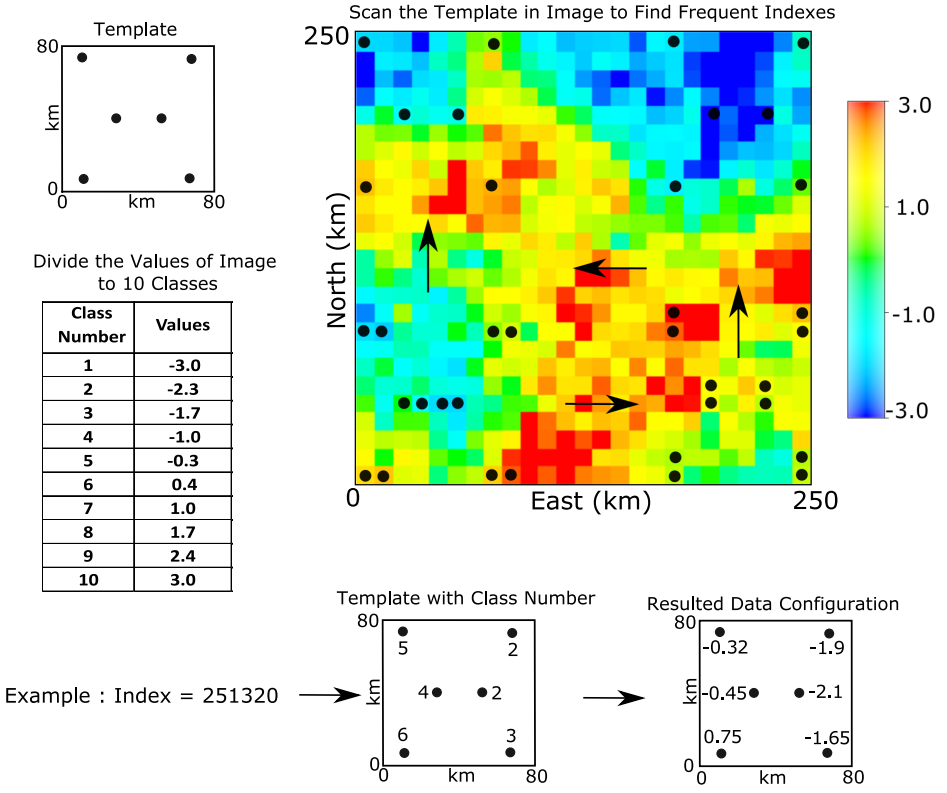

Figure 6.3: Schematic of index approach for finding optimum data configurations (high frequency in image) for the scan-based approach.

6.3.2 Error Correction

The resulting data configurations that have enough replicates should be used in the scan-based approach to calculate the true distribution of the mean. The bin size of the classes should be as small as possible. The smaller the bin size, the more classes, the more accurate and precise data configuration. However, using a large number of classes and data points is computationally demanding. There is a trade off between the number of classes, data locations and error between data configuration and pattern (replicate) in image. This error should increase with increasing number of data locations and bin size; otherwise, it would not be feasible to find sufficient replicates. Averaging each pattern with this error may lead to an inaccurate distribution of the mean. Figure 6.4 shows the error between data configuration and a replicate in the image for each location to the left, and the way of correcting this error at the right. The corrected true mean could be calculated by dividing the area of the data configuration into equal sub-areas including one data approximately in the middle of each area. For the entire each sub-area, the error is assumed to be the same. The Equation in Figure 6.4 (right) shows the true mean is equal to the expected value of the pattern (replicate in image) plus the expected value of the error ($E\{Pattern\} + E\{Error\}$).

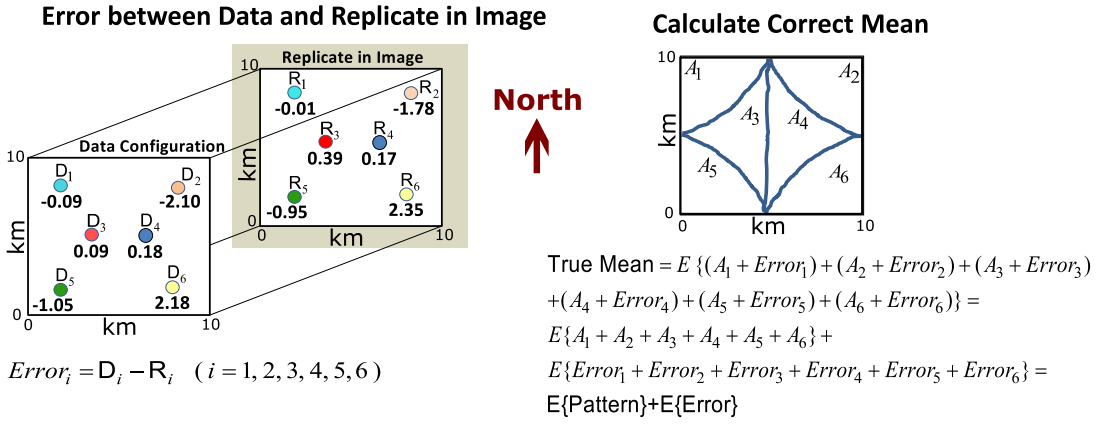

Figure 6.4: Error between data configuration and replicate in image (left), calculating the corrected true mean of the replicate (pattern) by adding the expected value of the error to the expected value of the pattern (right).

6.3.3 Stationary Area and Select Templates

Figure 6.5 shows an image applied for the scan-based approach. This image is color of the ocean of east Tasmania taken in December 2004 (NASA, n.d.). The original image is digitized and converted to a large numerical data having $nx = 5000 \times ny = 6000$ pixels. It is assumed a grid size of 1km because the length and width size of the original image is unknown. The data are transformed to normal score. The data

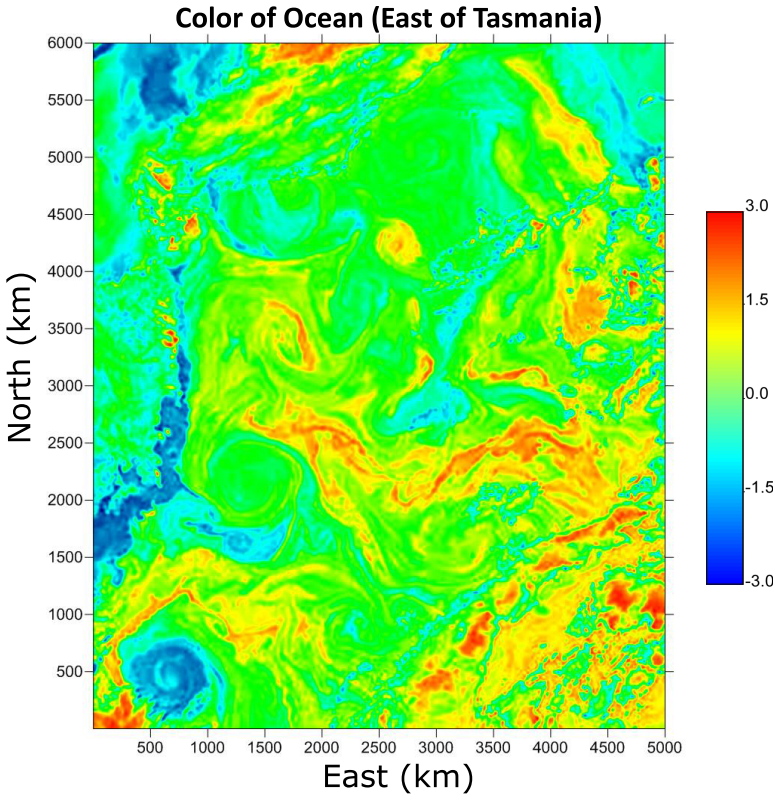

Figure 6.5: Color of the ocean of east Tasmania (NASA, n.d.).

configurations with stationary area should be taken from the image. The entire image is discretized to many square areas called patches (from small to big patches), then the normal score variogram of each patch is calculated. If the average of all variograms reaches to the expected sill in approximately half size of a patch, templates for the scan-based approach are selected within this patch size (area). Figure 6.6-a shows the discretized image (Figure 6.5) to patches $120km \times 120km$. The experimental variogram of each patch is calculated. Figure 6.6-b shows the average variograms for azimuths 0° and 90° reach approximately a sill of 60km for these patches ($120km \times 120km$): on

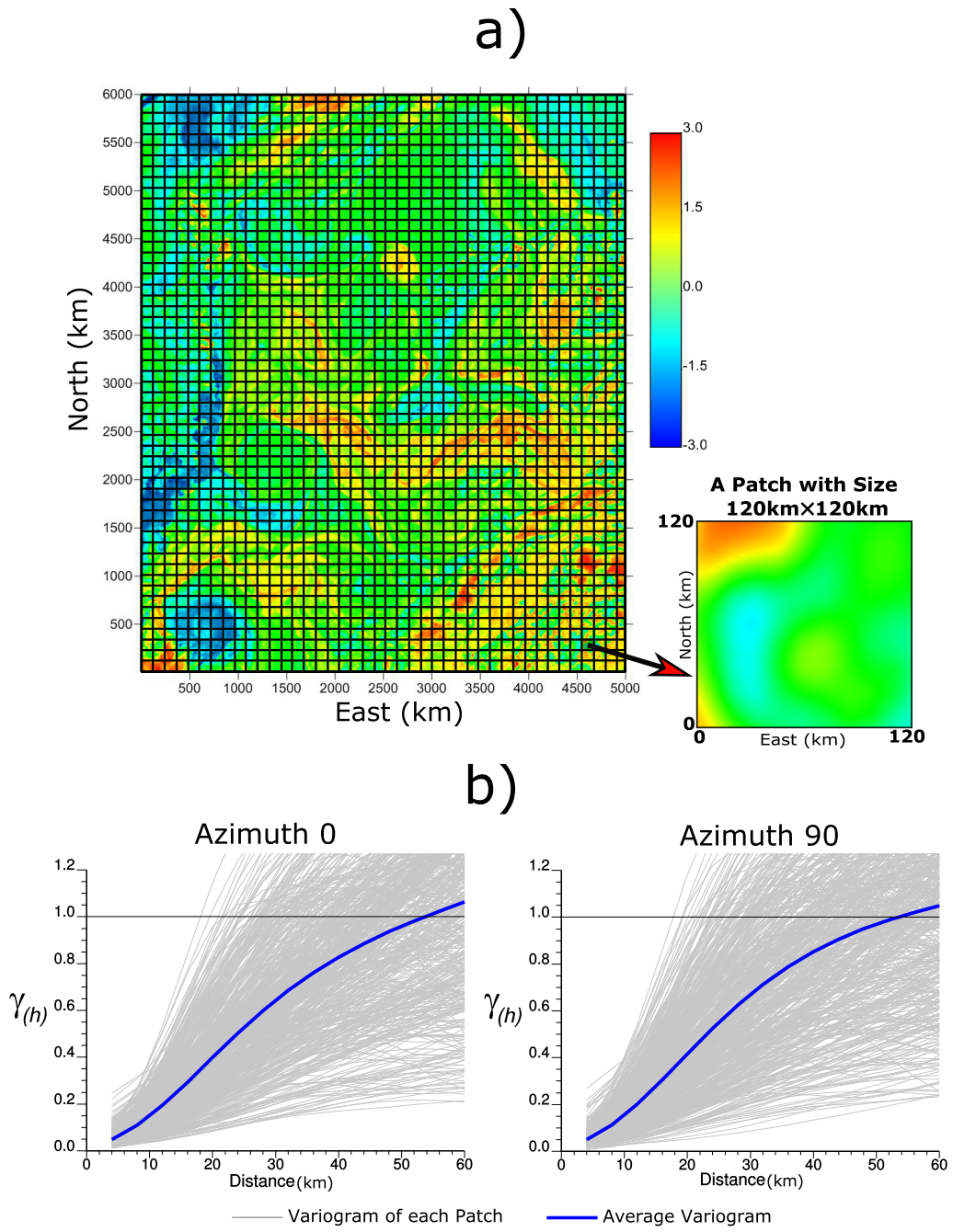

Figure 6.6: a) Discretized image of Figure 6.5 to patches $120km \times 120km$. b) The experimental variogram of each patch (gray lines) and the average variogram (blue line) are calculated for azimuths 0° and 90°.

average, there is no anisotropy. This area and the average variogram are used for the approach of quantifying histogram uncertainty. According to this area, four templates are selected with 6, 7, 8 and 9 data locations (see Figure 6.7). The result of the scan-based approach for templates smaller than 6 data locations might not be accurate. The

more data locations for a template, the more precise the distribution of the mean. The templates with more than 9 data locations are not used in this study since they are

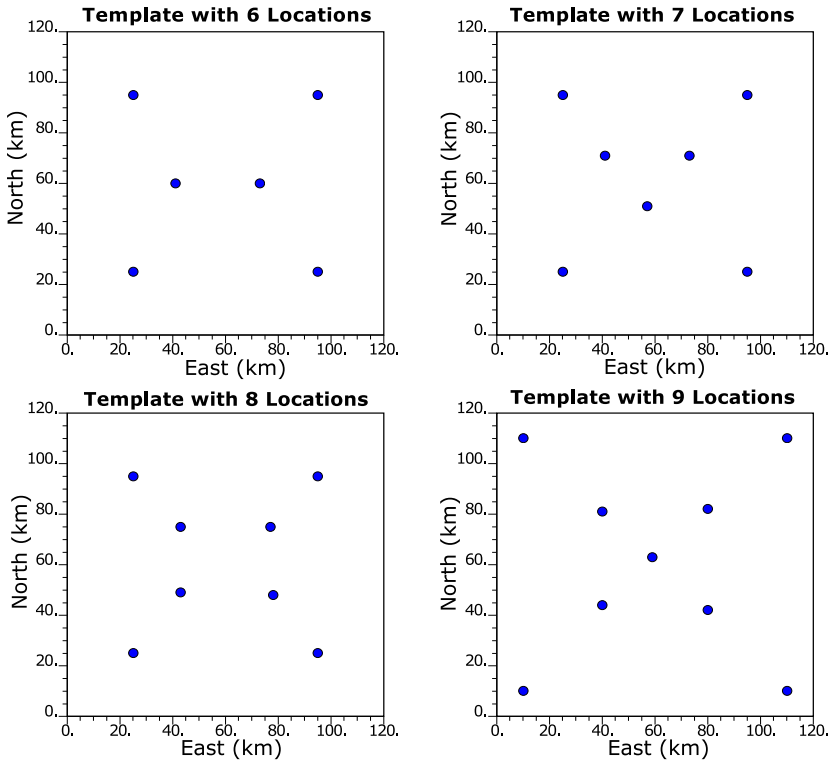

Figure 6.7: 4 templates with 6, 7, 8, 9 data locations used for scan-based approach.

computationally demanding. Moreover, as the number of data locations increases, the error between data configurations and patterns in the image should increase in order to find enough replicates. Increasing this error leads to incorrect distributions of the mean with the correction approach (Figure 6.4). The scan-based approach can be applied automatically on the image of Figure 6.5 for many data configurations obtained for each template (Figure 6.7) to calculate the true uncertainty in the mean. The index approach is used to select proper (frequent) data configurations within the image (see Section 6.3.1).

6.3.4 Quantify Uncertainty in the Mean by Parameter Uncertainty Approaches

The variance of the mean for the drawn data configurations from the image in Figure 6.5 based on the templates in Figure 6.7 are calculated by parameter uncertainty

approaches to compare with the true variance by the scan-based approach. The required variogram model for the spatial bootstrap, global kriging and posterior approach is achieved by the fitted variogram to the average variogram of the stationary area (patches $120km \times 120km$) in Figure 6.6-b:

$$\gamma(\mathbf{h}) = 0.665.Gau \underset{\substack{h_{major}=50.5km \\ h_{minor}=50.5km}}{(\mathbf{h})} + 0.335.Sph \underset{\substack{h_{major}=60.6km \\ h_{minor}=60.6km}}{(\mathbf{h})} \tag{6.2}$$

where Gau and Sph are Gaussian and spherical variogram structures. h_{major} and h_{minor} are major (azimuth 0°) and minor (azimuth 90°) variogram ranges for each variogram structure. The variance contributions of this variogram are scaled by the variance of each data configuration for the global kriging estimation of the uncertainty in the mean. Since there are just a few locations of data, quantile-quantile transformation for posterior approach may not be applicable (normal score transformation of the data and back transform to original units). If it is assumed that each data configuration has Gaussian distribution, the standard normal equation ($y = \frac{z-m}{\sigma}$) can be used. Where z is the value of each location of data configuration, y is the standardized value of each location, m and σ are different realizations of the mean and variance comes from the spatial bootstrap. Conditional SGS is applied with the standardized data configuration. Then, SGS realization is back-transformed, and calculate the arithmetic average of each realization to build the distribution of the mean. This process is repeated for 100 realizations to compute the variance of the mean by the posterior approach.

6.3.5 Results

Figure 6.8 shows the results of the true variance of the mean (Direct) for templates of Figure 6.7 versus the variance of the mean by the spatial bootstrap (SB), global kriging (GK) and posterior (Post) approaches. 150 data configurations that have enough replicates (more than average) in the image (Figure 6.5) with minimum error are selected for each template. The true variance of the mean is calculated for each data configuration. The spatial bootstrap (SB) estimates the variance of the mean much higher than

the true variance for all templates (prior uncertainty). The variance of the mean coming from the global kriging (GK) is always less than the true variance (approximately half). The posterior technique estimates the true variance of the mean more accurately

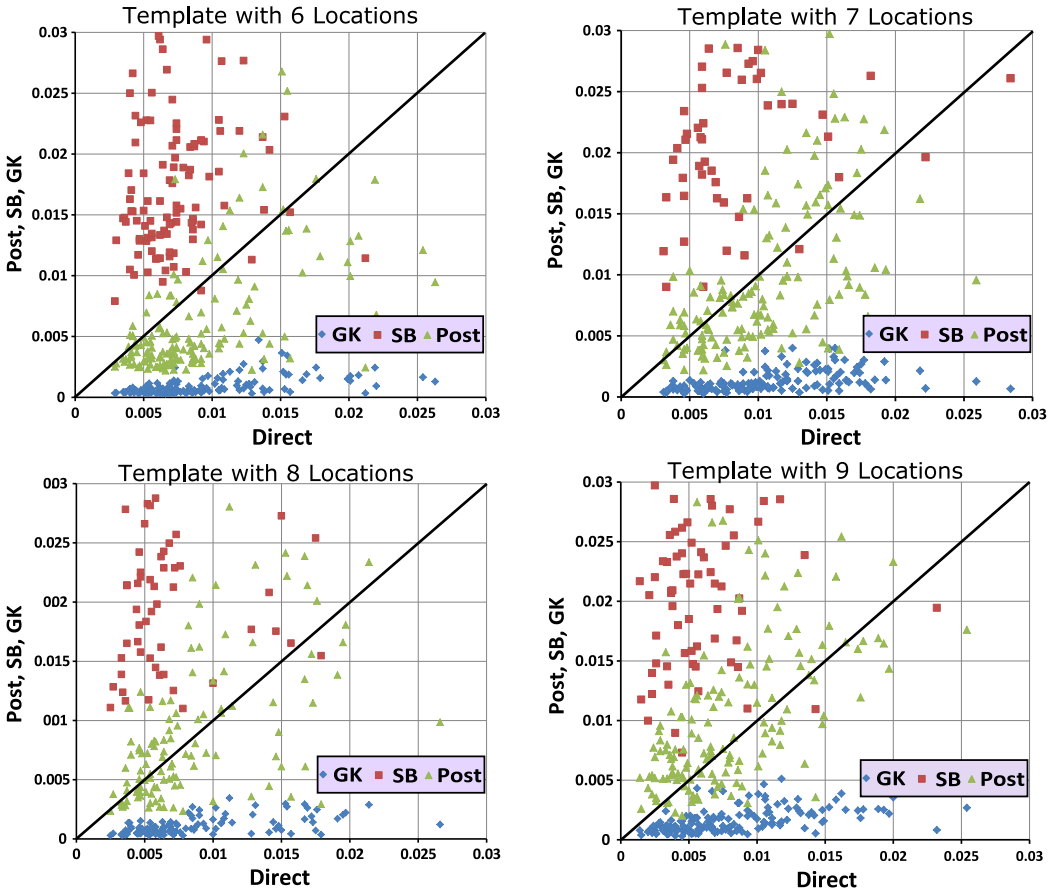

Figure 6.8: The true variance of the mean (Direct) resulted by scanning the image in Figure 6.5 for 150 data configurations of templates in Figure 6.7 versus the variance of the mean achieved by the global kriging (GK), spatial bootstrap (SB) and posterior (Post) approaches.

than other methods. Other images were considered with similar results, that is, GK gives too low uncertainty, SB is too high until conditioned to the data and clipped by reasonable domain limits. The updated or posterior results are reasonable. Figure 6.9 shows a schematic illustration of the variance of the mean calculated by the spatial bootstrap, global kriging and posterior approach versus the true variance.

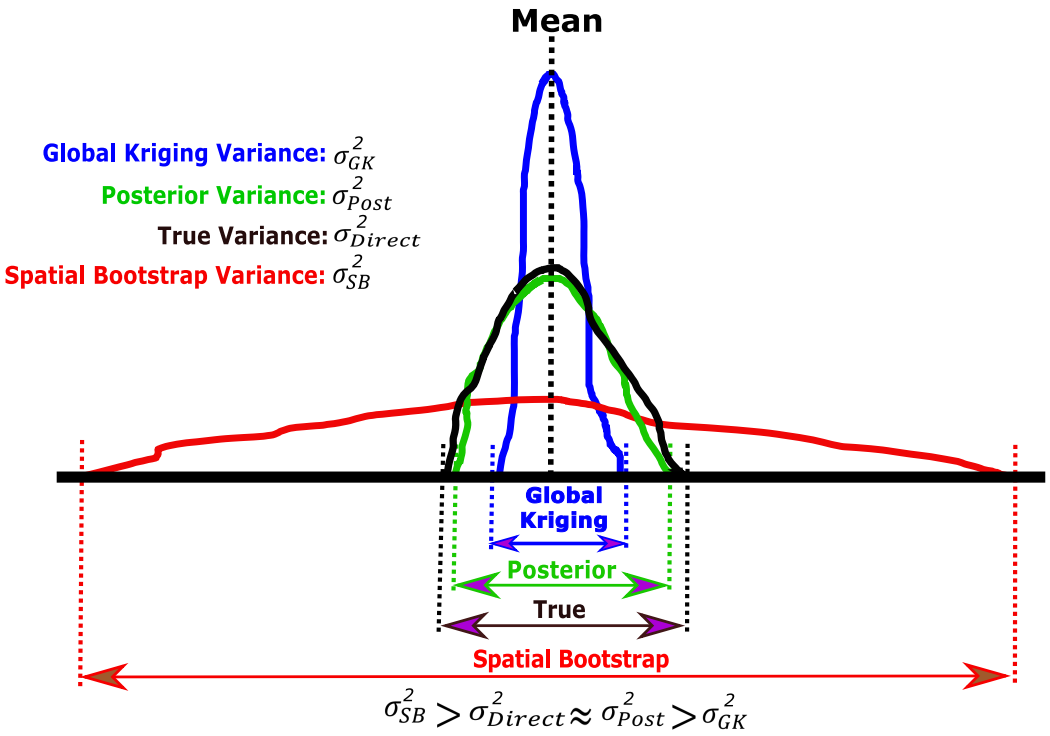

Figure 6.9: Schematic illustration of the variance of the mean by the spatial bootstrap, global kriging and the posterior approach based on Figure 6.8. The posterior approach is the most accurate estimator of the true variance of the mean.

6.4 Multivariate Parameter Uncertainty

The posterior histogram uncertainty for a single variable (univariate parameter uncertainty) has been discussed and showed to be close to the real histogram uncertainty: the spatial bootstrap is the prior histogram uncertainty (upper bound) which can be updated and improved (posterior uncertainty) by conditioning and accounting for the area of interest through transferring in geostatistical simulation modeling workflows. Univariate parameter uncertainty can be extended to multivariate parameter uncertainty. In this case, the joint prior parameter uncertainty is estimated by the multivariate spatial bootstrap and transferred through conventional cosimulation work flow to arrive at a final posterior uncertainty.

The spatial bootstrap for a single variable (Deutsch, 2004; Journel & Bitanov, 2004) could be extended to the multivariate spatial bootstrap to respect the corre-

lations between variables. The multivariate extension of the spatial bootstrap to n variables requires a data-data covariance matrix divided to direct and cross covariance submatrices from the direct and cross variograms between each pair of variables. This covariance matrix is the same as a full cokriging system of equations (left hand side). LU simulation is applied on the covariance matrix and back-transformed to the reference distribution of each variable. The procedure is summarized by:

1. Define stationary direct and cross covariance functions $C_{i,j}(\mathbf{h}), \quad i,j = 1,...n.$ Where n is the number of variables. There are n direct and $\frac{n(n-1)}{2}$ cross covariances.
2. Define the reference distribution of each variable $F_i(z), \quad i = 1,...,n.$
3. Construct the spatial data-to-data covariance matrix $\boldsymbol{C}$ that includes the data-data covariance submatrix of each variable plus the data-data cross covariance submatrix between the variables (full cokriging system of equations). There are totally n^2 submatrices. For k locations of data, each submatrix has a dimension of $k \times k$ (if the locations are identical for variables). Thus, $\boldsymbol{C}$ has a dimension of $(k \times n) \times (k \times n)$.
4. Compute the Cholesky decomposition of the covariance matrix as $\boldsymbol{C} = \mathbf{L}\mathbf{L}^T$.
5. Simulate a vector of uncorrelated standard normal deviate $\mathbf{w}$ with dimension $(k \times n)$.
6. Generate a vector of correlated Gaussian realization for each variable as $\mathbf{y}_{1,...,n} = \mathbf{L}\mathbf{w}$.
7. Transform the unconditional Gaussian values of each variable to original units $z_{1,...,n}(\mathbf{u}_\alpha) = F^{-1}_{1,...,n}(G(\mathbf{y}_{1,...,n}(\mathbf{u}_\alpha))), \quad \alpha = 1,...,k.$ Where $\mathbf{u}_\alpha$ is a data location.

Steps 5 to 7 are repeated for the multivariate spatial bootstrap that provides the prior histogram uncertainty for each variable.

The multivariate posterior uncertainty can be achieved by incorporating the multivariate prior uncertainty in geostatistical simulation: cosimulation by SGS could be

applied through full cokriging or intrinsic collocated cokriging (ICC) using super secondary variable (Babak & Deutsch, 2007a, 2007b; Manchuk & Deutsch, 2012). The multivariate spatial bootstrap realization is used as a reference distribution for normal score transformation of data and back transforming the Gaussian realization of each variable to original units. The multivariate spatial bootstrap realizations lead to uncertainty in the correlation between variables. By incorporating the multivariate spatial bootstrap in geostatistical simulation, uncertainty in correlation is also incorporated in the final model.

6.5 Realistic Example: Amoco Data Set

A real case study of 2D Amoco data (Chu et al., 1994) is considered for multivariate geostatistical modeling with histogram uncertainty (variogram uncertainty of this data set was considered in Chapters 3 and 4). SGS with full cokriging is used (Manchuk & Deutsch, 2012) for multivariate geostatistical modeling. The variables are averaged porosity and permeability over the main reservoir layer. Permeability is in milliDarcies. Figure 6.10 shows a location map (left Figure) and declustered histogram (right Figure) of porosity (Figure 6.10-a), permeability (Figure 6.10-b) and cross plots between these variables in original (left Figure) and normal score (right Figure) (Figure 6.10-c). There is a high correlation ($\rho(0) \approx 0.84$). Cell declustering provides a more reliable reference distribution (Pyrcz & Deutsch, 2014).

Since full cokriging is applied for multivariate geostatistical modeling, linear model of coregionalization (LMC) is fit to direct and cross variograms of porosity and permeability variables. Figure 6.11 shows direct and cross variogram models fitted to the directional experimental variograms for azimuths $0^\circ \pm 25^\circ$ and $90^\circ \pm 25^\circ$. The diagonal variograms are direct variograms and off diagonal is the cross variogram between porosity and permeability. The experimental variograms are calculated for normal score data. The LMC models the direct and cross variograms as:

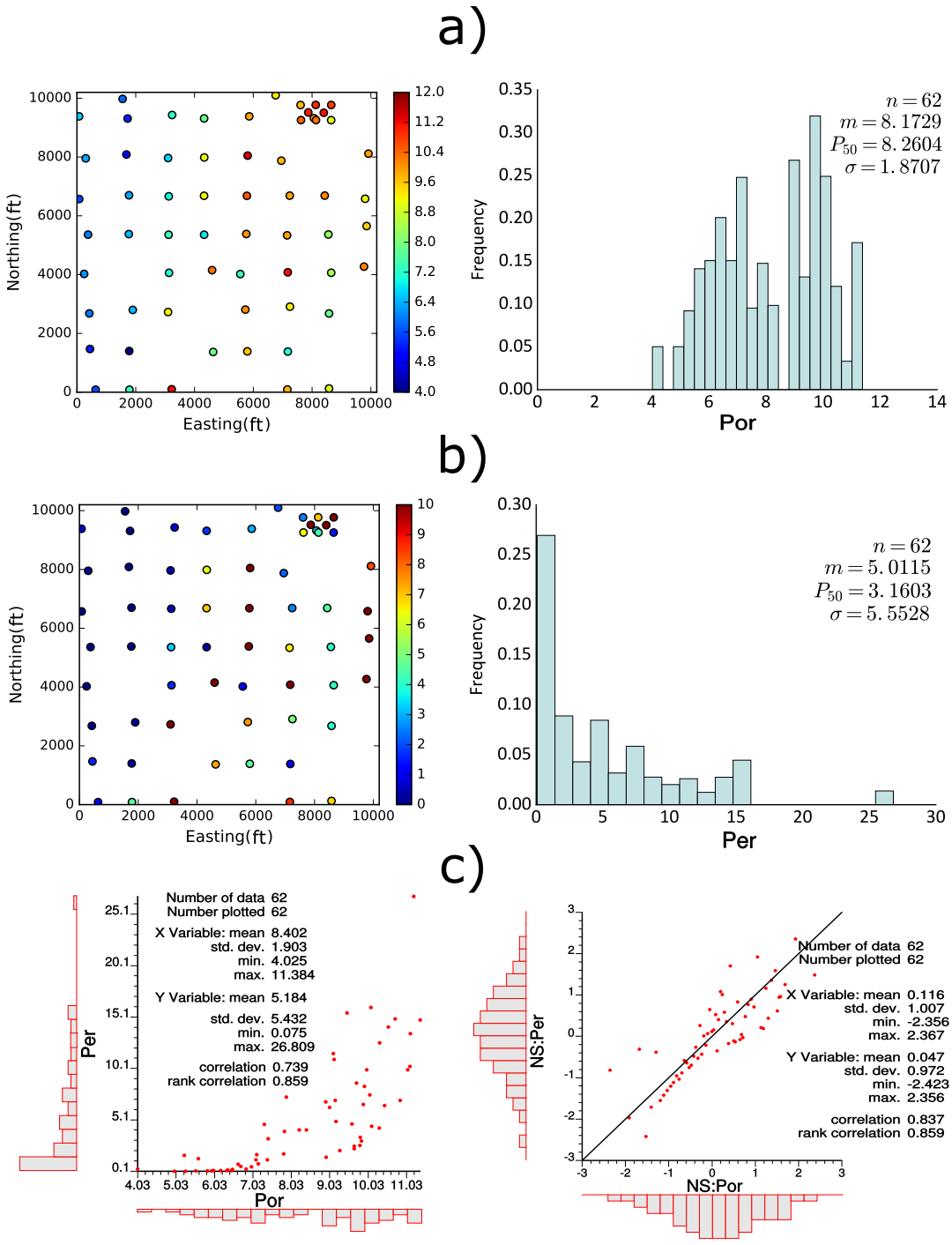

Figure 6.10: Location map (left Figure) and declustered histogram of porosity (right Figure) (a), permeability (b), and cross plots in original (left Figure) and normal score (right Figure) units (c) for Amoco data set.

$$LMC \begin{cases} \gamma_{Por}(\mathbf{h}) = 0.10.Exp_{\substack{h_{major}=2900ft \\ h_{minor}=1000ft}}(\mathbf{h}) + 0.90.Sph_{\substack{h_{major}=17000ft \\ h_{minor}=5000ft}}(\mathbf{h}) \\ \gamma_{Per}(\mathbf{h}) = 0.30.Exp_{\substack{h_{major}=2900ft \\ h_{minor}=1000ft}}(\mathbf{h}) + 0.70.Sph_{\substack{h_{major}=17000ft \\ h_{minor}=5000ft}}(\mathbf{h}) \\ \gamma_{Por,Per}(\mathbf{h}) = 0.17.Exp_{\substack{h_{major}=2900ft \\ h_{minor}=1000ft}}(\mathbf{h}) + 0.67.Sph_{\substack{h_{major}=17000ft \\ h_{minor}=5000ft}}(\mathbf{h}) \end{cases} \quad (6.3)$$

where Por and Per are porosity and permeability variables. Exp and Sph are exponential and spherical variogram models. h_{major} and h_{minor} are major (azimuth 0°) and minor (azimuth 90°) variogram ranges for each variogram structure.

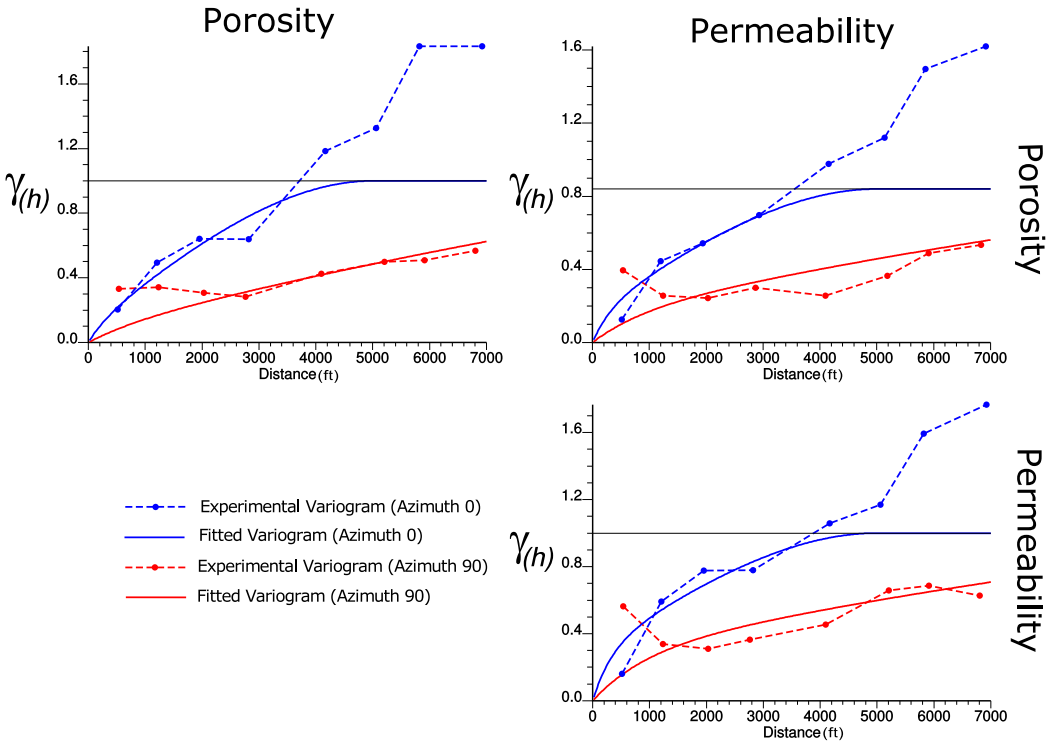

Figure 6.11: Direct and cross variogram models fitted to the directional experimental variograms of porosity and permeability for azimuths $0° \pm 25°$ and $90° \pm 25°$. The experimental variograms are calculated for normal score data.

The multivariate spatial bootstrap is applied with the LMC in Equation 6.3. The joint prior parameter uncertainty for porosity and permeability are calculated for 100 realizations. Figure 6.12-a shows crossplots between porosity and permeability for four realizations of the multivariate spatial bootstrap in original units. Each realization

shows different correlated histogram of variables. Different realizations lead to different correlation coefficients. Figure 6.12-b shows a histogram of 100 correlation coefficients from the multivariate spatial bootstrap in original units. This joint prior histogram uncertainty can be transferred to the final model. Figure 6.13-a and b show the joint prior histogram uncertainty for porosity and permeability, respectively. There is a very high histogram uncertainty for both variables. The standard deviation of the mean for porosity is 0.9145 and for permeability is 2.1538. This high uncertainty is reduced by conditioning data and clipping through geostatistical simulation (posterior uncertainty).

Figure 6.14 shows four cosimulation SGS realizations of porosity and permeability with full cokriging using the histogram uncertainty in Figure 6.13 as reference distributions for normal score transformation of the data and back transform the Gaussian realizations of each variable to original units. The realizations are correlated. The uncertainty in the correlation coefficient (6.12-b) is incorporated in SGS realizations by this approach. Figure 6.15 shows the posterior histogram uncertainty for 100 realizations of porosity (Figure 6.15-a) and permeability (Figure 6.15-b). Due to the trend in the data set (see Figure 6.10), the reference distributions (red lines) are not exactly in the middle of the posterior histogram realizations. Compared with prior histogram uncertainty in Figure 6.13, the uncertainty goes down considerably. The standard deviation of the mean of realizations for porosity and permeability decreases from 0.91 to 0.12, and from 2.15 to 0.48. The standard deviation of the standard deviation of realizations for porosity and permeability also reduces from 0.31 to 0.15, and 1.42 to 0.68. The posterior uncertainty accounts for conditioning, spatial correlation of the data, the area and correlation between variables. This leads to more accurate histogram uncertainty.

6.6 Remarks

The approaches to quantify histogram uncertainty available to the practitioner have drawbacks; overestimating or underestimating the uncertainty. A new framework is

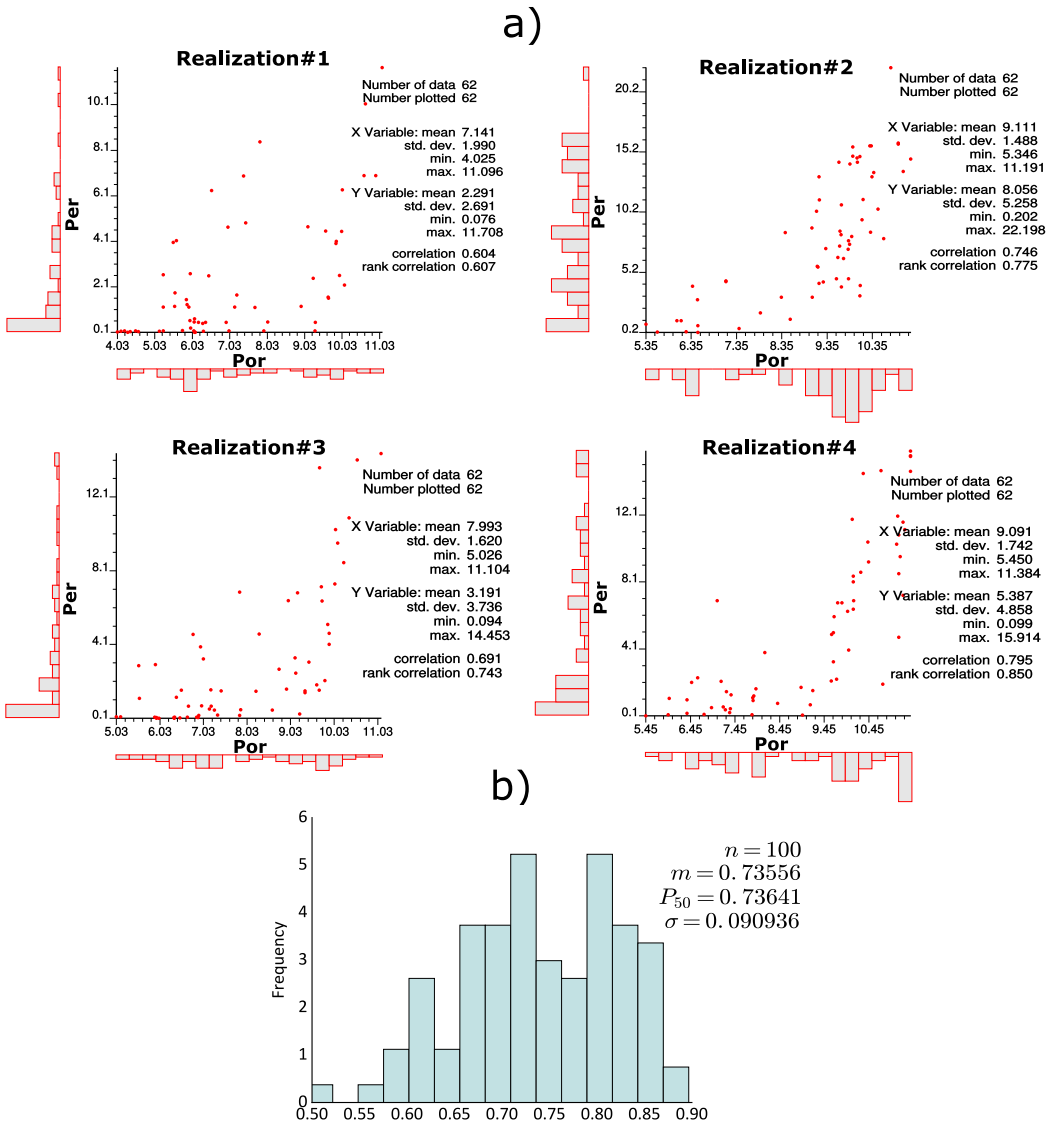

Figure 6.12: a) Crossplots between porosity and permeability for four realizations of the multivariate spatial bootstrap in original units. b) Histogram of 100 correlation coefficients resulted by the multivariate spatial bootstrap in original units.

proposed to transfer prior uncertainty in the histogram through geostatistical simulation to achieve posterior uncertainty. The prior histogram uncertainty is calculated by the spatial bootstrap. The posterior histogram uncertainty is obtained by conditioning the prior histogram uncertainty and accounting for the area to attain lower and more accurate histogram uncertainty: the spatial bootstrap realizations are used as reference distributions. The simulation volume could also be clipped to the volume of interest.

In order to evaluate the posterior approach and the current techniques of quantifying

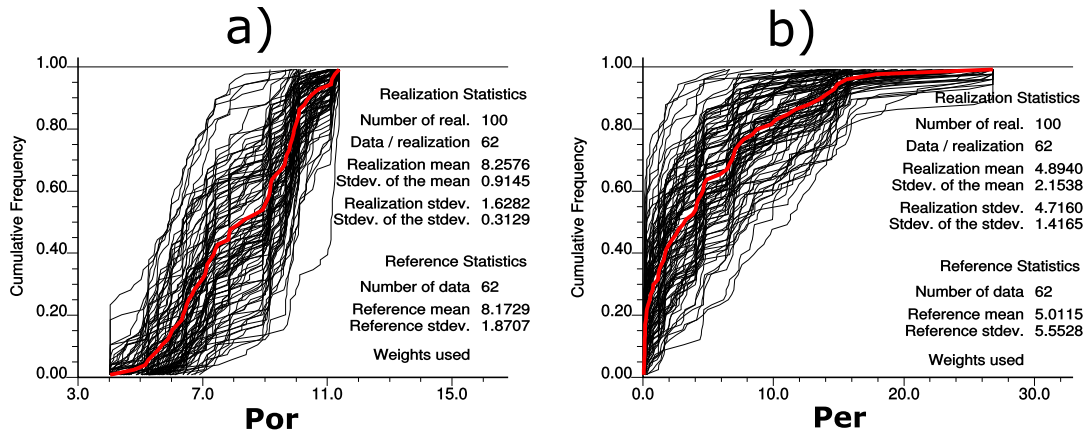

Figure 6.13: The joint prior histogram uncertainty for porosity (a) and permeability (b) (black lines). The red lines are the reference distributions for each variable.

parameter uncertainty such as spatial bootstrap and global kriging, the uncertainty in the mean for a single variable is calculated and compared with possible true uncertainty. A scan-based approach is considered. This approach looks for similar patterns of a data configuration within a large image. The results of the scan-based approach confirms that the posterior approach is a reliable estimator of histogram uncertainty. The global kriging is found to underestimate the uncertainty (lower bound). The spatial bootstrap is found to overestimate the uncertainty (upper bound).

The posterior histogram uncertainty for a single variable is extended to multiple variables by calculating the joint prior parameter uncertainty through multivariate spatial bootstrap. The multivariate spatial bootstrap respects the correlation between the histogram realization of the variables. The uncertainty in the correlation between variables as well as statistics (means and variance) are accounted in the final model. Considering histogram uncertainty plus variogram uncertainty (previous Chapters) leads to more reliable models for decision making.

Only continuous variables (porosity, permeability...) are considered in this Chapter. However, categorical variables such as facies should be incorporated in the final model. The prior histogram uncertainty in the proportion of each facies could be calculated by unconditional simulation at data locations. Conditional realizations with different realizations of prior uncertainty in proportions of facies would provide posterior

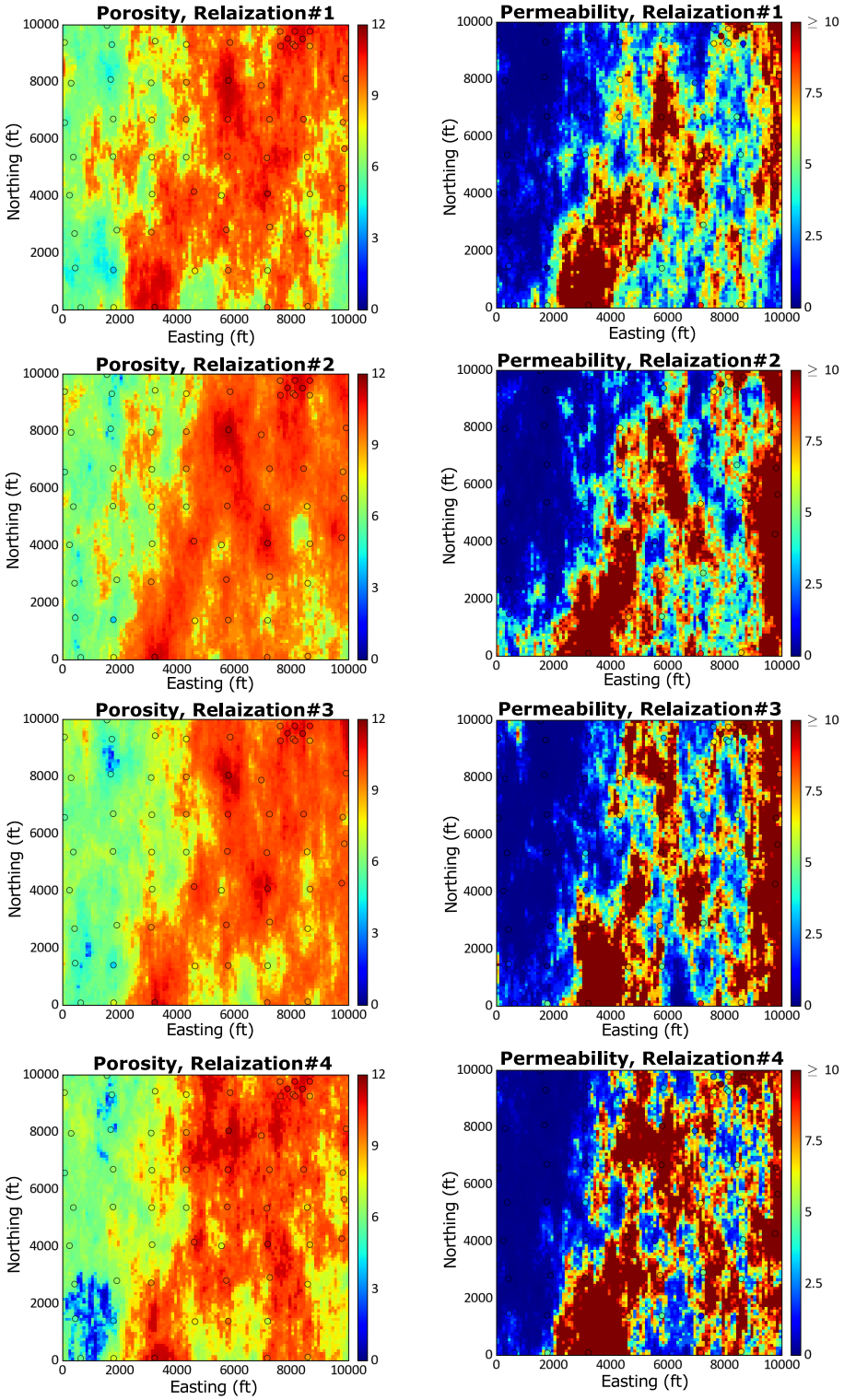

Figure 6.14: Four cosimulation SGS realizations of porosity and permeability with full cokriging using the histogram uncertainty in Figure 6.13 as reference distributions.

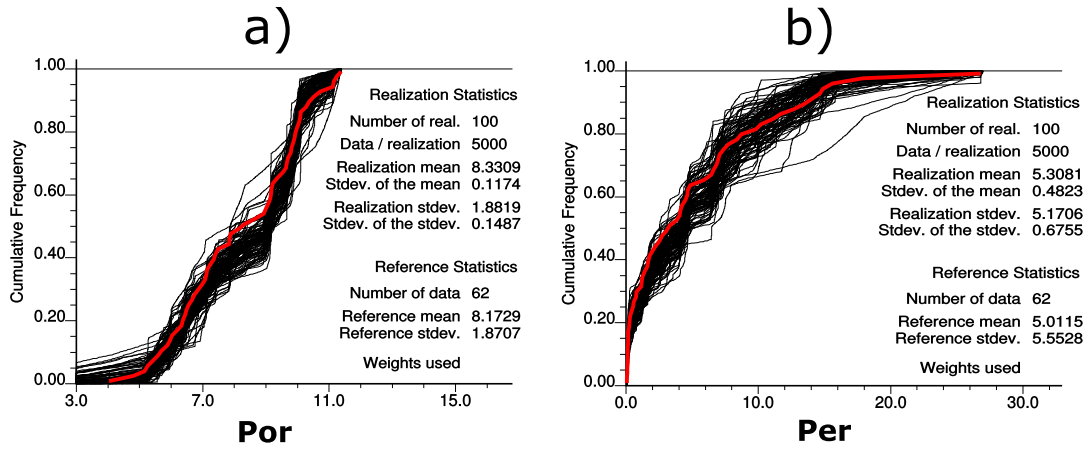

Figure 6.15: Posterior histogram uncertainty for porosity (a) and permeability (b) (black lines). The red lines are the reference distributions for each variable.

uncertainty.

Chapter 7

Case Study - Uncertainty in Hydrocarbon Initially In Place (HIIP)

The approaches of quantifying parameter uncertainty (histogram and variogram) are developed in previous Chapters. This Chapter presents geostatistical reservoir modeling with parameter uncertainty for a real case study with only four widely-spaced wells and 3D seismic data. Uncertainty in gross rock volume (GRV) is achieved by considering uncertainty in the top and thickness of the reservoir using seismic data. Uncertainty in the hydrocarbon water contact (HWC) is also considered. The horizontal variograms from well data, which are unreliable and noisy, are improved by applying constraints from the seismic variogram and the vertical variogram of well data. The joint prior histogram uncertainty is calculated by the multivariate spatial bootstrap. All sources of uncertainty are then incorporated in geostatistical modeling to calculate full uncertainty in hydrocarbon initially in place (HIIP). A sensitivity analysis is considered to define the impact of each source of uncertainty in the final results.

7.1 Netherlands Offshore F3 Block

The case study is "F3" block that is an offshore reservoir located in the Dutch sector of the North Sea. The F3 block is an exploration target for oil and gas made up of Upper-Jurassic to Lower Cretaceous strata. A 3D seismic survey covers the entire

block. There are four vertical wells F02-1, F03-2, F03-4, F06-1 within this block (see Figure 7.1). The 3D seismic data and wells are available in the public domain. The company dGB (dGB, n.d.) used this data to produce a demo survey for OpendTect software (OpendTect, n.d.). Since the base version of OpendTect software is free for academic tasks, all required data for geostatistical modeling in this Chapter such as well data, seismic surfaces, acoustic impedance are retrieved from the demo survey of this software (OpendTect, n.d.) with permission from dGB company.

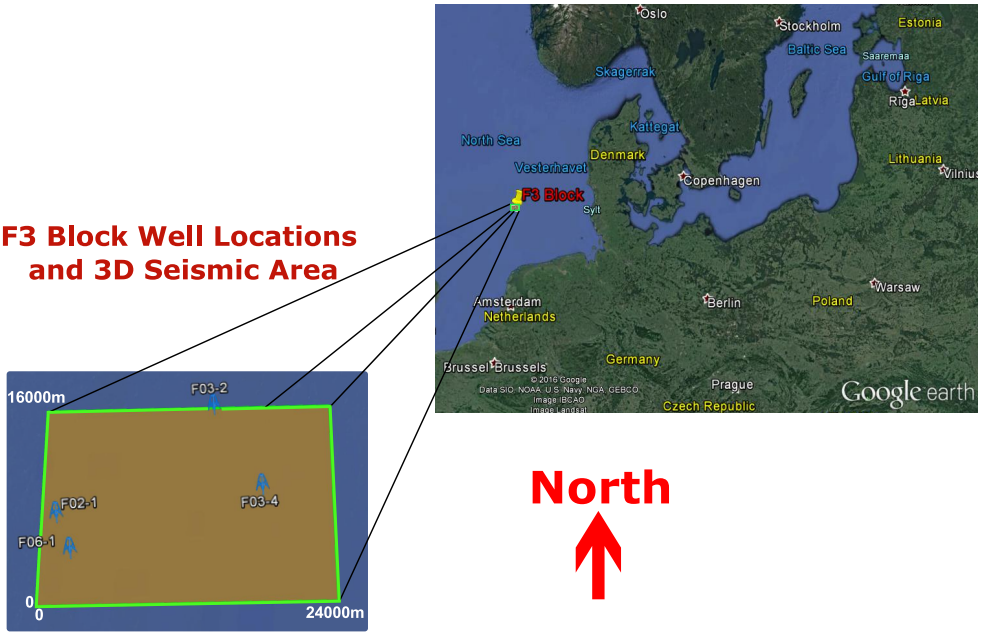

Figure 7.1: 3D seismic area and four vertical wells F02-1, F03-2, F03-4, F06-1 in the F3 block of the Dutch sector of the North Sea (OpendTect, n.d.).

Sonic and gamma-ray logs are available for all wells; however, only the wells F02-1 and F03-2 have density logs. These logs are used to predict density from sonic and gamma-ray logs of the wells F03-4 and F06-1 (OpendTect, n.d.). Total porosity (PHIT) is calculated from density by Equation 7.1:

$$PHIT = \frac{\rho_{ma} - \rho_{log}}{\rho_{ma} - \rho_f} \tag{7.1}$$

where $\rho_{ma} = 2.65 \frac{gr}{cm^3}$ is matrix density, $\rho_f = 1.05 \frac{gr}{cm^3}$ is fluid density and ρ_{log} is density from the log at the point of interest (OpendTect, n.d.). Other required variables for

computing HIIP are calculated from gamma-ray and total porosity logs. Shale volume (V_{sh}) could be calculated by:

$$V_{sh} = \frac{GR_{log} - GR_{min}}{GR_{max} - GR_{min}} \tag{7.2}$$

(Asquith, Krygowski, Henderson, & Hurley, 2004). Where $GR_{min} = 25\,API$ is minimum gamma-ray (log response in clean beds), $GR_{max} = 75\,API$ is maximum gamma-ray (log response in complete shale), and GR_{log} is log response in the zone of interest. Effective porosity (PHIE), which only accounts for the interconnected pores of the formation and contains hydrocarbon or water could be calculated by:

$$PHIE = PHIT - \phi_{sh} \times V_{sh} \tag{7.3}$$

(AlRuwaili & AlWaheed, 2004). Where ϕ_{sh} is the shale-porosity. Since the estimation of ϕ_{sh} is difficult, ϕ_{sh} is usually replaced with PHIT (AlRuwaili & AlWaheed, 2004). Three variables PHIT, PHIE and Vsh calculated at each well are used for geostatistical modeling and estimation of HIIP uncertainty. There is no water saturation data. Vsh is assumed as water saturation due to the fact that shale likely contains more irreducible water saturation. The calculated variables are upscaled to 5m. This is the vertical grid size for geostatistical modeling of this case study. Figure 7.2 shows log tube of the variables PHIT (a), PHIE (b) and Vsh (b) at well locations after upscaling.

The original 3D seismic data, which covers the entire reservoir and four wells, is processed and inverted to acoustic impedance (dGB, n.d.). The seismic survey parameters are: Inline range: from 100 to 750 with Step 1, Crossline range: from 300 to 1250 with Step 1, Bin size: 25m long (Inline) and 25m wide (Crossline), Z range (time): from 0 to 1848 millisecond (ms) with 4ms Step. The acoustic impedance is produced after filtering the original seismic data because it is rather noisy (OpendTect, n.d.). There are several 3D seismic surfaces in the demo survey; Truncation and MFS4 surfaces are assumed as top and bottom of the reservoir containing oil and gas layers although these surfaces may not be the real top and bottom surfaces of the hydrocarbon-bearing

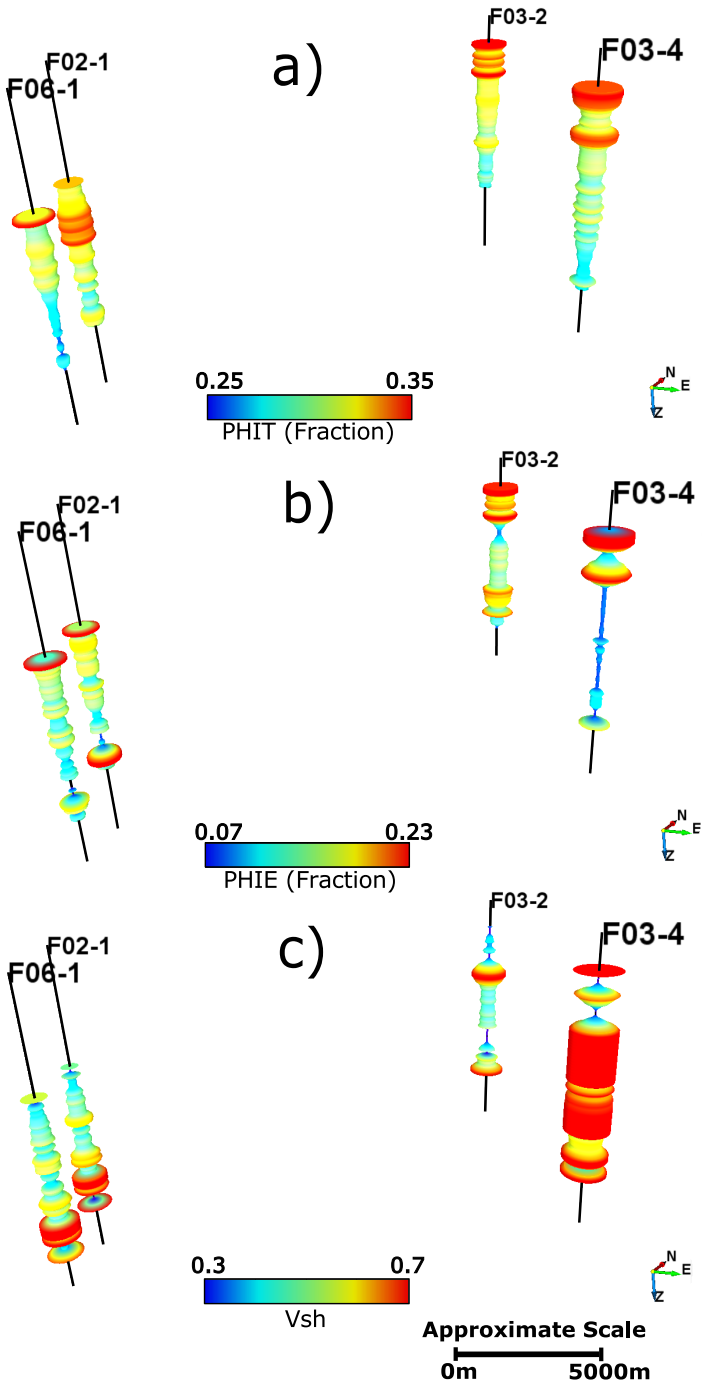

Figure 7.2: Log tube of the variables PHIT (a), PHIE (b) and Vsh (b) for four wells in F3 Block: east 0-24000m and north 0-16000m (see Figure 7.1).

reservoir (there is no information which surfaces are the real top and bottom surfaces in the data set). Figure 7.3 shows Truncation (top) and MFS4 (bottom) surfaces in 3D acoustic impedance in time domain (ms). The acoustic impedance and seismic

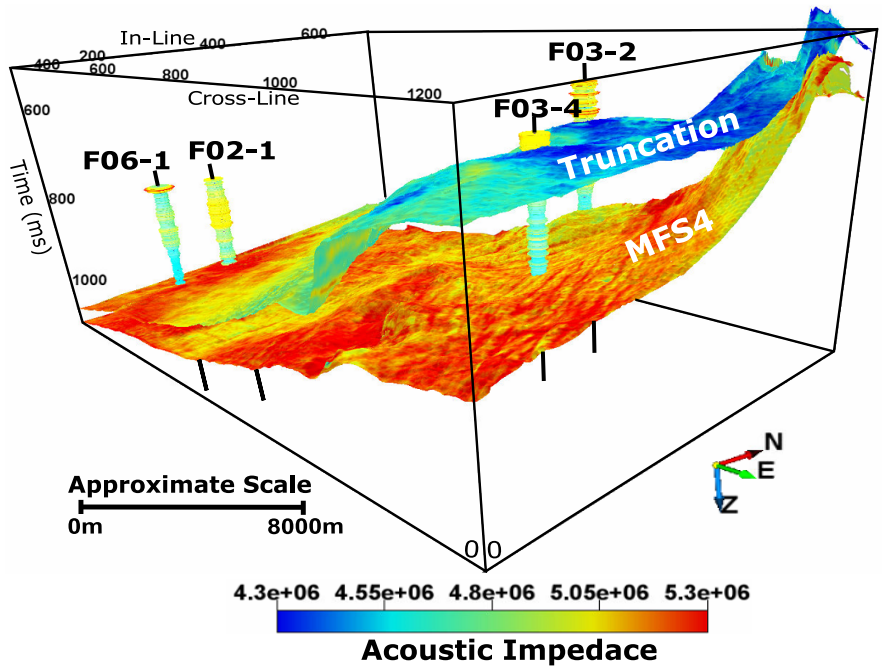

Figure 7.3: Truncation (top) and MFS4 (bottom) seismic surfaces in 3D acoustic impedance in time domain (ms).

surfaces should be converted from time to depth by a velocity model in order to be used as secondary data in geostatistical modeling. The velocity model could be built by time-depth model from checkshot survey, vertical seismic profiles (VSP) in wells, or by seismic refraction operation.

7.2 Gross Rock Volume (GRV) with Parameter Uncertainty

Gross rock volume (GRV) is the volume of reservoir between top and bottom surfaces. These surfaces could be achieved from seismic data: for example, Truncation and MFS4 seismic surfaces in Figure 7.3; however, they are subject to uncertainty away from the well data. This uncertainty should be incorporated in GRV model.

There are two sources of uncertainty in seismic surfaces: 1- Time interpretation uncertainty: this uncertainty happens while picking stratigraphic surfaces on seismic data. The stratigraphic surfaces are usually acquired in depth from well data and converted to time domain by checkshot survey or VSP to be visible in time domain for

seismic data, then the surfaces are followed over the entire 3D seismic area. This is usually done by auto-picking software that are subject to uncertainty. 2- Time-to-depth conversion: seismic surfaces are interpreted in time domain and must be converted to depth by a velocity model. This is also subject to uncertainty. A combination of time interpretation and time to depth uncertainties leads to uncertainty in the seismic surface. Usually seismic surfaces are calibrated with well observations to remove the mismatch. However, valuable hints can be attained from the mismatch between seismic surfaces and actual well observations for quantifying uncertainty in surfaces (Alshehri, 2010).

The conventional approach for modeling GRV is to model top surface and thickness; the thickness at each location is added to the top surface to build bottom surface. This naturally accounts for the correlation between the two surfaces and avoids any artifact crossing of surfaces. The volume of the reservoir above hydrocarbon (oil or gas) water contact (HWC) is considered for GRV.

7.2.1 Tie Seismic to Well

The mismatch between seismic surface (top and thickness) and actual well observations should be corrected although it is used for quantifying parameter uncertainty. Figure 7.4 shows a 2D map view of seismic top surface (Truncation) after converting from time to depth by velocity model achieved from check shot survey of wells. The values are depths (in meter) below sea level; low and high values signify closer and farther depths from the sea level, respectively. There is a mismatch $\Delta = actual - seismic$ at each well location. The Δ is negative for wells F06-1 ($\Delta = -47.6m$) and F02-1 ($\Delta = -25.6m$), and positive for F03-2 ($\Delta = 1.4m$) and F03-4 ($\Delta = 18.9m$). Figure 7.5 shows 2D map view of thickness resulted by difference between Truncation and MFS4 surfaces. The thickness from seismic surfaces in time domain is converted to depth by velocity model of the wells (check shot survey). There is a mismatch $\Delta = actual - seismic$ at each well location. The Δ is negative for wells F06-1 ($\Delta = -13.1m$) and F02-1 ($\Delta = -0.2m$), and positive for F03-2 ($\Delta = 17.5m$) and F03-4 ($\Delta = 13.6m$).

The seismic surface (in depth domain) could be tied to the wells by simple kriging

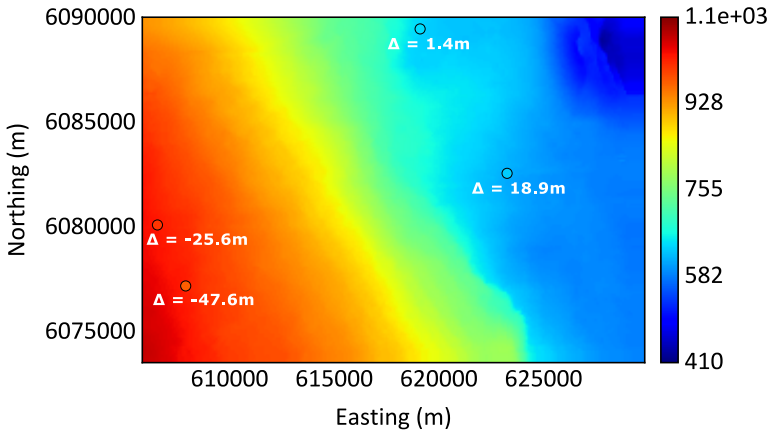

Figure 7.4: 2D map view of seismic top surface (Truncation) after converting from time to depth. The values are in meter and signify depths below see level. The mismatch Δ is negative for wells F06-1 ($\Delta = -47.6m$) and F02-1 ($\Delta = -25.6m$) and positive for F03-2 ($\Delta = 1.4m$) and F03-4 ($\Delta = 18.9m$).

(SK) of Δ at each well. The variogram model of Δ is required for kriging. There are only four conditioning Δ values: thus, the experimental variogram of Δ for top and thickness can not be calculated. It is proposed to attain the approximate variogram of Δ by the variogram of the detrended seismic. The detrended model is calculated by the difference between seismic and trend model $detrend = seismic - trend$. It is required to model the trend of top surface and thickness from seismic data. Radial basis function is used for computing the trend models (see Qu and Deutsch (2014) for more information). The variogram of detrended top and thickness may be assumed as an approximate variogram of Δ.

Figure 7.6-a shows the trend model of top surface (Figure 7.4), and Figure 7.6-b shows the detrended model. Figure 7.7 shows the experimental variograms and fitted models of the detrended model (Figure 7.6-b) for azimuths $-20°$ (major direction of continuity) and azimuth $70°$ (minor direction of continuity). The 2D variogram model has one Gaussian structure with maximum range of 5800m and minimum range of 4500m and very small nugget effect (0.000001) to avoid precision problems in simulation.

Figure 7.8-a shows the trend model of thickness (Figure 7.5), and Figure 7.8-b shows the detrended model. Figure 7.9 shows the experimental variograms and fitted

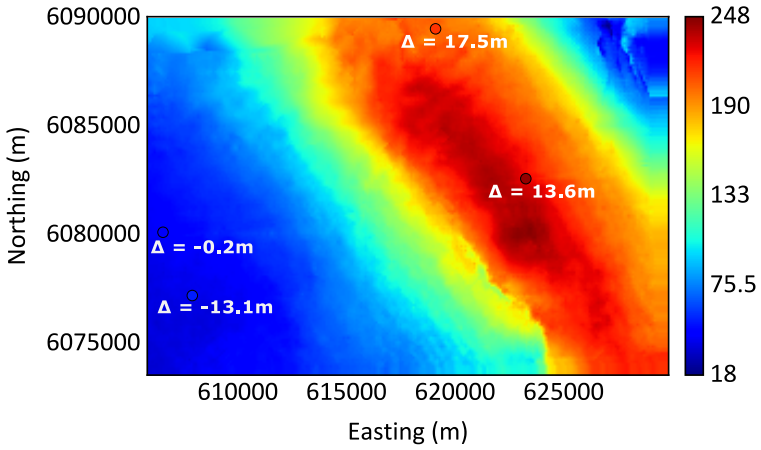

Figure 7.5: 2D map view of thickness by difference between Truncation and MFS4 surfaces and converting to depth. The values are in meter. The mismatch Δ is negative for wells F06-1 ($\Delta = -13.1m$) and F02-1 ($\Delta = -0.2m$) and positive for F03-2 ($\Delta = 17.5m$) and F03-4 ($\Delta = 13.6m$).

models of the detrend model (Figure 7.8-b) for azimuths $-20°$ (major direction of continuity) and azimuth $70°$ (minor direction of continuity). The 2D variogram model has two Gaussian structures with maximum and minimum range of 1500m for the first structure, and maximum and minimum range of 8000m and 6000m for the second structure and very small nugget effect (0.000001).

SK of Δ by the variogram of detrended model (Figures 7.7 and 7.9) and adding to the seismic surface leads to tie seismic to wells. This leads to $\Delta = 0$ at well locations. The mean value for simple kriging estimator should be set to zero due to returning the seismic surface for locations far away from data where the estimate is the mean. Figure 7.10 shows a 2D map view of tied seismic top surface to well. The mismatch Δ between well and seismic is zero at each well location. Figure 7.11 shows a 3D map view of tied seismic top surface to well. Figure 7.12 shows a 2D map view of tied seismic thickness to well. Similar to the top surface, the mismatch Δ between well and seismic is zero at each well. Figure 7.13 shows a 3D map view of bottom surface by adding the tied thickness (Figure 7.12) to the tied top surface (Figure 7.10).

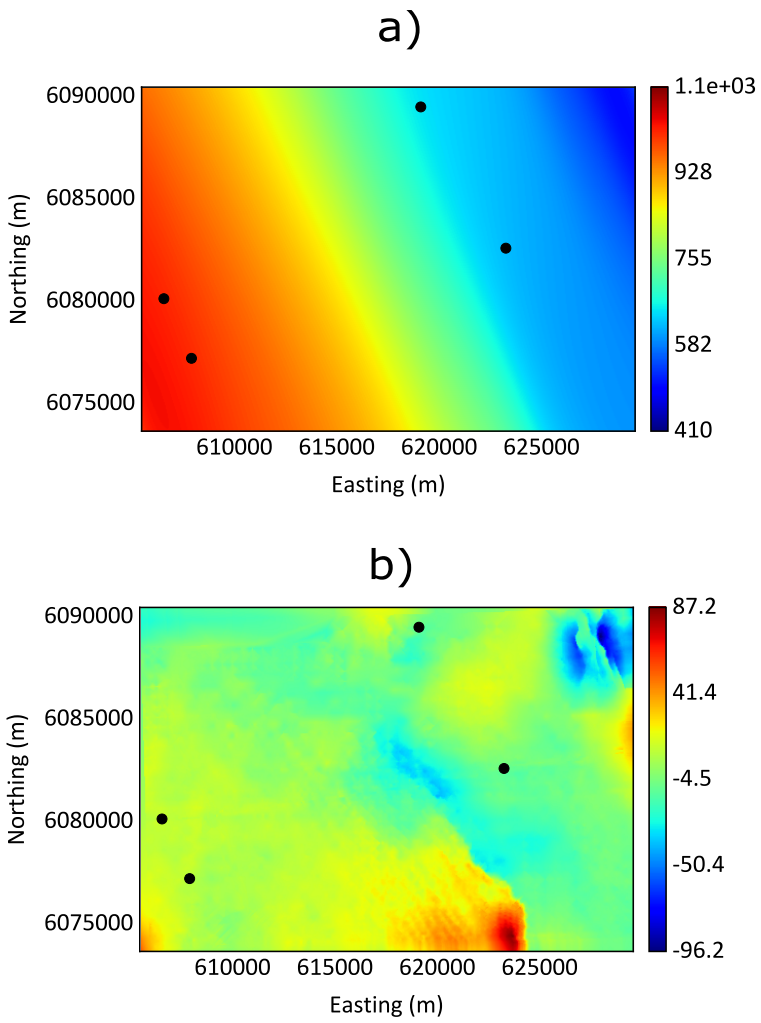

Figure 7.6: a) 2D map view of trend modeling of top surface (Figure 7.4). b) The detrended model by difference between seismic (Figure 7.4) and trend model (a). The values are in meter.

7.2.2 Uncertainty in GRV

Conditional sequential Gaussian simulation (SGS) is used to assess uncertainty in GRV. Uncertainty in GRV can be calculated by quantifying uncertainty in the reference top surface and thickness: the top surface and thickness from seismic interpretation after tie to well are considered as reference structures (Figure 7.10 and 7.12). The realizations of the top and thickness should have zero deviation from the reference structures at well locations, and deviation should increase for locations far away from the wells. The deviation is assumed to be a Gaussian distribution and could be simulated by con-

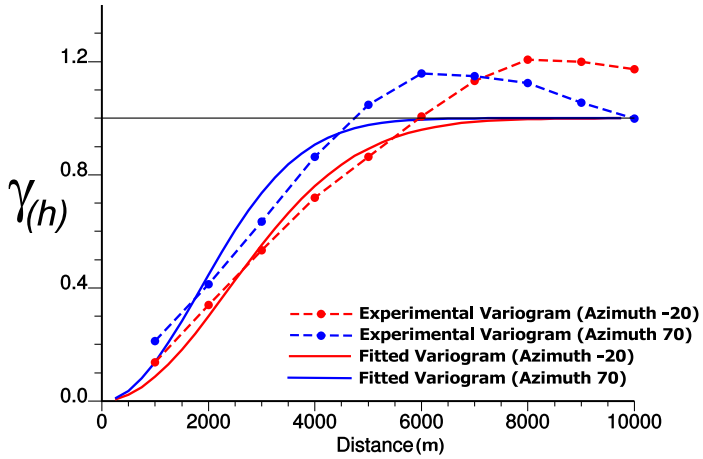

Figure 7.7: The experimental variograms and fitted models of the detrended model (Figure 7.6-b) for azimuths $-20°$ (major direction of continuity) and azimuth $70°$ (minor direction of continuity).

ditional SGS with zero value at well locations. SGS realizations are non-standardized by multiplying them with some standard deviation that could be calculated from the standard deviation of the mismatch between well and seismic (σ_Δ) (standard deviations of Δ in Figures 7.4 and 7.5):

$$\sigma_\Delta = \sqrt{\sigma_{TP}^2 + \sigma_{TtD}^2} \tag{7.4}$$

(Alshehri, 2010). Where σ_{TP}^2 is the variance of time picking, and σ_{TtD}^2 is time-to-depth variance; the variance of Δ comes from both time picking and time to depth conversions based on two assumptions: 1- The distributions of TP and TtD are Gaussian, 2- The errors in TP and TtD are independent from each others (Alshehri, 2010). By adding the deviation (non-standardized SGS realization with σ_Δ) to the reference structure (top and thickness), the uncertainty in GRV is calculated. This technique calculates uncertainty in GRV without considering parameter uncertainty (Alshehri, 2010).

Parameter uncertainty could be incorporated in GRV by considering variogram and histogram uncertainty of Δ for the top and thickness: prior uncertainty in the experimental variogram and histogram of Δ are calculated (Chapters 3 and 6) and incorporated in conditional SGS with zero value at well locations. The posterior uncertainty in Δ is then added to the reference structure (top or thickness). This process should

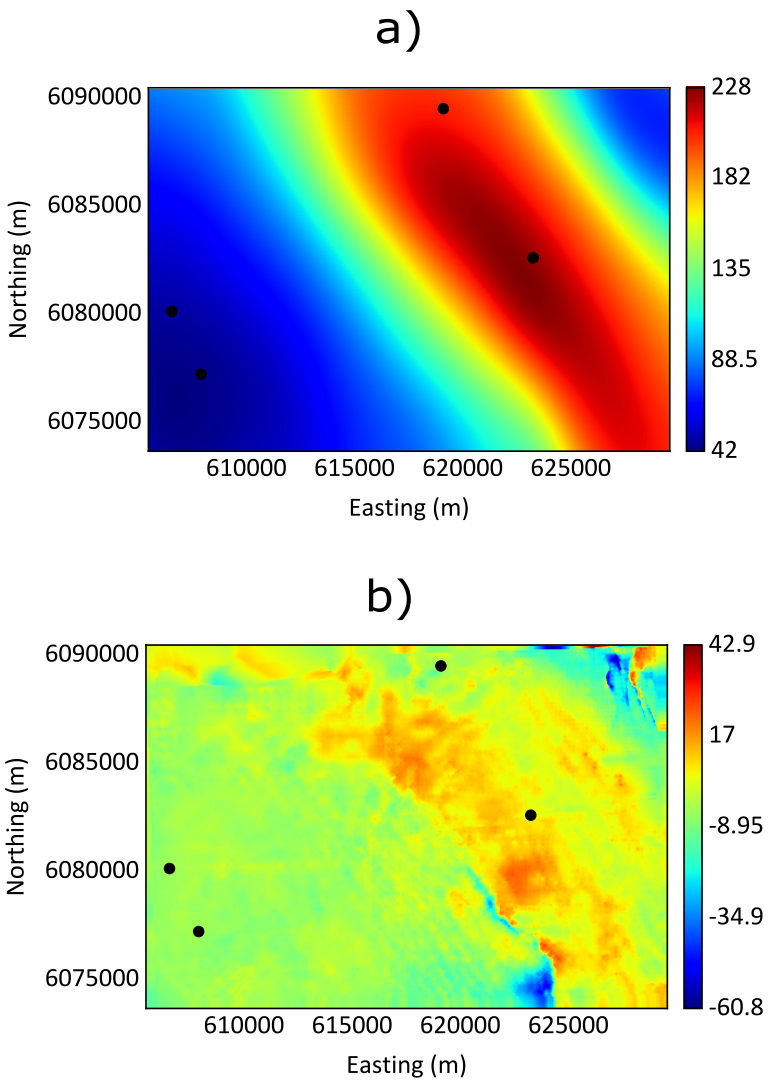

Figure 7.8: a) 2D map view of trend modeling of thickness (Figure 7.5). b) The detrended model by difference between seismic (Figure 7.5) and trend model (a). The values are in meter.

be applied for the top and thickness independently. This approach leads to honor well observations of the top and thickness for all SGS realizations. The variogram uncertainty cannot be calculated due to only four wells: the variograms of detrended seismic are assumed as fixed reference variogram models for top and thickness. Moreover, histogram uncertainty by the spatial bootstrap may not be applicable for this data due to unreliable reference distribution of Δ with only four values. A parametric approach is proposed to attain histogram uncertainty (only uncertainty in the mean) of Δ. The Steps of this technique are as follow:

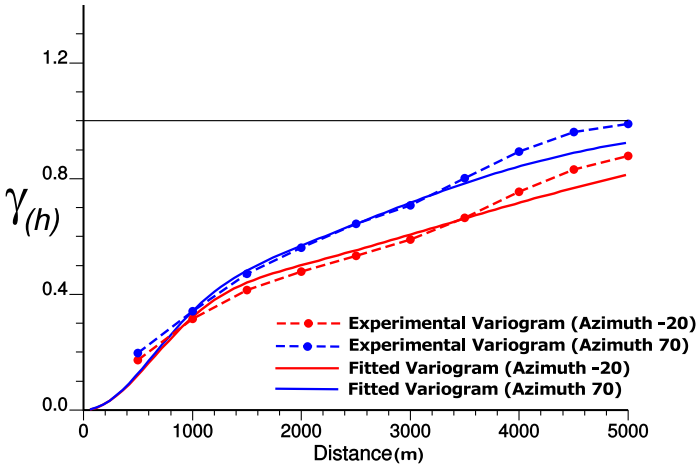

Figure 7.9: The experimental variograms and fitted models of the detrended model (Figure 7.8-b) for azimuths $-20°$ (major direction of continuity) and azimuth $70°$ (minor direction of continuity).

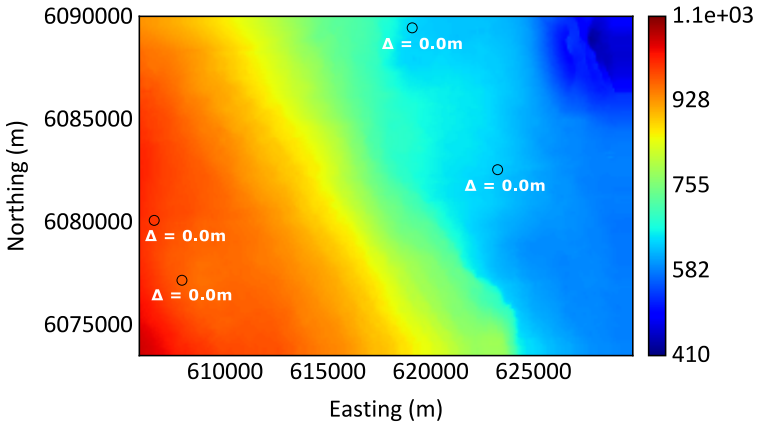

Figure 7.10: 2D map view of tied seismic top surface (Truncation) to well observation. The mismatch Δ between well and seismic is zero at each well location. The values are in meter (depth below sea level).

1. Build a Gaussian distribution for the mean of Δ. This Gaussian distribution $\overline{\Delta}$ has a mean of zero and variance of $\frac{\sigma^2_\Delta}{n}$, where n is number of independent data; this is consistent with the spatial bootstrap of Δ for quantifying the uncertainty in the mean. It is assumed four well data are independent from each other ($n = 4$).

2. Draw mean values from $\overline{\Delta}$ distribution: $\overline{\Delta}^{\,l}$, $l = 1, ..., L$, where L is number of drawn mean values.

3. Attain conditioning well data in the Gaussian space with zero value in original units for each realization of the mean ($\overline{\Delta}^{\,l}$). SGS should be applied with zero

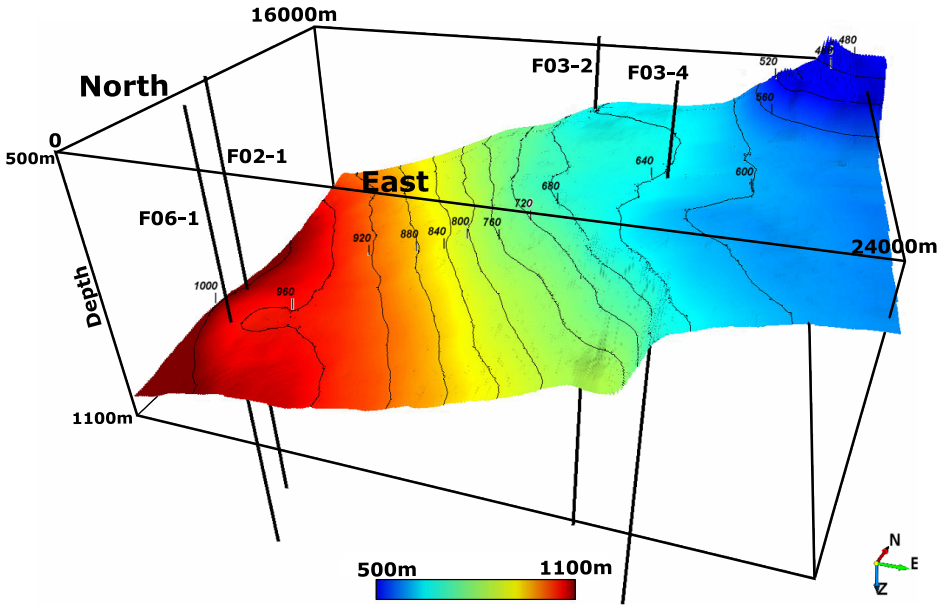

Figure 7.11: 3D map view of tied seismic top surface (Truncation) to well observation. The values are in meter (depth below sea level).

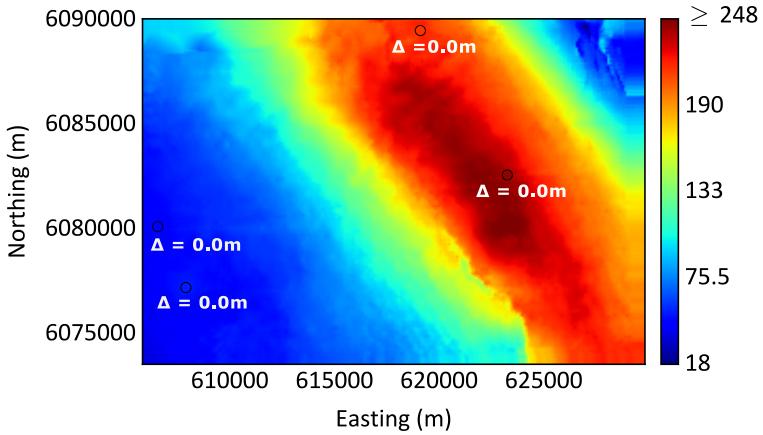

Figure 7.12: 2D map view of tied seismic thickness to well observation. The values are in meter. The mismatch Δ between well and seismic is zero at each well location.

value at well locations. The zero value should be transformed to the Gaussian space based on a reference distribution, and back-transformed to the original units after simulation (zero value at each well location after back transformation). Since the mean of each histogram realization of Δ (Δ is assumed Gaussian) is calculated ($\overline{\Delta}^{\,l}$), and the variance of each realization is σ^2_Δ (assumed constant),

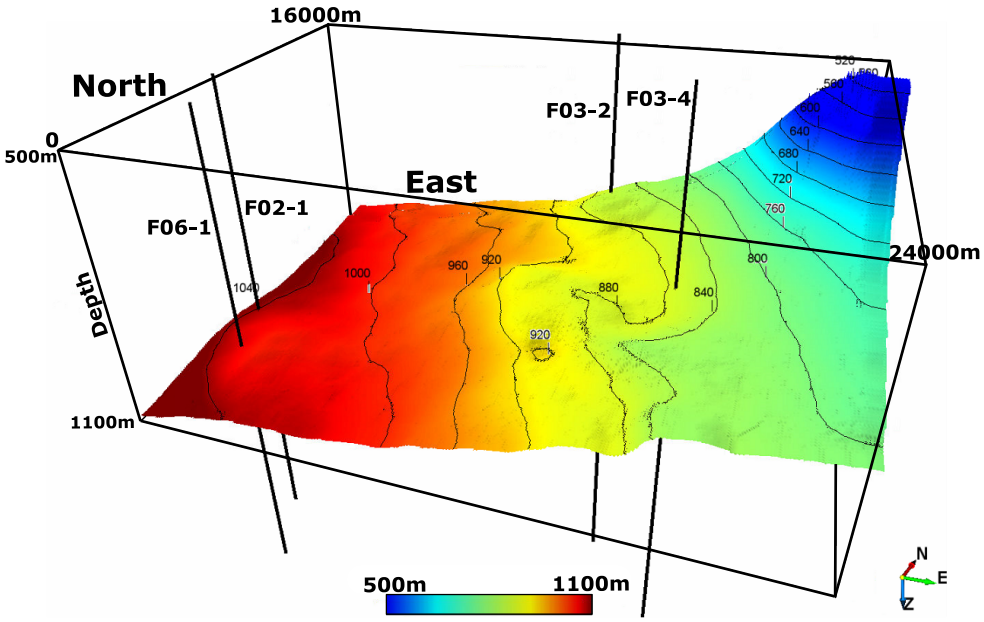

Figure 7.13: 3D map view of tied seismic bottom surface to well observation by adding the tied thickness to the tied top surface. The values are in meter (depth below sea level).

zero value at well locations can be transformed to the Gaussian space by:

$$y^l = \frac{z - m}{\sigma} = \frac{-\overline{\Delta}^{\,l}}{\sigma_\Delta} \qquad , \qquad l = 1, ..., L \tag{7.5}$$

where $z = 0$, which is the conditioning data for each well location in the original space. y^l is the conditioning data for each SGS realizations in the Gaussian space. For example, four well locations have Gaussian conditional value y^l for l^{th} SGS realization.

4. The Gaussian SGS realizations are back-transformed to the original units and added to the reference structure (top or thickness):

$$s^l = (y_{SGS} \times \sigma_\Delta + \overline{\Delta}^{\,l}) + s_{ref} \qquad , \qquad l = 1, ..., L \tag{7.6}$$

where y_{SGS} is the Gaussian SGS realization, s_{ref} is the reference structure (top or thickness), and s^l is the final realization of top surface or thickness accounting for parameter uncertainty. Since SGS realization has zero value at well locations

after beck transformation to the original space, the final realization s^l has the deviation of zero from the reference structure at well locations.

Step 1 to 5 are repeated for top and thickness separately to attain parameter uncertainty in GRV. Realizations of thickness are added to the realizations of the top surface to achieve bottom surface realizations. This methodology ensures that all realizations honor Truncation and MFS4 surfaces of four wells. Figure 7.14 and 7.15 show 20 realizations of top (Truncation) and bottom surfaces (MFS4). As shown for well F06-1, the uncertainty at the well locations is zero.

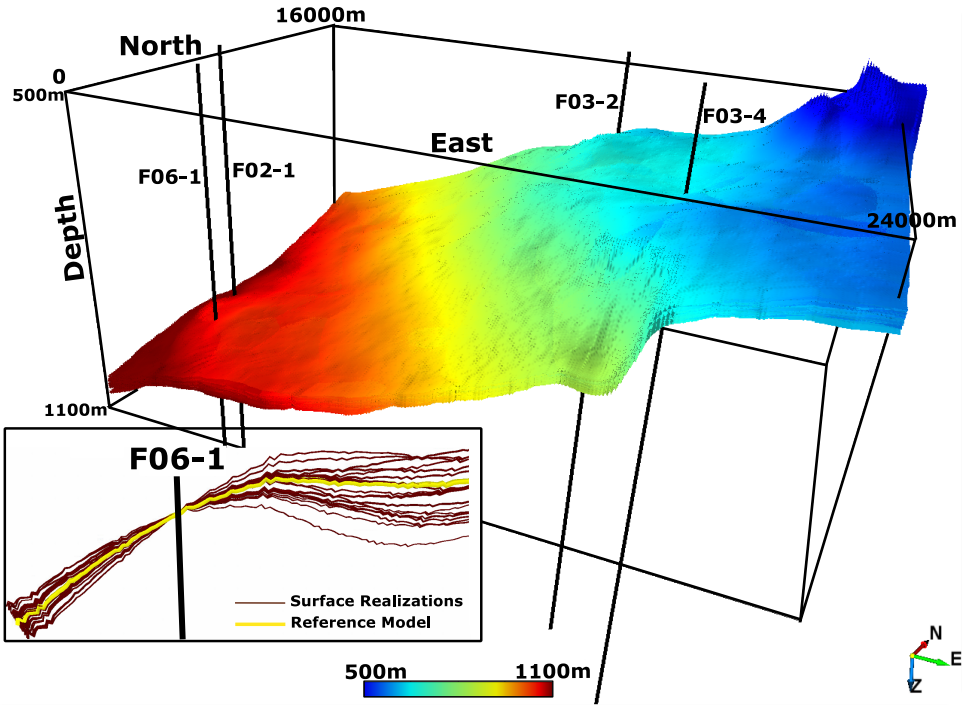

Figure 7.14: 3D map view of 20 realizations of top surface (Truncation). The values are in meter (depth below sea level). All surface realizations are consistent with the reference model (Truncation) at well locations (zero uncertainty).

The reservoir is not only bounded by the top and bottom surfaces, but also it must be above hydrocarbon (oil or gas) water contact (HWC). The volume above HWC surface contributes to calculate HIIP because there may be only water below HWC. The depth of HWC could be obtained from the available wells (well logs), or it can be assumed at the lowest known hydrocarbon level if it is not detected. HWC may also be

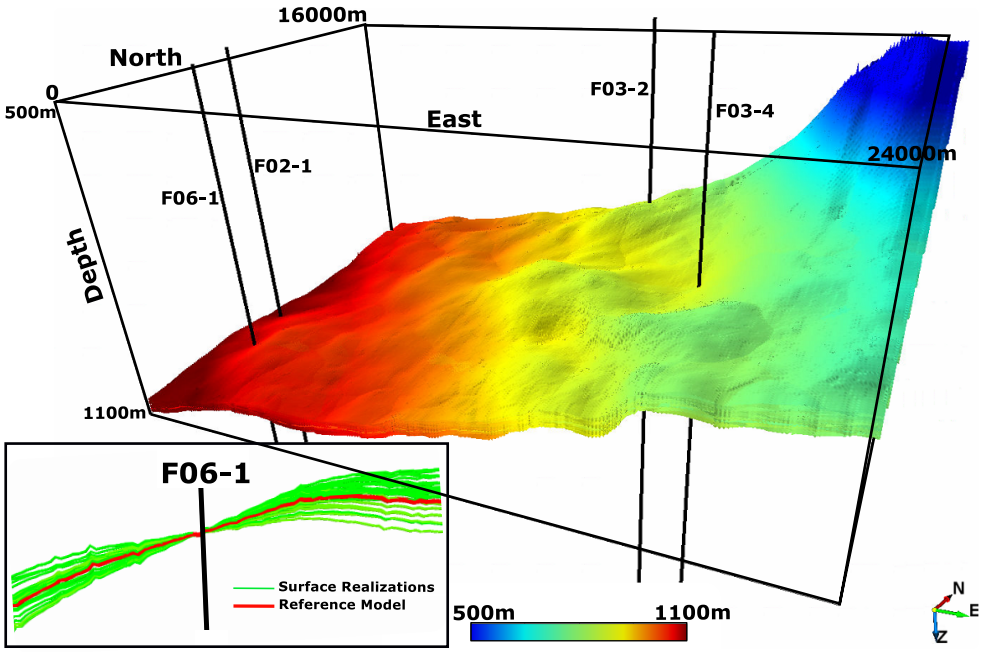

Figure 7.15: 3D map view of 20 realizations of bottom surface (MFS4). The values are in meter (depth below sea level). All surface realizations are consistent with the reference model (MFS4) at well locations (zero uncertainty).

uncertain. Monte Carlo simulation (MCS) is widely used to quantify this uncertainty by assuming a triangular or uniform distribution for HWC (see Alshehri (2010) for more information). There is no information of HWC for this case study; thus, HWC is assumed to be at 870m. The uncertainty in HWC is calculated by building a uniform distribution with minimum value of HWC-10m and maximum value of HWC+10m. Figure 7.16 shows the uncertainty at the top and bottom surfaces for 20 realizations and the uncertainty in HWC. Monte Carlo simulation is applied to draw realizations from the uniform distribution of HWC between 860m to 880m. These realizations plus the surface realizations (top and bottom) are used to calculate the uncertainty in GRV. Figure 7.17 shows histogram of uncertainty in GRV by 100 realizations in billion cubic meter (BCM). The volume of the reservoir is high and seems unrealistic in comparison with the conventional hydrocarbon reservoir. This high volume is due to high reservoir thickness in north and south east (see Figures 7.12 and Figure 7.16). Since there is no information of the real top and bottom surface for this reservoir, Truncation and

MFS4 surfaces are assumed as the top and bottom surfaces of the hydrocarbon-bearing reservoir, which they might not be the real top and bottom surfaces.

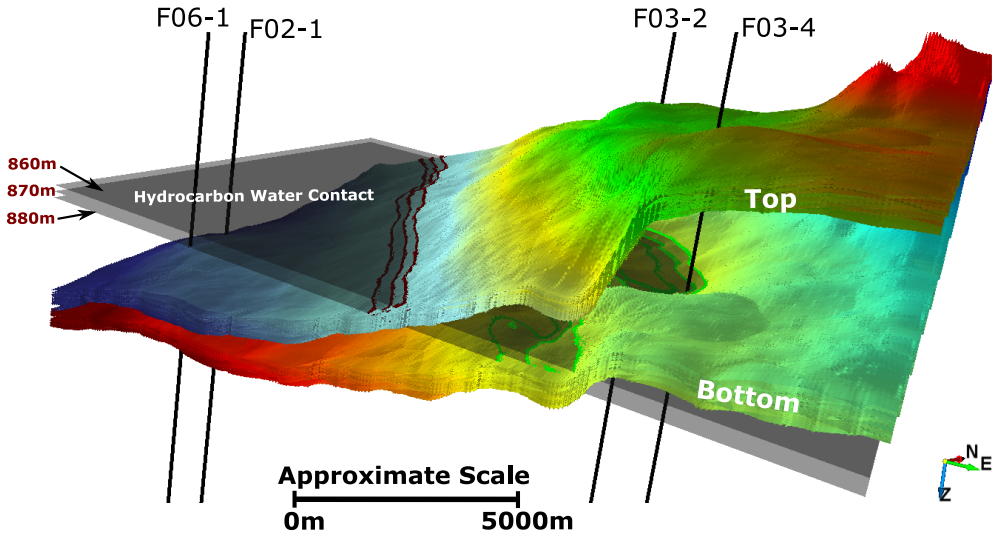

Figure 7.16: Uncertainty for the top (Truncation) and bottom (MFS4) surfaces for 20 realizations and the uncertainty in HWC=870m, HWC$\pm 10m$.

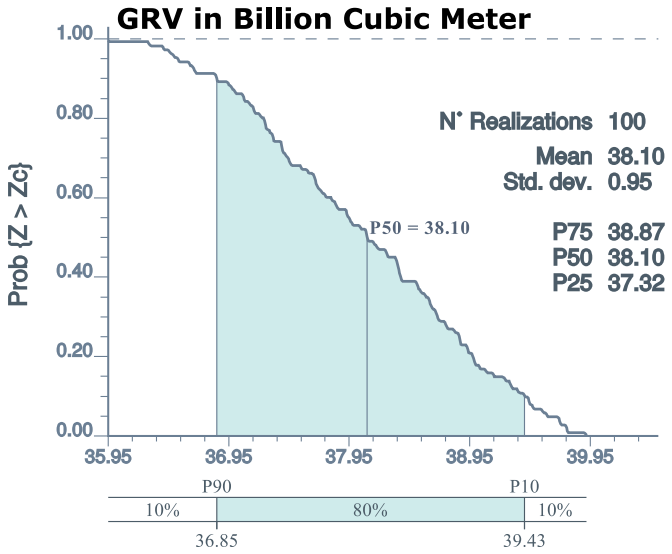

Figure 7.17: Histogram of uncertainty in GRV for 100 realizations in billion cubic meter (BCM).

7.3 Variogram Calculation and Uncertainty in Variogram

The experimental variogram of the well data is calculated and improved by secondary data for geostatistical modeling. Since there are only four wells, the horizontal variograms of the well data are not reliable. The vertical variogram from the well data and the horizontal variogram from the seismic data are used for improving the horizontal well variogram (Chapter 3, 4 and 5). The 3D acoustic impedance is converted from time domain (ms) to depth domain (m) by building a velocity model from checkshot survey of the well data. The 3D acoustic impedance model produced by dGB company has a very high resolution due to interpolation acoustic impedance for a very small vertical scale: the original seismic data cannot give information for a very fine vertical scale. The vertical scale of the acoustic impedance is upscaled to 8m for using in geostatistical modeling of this case study. Figure 7.18 shows the normal score acoustic impedance model with the vertical scale of 8m (after upscaling) after converting from time to depth. The top and bottom surfaces of the acoustic impedance model are clipped. There is a clear geometric anisotropy. The calculated variables of PHIT, PHIE and Vsh from the four wells are upscaled to 5m which is the vertical grid size of geostatistical modeling in this study; this is the vertical scale of geostatistical mapping of a similar reservoir to this case study in West Netherlands Basin (Vis et al., 2010). Figure 7.19 shows the correlation matrix between PHIT, PHIE and Vsh and seismic data (acoustic impedance) for the well data in original units. The diagonal elements show the histograms of the variables. The seismic data has the highest correlation with PHIT ($\rho(0) = -0.56$). Since PHIE is calculated by PHIT and Vsh, it has high correlation with these variables especially with Vsh ($\rho(0) = -0.97$) and low correlation with the seismic data ($\rho(0) = -0.36$).

The variograms of normal score PHIT, PHIE and Vsh are calculated for geostatistical modeling. A linear model of coregionalization (LMC) (Pyrcz & Deutsch, 2014) is required for the multivariate spatial bootstrap to calculate the joint prior histogram uncertainty of the variables (see Chapter 6). Since directional horizontal variograms cannot be calculated from the four sparse wells, variogram uncertainty is calculated

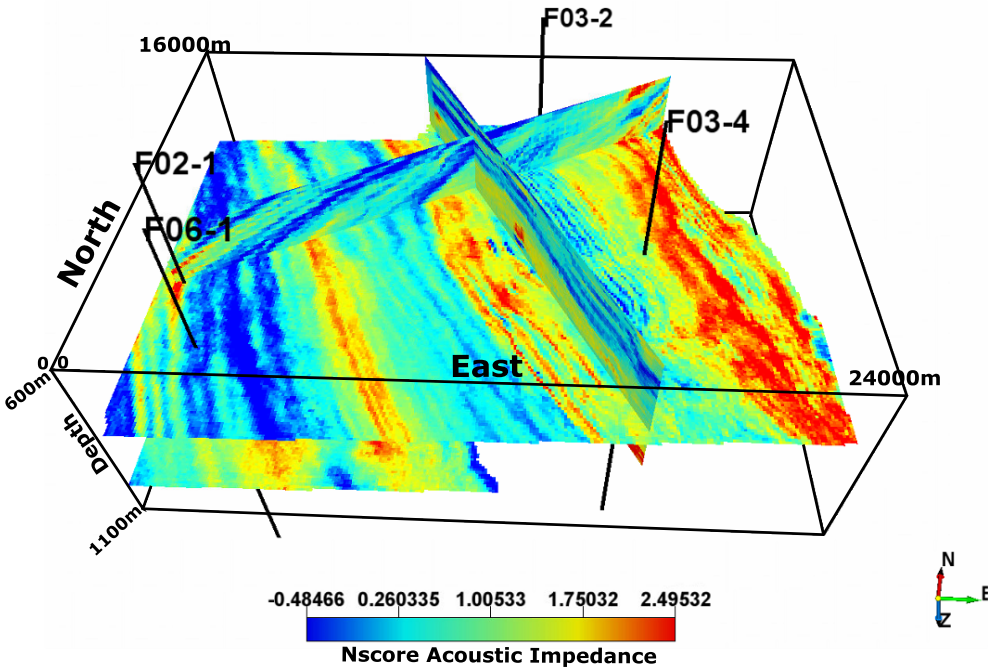

Figure 7.18: Normal score acoustic impedance in depth domain (vertical scale of 8m after upscaling).

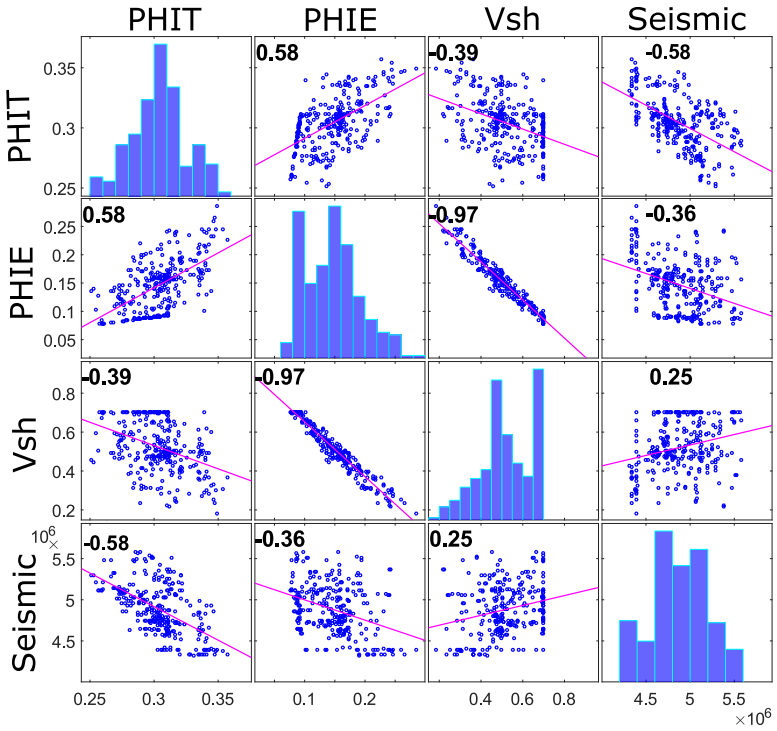

Figure 7.19: Correlation matrix between PHIT, PHIE and Vsh and seismic data (acoustic impedance) in original units.

for the omnidirectional horizontal variograms (azimuth $0° \pm 90°$). Figure 7.20 shows the omnidirectional horizontal direct and cross variograms for PHIT, PHIE and Vsh variables in normal score and fitted models. The omnidirectional horizontal variograms are noisy due to the sparse well data. Figure 7.21 shows the vertical direct and cross variograms for PHIT, PHIE and Vsh variables in normal score and fitted models. The vertical variogram are well-defined because of regular sampling in direction of drilling. Due to the vertical trend in data set (see Figure 7.2), the vertical variograms (Figure 7.21) do not reach to stationary sills (the variance of data for direct variograms and the covariance between variables at $\mathbf{h}$=0 for cross variograms). All variogram models are fitted to the stationary sills since a stationary variogram is required for geostatistical modeling. The sill of the cross variograms for normal score data are the correlation coefficients between variables (Pyrcz & Deutsch, 2014) and the sill of the direct variograms are 1. The linear model of coregionalization (LMC) for PHIT, PHIE and Vsh are written as:

$$
LMC \begin{cases}
\gamma_{PHIT}(\mathbf{h}) = 0.10.Sph\limits_{\substack{h_{major}=6000m \\ h_{minor}=6000m \\ h_{ver}=40m}}(\mathbf{h}) + 0.90.Sph\limits_{\substack{h_{major}=15000m \\ h_{minor}=15000m \\ h_{ver}=250m}}(\mathbf{h}) \\
\gamma_{PHIE}(\mathbf{h}) = 0.45.Sph\limits_{\substack{h_{major}=6000m \\ h_{minor}=6000m \\ h_{ver}=40m}}(\mathbf{h}) + 0.55.Sph\limits_{\substack{h_{major}=15000m \\ h_{minor}=15000m \\ h_{ver}=250m}}(\mathbf{h}) \\
\gamma_{V_{sh}}(\mathbf{h}) = 0.55.Sph\limits_{\substack{h_{major}=6000m \\ h_{minor}=6000m \\ h_{ver}=40m}}(\mathbf{h}) + 0.45.Sph\limits_{\substack{h_{major}=15000m \\ h_{minor}=15000m \\ h_{ver}=250m}}(\mathbf{h}) \\
\gamma_{PHIT,PHIE}(\mathbf{h}) = 0.20.Sph\limits_{\substack{h_{major}=6000m \\ h_{minor}=6000m \\ h_{ver}=40m}}(\mathbf{h}) + 0.38.Sph\limits_{\substack{h_{major}=15000m \\ h_{minor}=15000m \\ h_{ver}=250m}}(\mathbf{h}) \\
\gamma_{PHIT,Vsh}(\mathbf{h}) = -0.05.Sph\limits_{\substack{h_{major}=6000m \\ h_{minor}=6000m \\ h_{ver}=40m}}(\mathbf{h}) - 0.35.Sph\limits_{\substack{h_{major}=15000m \\ h_{minor}=15000m \\ h_{ver}=250m}}(\mathbf{h}) \\
\gamma_{PHIE,Vsh}(\mathbf{h}) = -0.46.Sph\limits_{\substack{h_{major}=6000m \\ h_{minor}=6000m \\ h_{ver}=40m}}(\mathbf{h}) - 0.496.Sph\limits_{\substack{h_{major}=15000m \\ h_{minor}=15000m \\ h_{ver}=250m}}(\mathbf{h})
\end{cases}
\tag{7.7}
$$

where Sph is the spherical variogram structure, h_{major} and h_{minor} are the horizontal variogram ranges for major and minor directions of continuity that are equal for omnidirectional variogram, and h_{ver} is the vertical variogram range. The variogram models

in Equation 7.7 are used in the multivariate spatial bootstrap for calculating the joint prior parameter uncertainty (see Section 7.5). Variogram uncertainty is not applied for the spatial bootstrap (see Chapter 6).

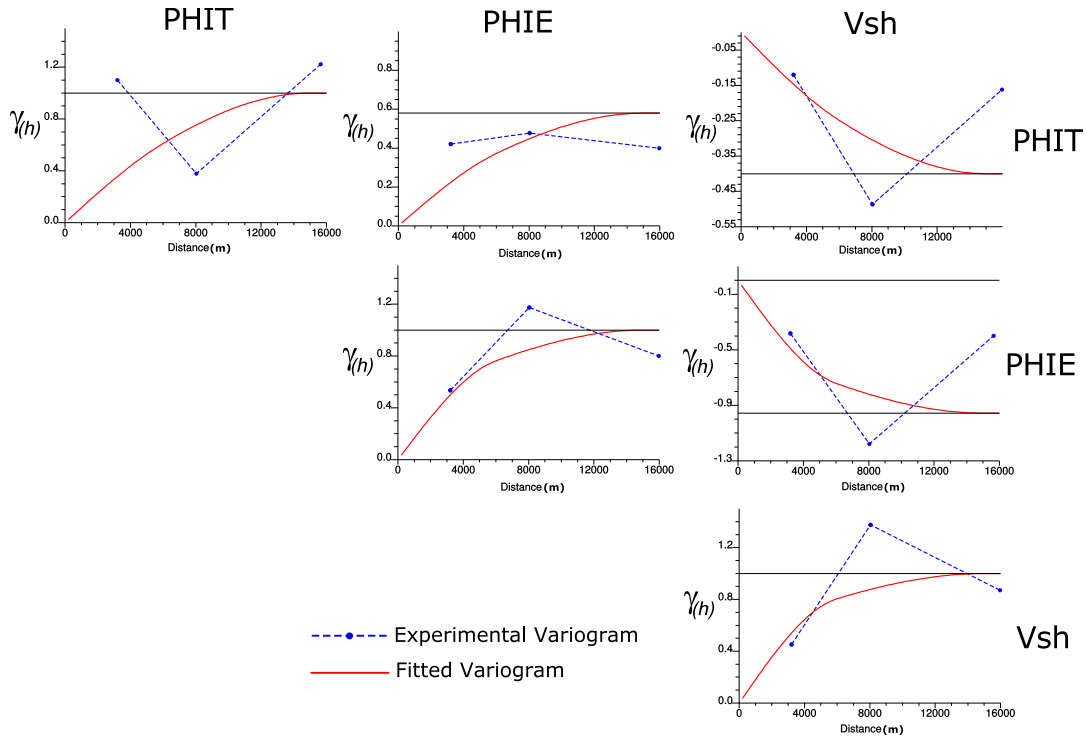

Figure 7.20: The omnidirectional horizontal direct and cross variograms (azimuth $0° \pm 90°$) for PHIT, PHIE and Vsh in normal score. The diagonal variograms are direct variograms and off diagonal variograms are the cross variograms between variables.

The omnidirectional horizontal variograms are noisy and unreliable; variogram uncertainty for omnidirectional horizontal variograms are calculated by the DoF approach and reduced by the seismic-derived and vertical-derived variograms (see Chapters 3, 4 and 5). Since PHIE and Vsh have very high correlation ($\rho(0) = -0.97$), the variogram realizations of these variables must be approximately equal. PHIE has higher correlation with the seismic data than Vsh; thus, the variogram uncertainty of PHIE is calculated, improved and assigned to Vsh. The omnidirectional horizontal variogram of the seismic data could also be calculated and used for improving the variogram of the well data. However, there is a clear geometric anisotropy in the seismic data for azimuths $-20°$ and $70°$. Although this anisotropy cannot be obtained from the well

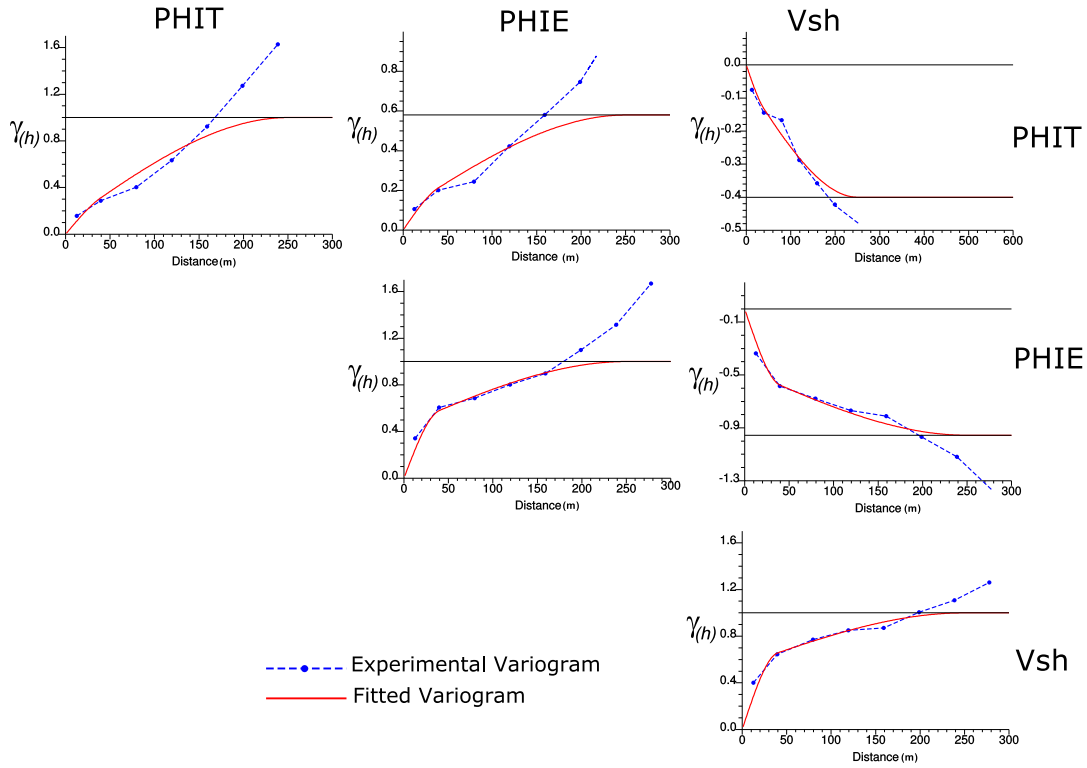

Figure 7.21: The vertical direct and cross variograms for PHIT, PHIE and Vsh in normal score. The diagonal variograms are direct variograms and off diagonal variograms are the cross variograms between variables.

data, it could be incorporated in the final horizontal variogram realizations by calculating the ratio of the major to minor variogram ranges of the seismic data and enforce it to all omnidirectional horizontal variogram realizations (see Section 7.3.3).

7.3.1 Improve Omnidirectional Horizontal Variogram Uncertainty of PHIT

Figure 7.22-a shows 100 omnidirectional horizontal variogram realizations of PHIT, Figure 7.22-b shows the correlation matrix of lag distances for variogram realizations, and Figure 7.22-c shows the fitted variogram realizations (see Chapters 3, 4 and 5). There is a high uncertainty in the variogram that could be reduced by the vertical variogram of the well data (vertical-derived variogram) and the omnidirectional horizontal variogram of the seismic data (seismic-derived variogram). Figure 7.23 shows the vertical-derived variogram realizations (100 realizations) for PHIT. According to a similar reservoir to this case study in West Netherlands Basin (Vis et al., 2010), the

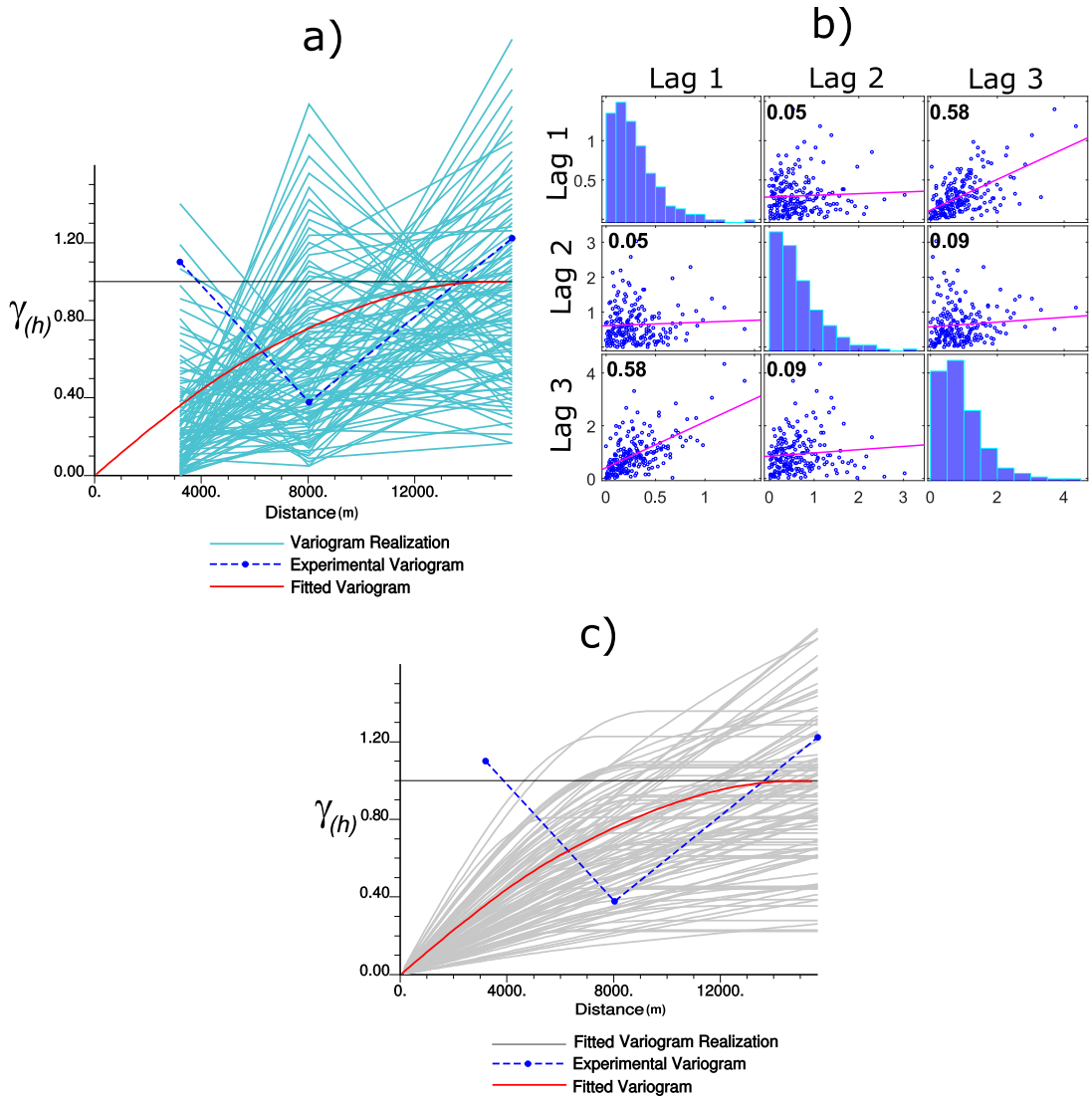

Figure 7.22: a) 100 omnidirectional horizontal variogram realizations of PHIT. b) Correlation matrix of lag distances for variogram realizations. c) Fitted variogram realizations.

horizontal to vertical anisotropy ratio (H:V) of 77 wells is 100 ($\frac{10000m}{100m}$). For calculating the vertical-derived variogram realizations, the distribution of uncertainty in H:V is build by the triangular distribution with $Min = 75$ and $Mode = 100$ and $Max = 125$. The Gaussian distribution of the sill is calculated by $\Delta_v = 0.15$ (see Chapter 5). The constraints from the omnidirectional horizontal variogram of the acoustic impedance in Figure 7.18 could be achieved to improve the omnidirectional horizontal variogram realizations of PHIT. The seismic variogram should be down scaled to the scale of the well data. The small scale variogram of the seismic data for down scaling could comes

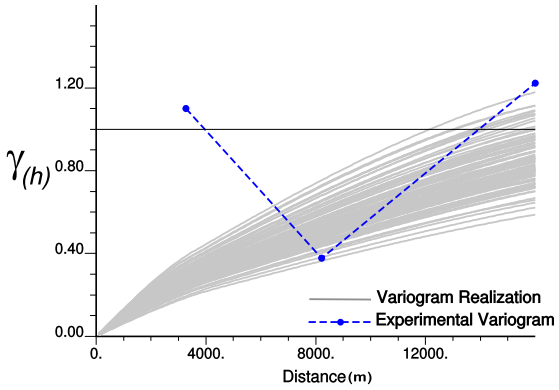

Figure 7.23: 100 vertical-derived variogram realizations for PHIT.

from the variogram of the calculated acoustic impedance based on the well logs (velocity and density). Since the vertical scale of well and seismic are not very different (5m versus 8m), the down scaled seismic variogram is changed a little for variance contributions (see Chapter 5). The cross covariance between well and seismic data is required for computing the the upper and lower limits of the seismic-derived variogram. The cross covariance is obtained by an intrinsic model (Markov model) (Pyrcz & Deutsch, 2014). The correlation between well and seismic at the scale of well data is required. This correlation could be calculated by fitting an Equation to the different upscaled volumes and the observed correlations from the data set; the higher scale, the higher correlation. Figure 7.24-a shows the cross plot between the positive correlation of the upscaled PHIT and the seismic data versus the vertical scale (volume), and the fitted Equation. The positive correlation between well and seismic data for the vertical scale of 5m (the vertical scale of the well data) is approximately 0.57. The seismic-derived upper and lower limits are calculated in Figure 7.24-b. These limits are enforced on the omnidirectional horizontal variogram realizations of PHIT (Figure 7.22-c), and the vertical-derived variogram realizations (Figure 7.23) up to correlation 0.2 ($\rho(\mathbf{h}) = 0.2$) by the rejection sampling approach with the correlation matrix of the lag distances (see Chapter 4 and 5). The final step is merging the omnidirectional horizontal variogram and the vertical-derived variogram after applying the seismic-derived limits.

Figure 7.25-a shows enforcing the upper and lower limits of the seismic-derived variogram on the omnidirectional horizontal variogram realizations (Figure 7.22-c) without

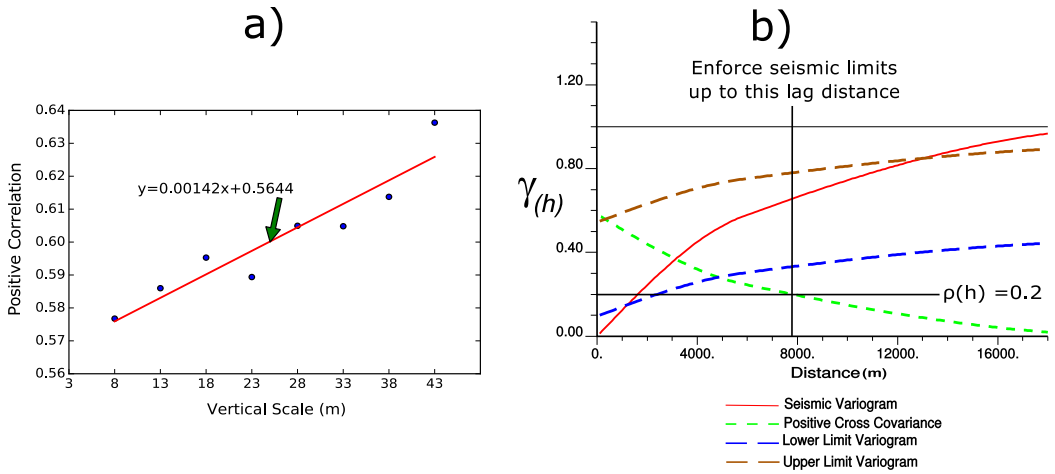

Figure 7.24: a) Cross plot between the positive correlation of the upscaled PHIT and seismic data versus the vertical scale (volume), and the fitted Equation. The positive correlation between well and seismic data for the vertical scale of 5m is approximately 0.57. b) Calculated seismic-derived variogram used up to correlation 0.2 ($\rho(\mathbf{h}) = 0.2$).

merging with the vertical-derived variogram. The variogram uncertainty is improved although the improvement is not considerable. Figure 7.25-b shows merging the omnidirectional horizontal variogram realizations with the vertical-derived variogram without enforcing the seismic limits. This is the case when seismic data is not used. The uncertainty goes down considerably because the uncertainty in the vertical-derived variogram is lower than the omnidirectional horizontal variogram and merging variogram distributions leads to less uncertainty (Chapter 5). Figure 7.25-c shows the final improved variogram realizations considering the horizontal variogram of the seismic data and the vertical variogram of well data: enforce upper and lower limits of the seismic-derived variogram on the omnidirectional horizontal variogram and the vertical-derived variogram realizations, merge variogram distributions and sample from the merged distribution by LU simulation. These realizations are standardized to sill 1 (Figure 7.25-d) and fitted with the fixed vertical variogram to attain the 3D variogram realizations of PHIT for univariate geostatistical modeling.

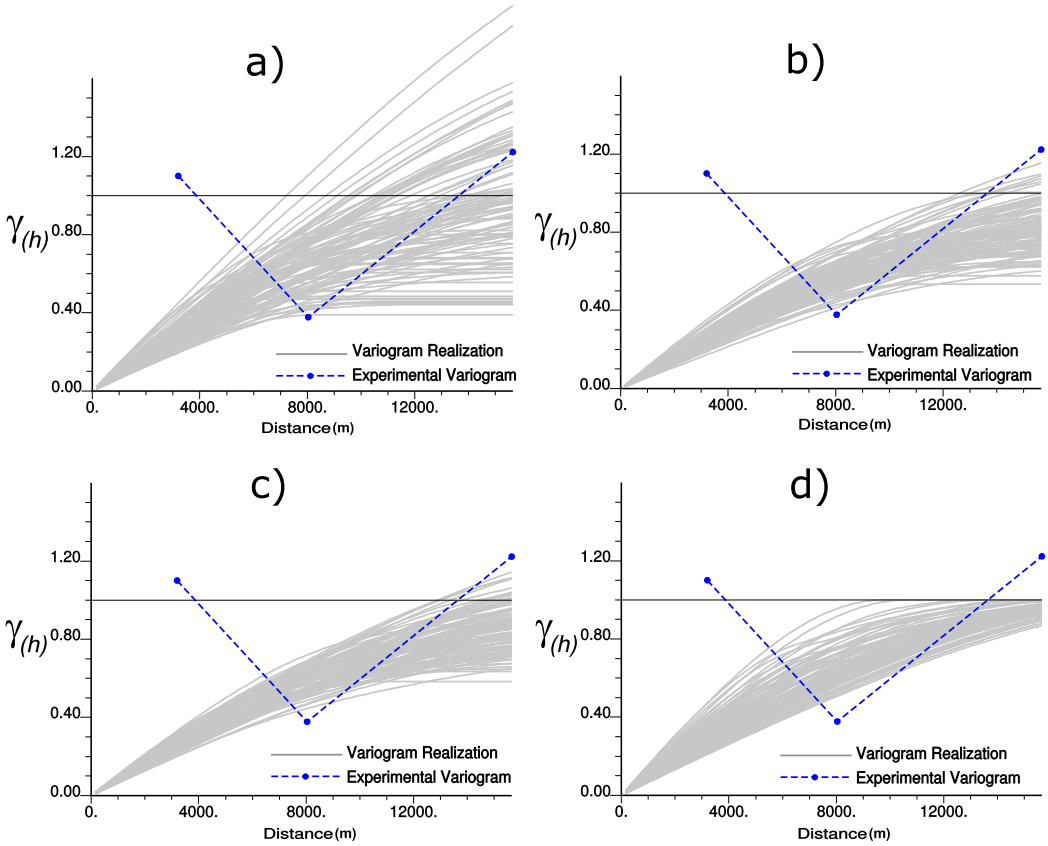

Figure 7.25: Improved omnidirectional horizontal variogram realizations for PHIT. a) Enforce the upper and lower limits of the seismic-derived variogram on omnidirectional horizontal variogram realizations without merging with the vertical-derived variogram. b) Merge the omnidirectional horizontal variogram realizations with the vertical-derived variogram without enforcing the seismic-derived variogram. c) Improved variogram realizations by enforcing the seismic-derived variogram, and the vertical-derived variogram. d) Standardized fitted variogram realizations.

7.3.2 Improve Omnidirectional Horizontal Variogram Uncertainty of PHIE

The similar approach is applied for improving the omnidirectional horizontal variogram realizations of PHIE. Figure 7.26-a shows 100 omnidirectional horizontal variogram realizations of PHIE, Figure 7.26-b shows the correlation matrix of lag distances for variogram realizations, and Figure 7.26-c shows the fitted variogram realizations. Figure 7.27 shows the vertical-derived variogram realizations (100 realizations) of the omnidirectional horizontal variogram for PHIE; the distribution of uncertainty in H:V is build

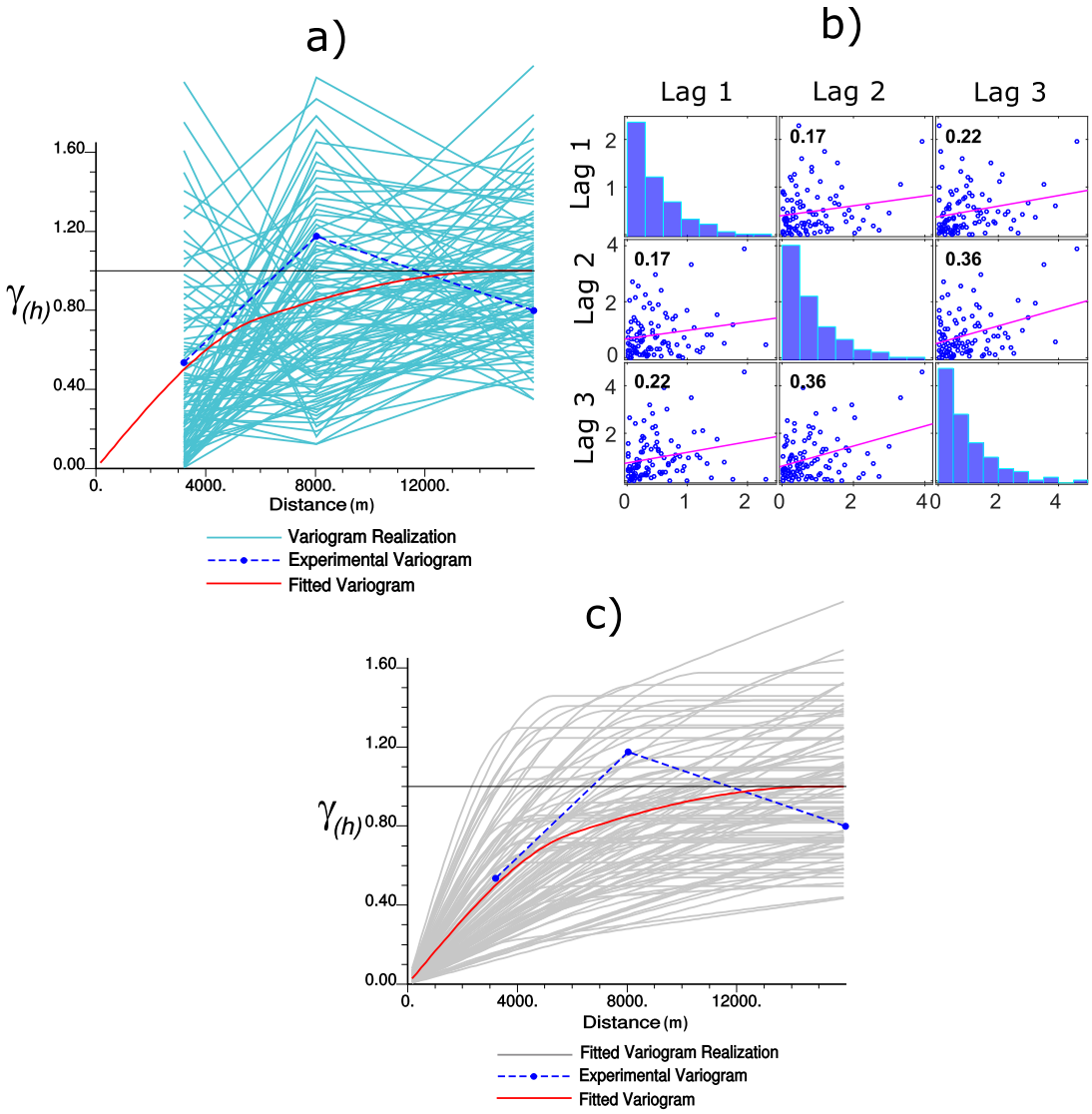

Figure 7.26: a) 100 omnidirectional horizontal variogram realizations of PHIE. b) Correlation matrix of lag distances for variogram realizations. c) Fitted variogram realizations.

by the triangular distribution with $Min = 75$, $Mode = 100$ and $Max = 125$. The Gaussian distribution of the sill is calculated by $\Delta_v = 0.15$. The correlation between well and seismic at the scale of the well data is calculated by fitting an Equation between the different vertical upscales and the observed correlations from the data set (Figure 7.28-a). The seismic-derived upper and lower limits are calculated in Figure 7.28-b. These limits are enforced on the omnidirectional horizontal variogram realizations of PHIE (Figure 7.26-c), and the vertical-derived variogram realizations (Figure

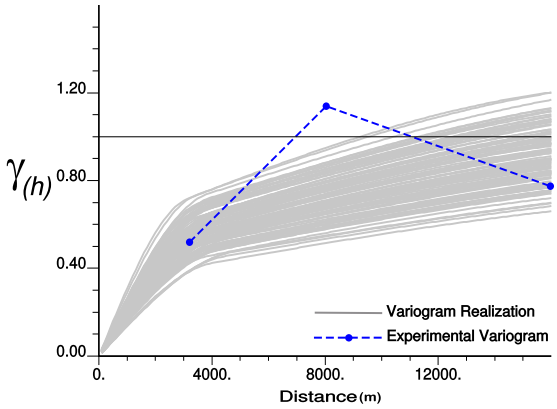

Figure 7.27: 100 vertical-derived variogram realizations for PHIE.

7.27) up to correlation 0.2 ($\rho(\mathbf{h}) = 0.2$). Figure 7.29-a shows enforcing the upper and lower limits of the seismic-derived variogram on the omnidirectional horizontal variogram realizations (Figure 7.26-c) without merging with the vertical-derived variogram. The variogram uncertainty is still high because the seismic-derived variogram is calculated based on low correlation between well and seismic data ($\rho(0) = 0.38$). Figure 7.29-b shows merging the omnidirectional horizontal variogram realizations with the vertical-derived variogram. Figure 7.29-c shows the final improved variogram realizations considering the horizontal variogram of seismic data and the vertical variogram of well data. The final omnidirectional horizontal variogram is more influenced by the vertical-derived variogram because of low correlation between well and seismic. These realizations are standardized to sill 1 (Figure 7.29-d) and fitted with the fixed vertical variogram to attain the 3D variogram realizations of PHIE for univariate geostatistical modeling.

7.3.3 Final 3D Variogram Realizations for Geostatistical Modeling

Since multivariate geostatistical modeling is considered, the omnidirectional horizontal variogram realizations of PHIT and PHIE, Vsh (Vsh=PHIE) should not be used independently without considering the correlation between variables: if the variables are correlated, the correlations should exist in the variogram realizations between variables. LU simulation is used to respect the correlation between variables for variogram realizations (see Chapter 5, Section 5.6). The negative correlation between variables

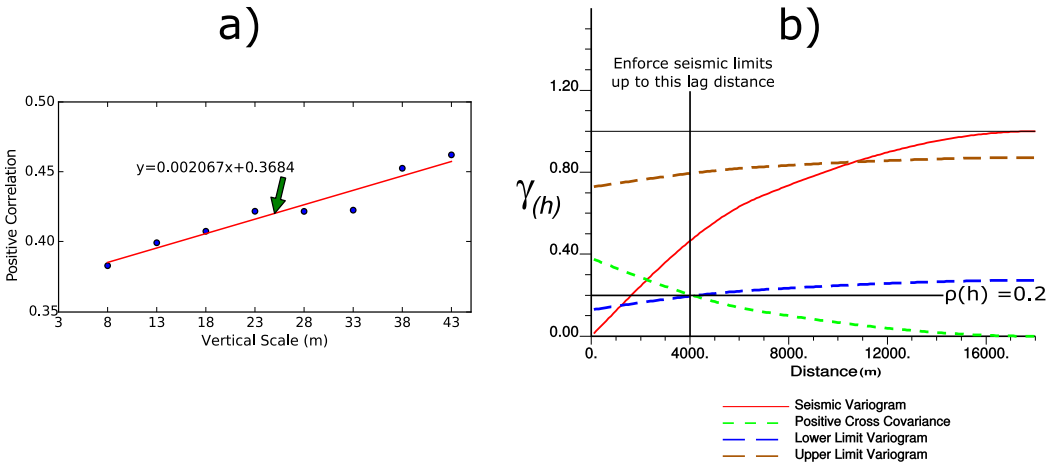

Figure 7.28: a) Cross plot between the positive correlation of the upscaled PHIE and seismic data versus the vertical scale (volume), and the fitted Equation. The positive correlation between well and seismic data for the vertical scale of 5m is approximately 0.38. b) Calculated seismic-derived variogram used up to correlation 0.2 ($\rho(\mathbf{h}) = 0.2$).

should be set to positive: even if variables have negative correlations, the variogram realizations should have positive correlations. Figure 7.30 shows the correlation matrix between the values of omnidirectional horizontal variogram realizations of PHIT, PHIE and Vsh for lag distances 8000m after LU simulation. The positive correlation between variables are preserved (compare Figure 7.30 with Figure 7.19).

After calculating the omnidirectional horizontal variogram realizations and preserve the correlation between variables, the clear anisotropy in the seismic data (Figure 7.18) should be enforced on the variogram realizations of all variables. For the seismic data (acoustic impedance), variogram range of azimuth $-20°$ is approximately 3 times bigger than the variogram range of azimuth $70°$. The omnidirectional horizontal variogram realizations are scaled based on this ratio: multiplying range of all variogram realizations by $\sqrt{3}$ and $\frac{1}{\sqrt{3}}$ for azimuths $-20°$ and $70°$, respectively. Then, 3D variogram realizations of each variable are attained by auto variogram modeling with the fixed vertical variogram and used in multivariate geostatistical modeling to incorporate variogram uncertainty.

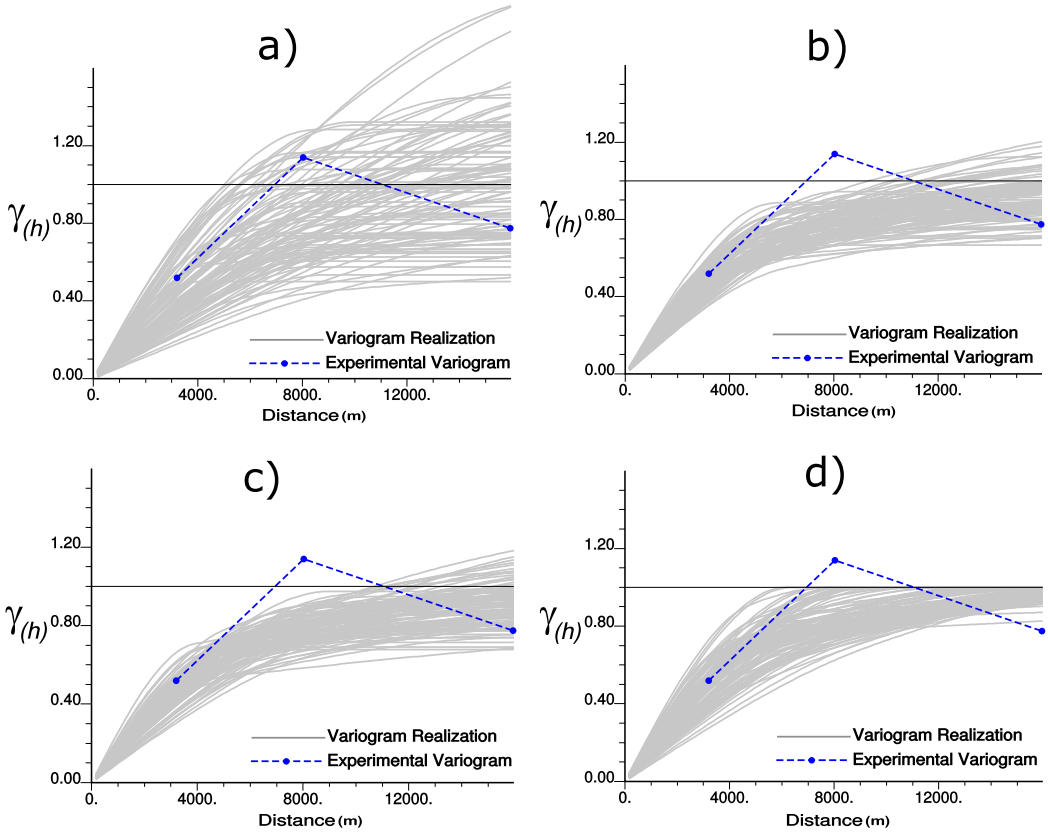

Figure 7.29: Improved omnidirectional horizontal variogram for PHIE. a) Enforce the upper and lower seismic-derived limits on the omnidirectional horizontal variogram realizations without merging with the vertical-derived variogram. b) Merge the omnidirectional horizontal variogram realizations with the vertical-derived variogram without enforcing the seismic limits. c) Improved variogram realizations by enforcing upper and lower seismic-derived limits, and merge variogram distributions. d) Standardized fitted variogram realizations.

7.4 Univariate Geostatistical Modeling without Parameter Uncertainty

Before incorporating variogram and histogram uncertainties in multivariate geostatistical modeling, each variable is modeled independently with the fixed variogram model and fixed reference distribution in SGS without using seismic data for the entire 3D acoustic impedance model (see Figure 7.18). The main aim is checking the histogram reproduction of variables PHIT, PHIE, Vsh after modeling. Figure 7.2 shows a clear vertical trend especially for PHIE, Vsh (see vertical variogram in Figure 7.21). The

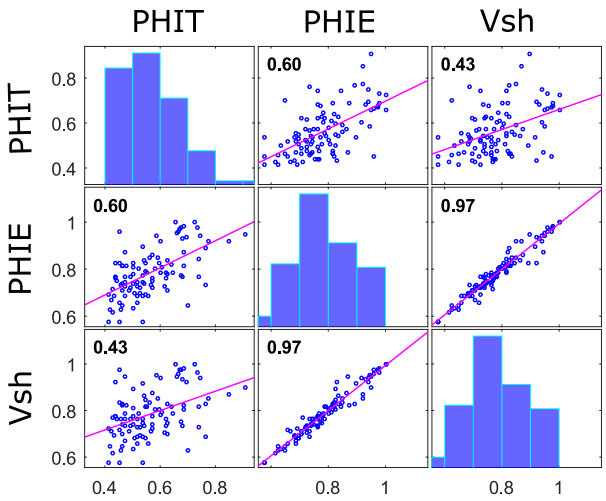

Figure 7.30: Correlation matrix of 100 omnidirectional horizontal variogram realization values for PHIT, PHIE and Vsh at lag distance 8000m after LU simulation to preserve the positive correlations between the variables.

trend usually has an impact on histogram reproduction. Figure 7.31-a shows 100 histogram of SGS realizations for PHIT with the fixed reference distribution (red line) and fixed variogram model from Equation 7.7. The histogram reproduction is good in general but for high values (larger than 0.3), they tend to be lower than the reference histogram. Figure 7.31-b shows histogram reproduction for unconditional simulation. Because of no conditioning data, there is higher uncertainty in histogram of realizations; however, the reference histogram is approximately in the middle. Figure 7.31-c shows histogram reproduction for a pure nugget effect variogram (nugget effect 1). All SGS realizations are normal with mean 0 and standard deviation 1; so, histogram reproduction is consistent with the reference distribution. Figure 7.32 and 7.33 show the same approach for assessment histogram reproduction of PHIE and Vsh. Since the trend is stronger for these variables especially for Vsh and there is a zonal anisotropy in well F03-4 (Figure 7.2), 100 histogram of SGS realizations are shifted to the left of the reference distribution for PHIE (see Figure 7.32-a), and to the right of the reference distribution for Vsh (Figure 7.33-a). By unconditional simulation (see Figure 7.32-b for PHIE and Figure 7.33-b for Vsh), or SGS using a pure nugget effect variogram (see Figure 7.32-c for PHIE and Figure 7.33-c for Vsh), the problem of shifting histogram realizations is resolved. This confirms that there is neither a problem with the geo-

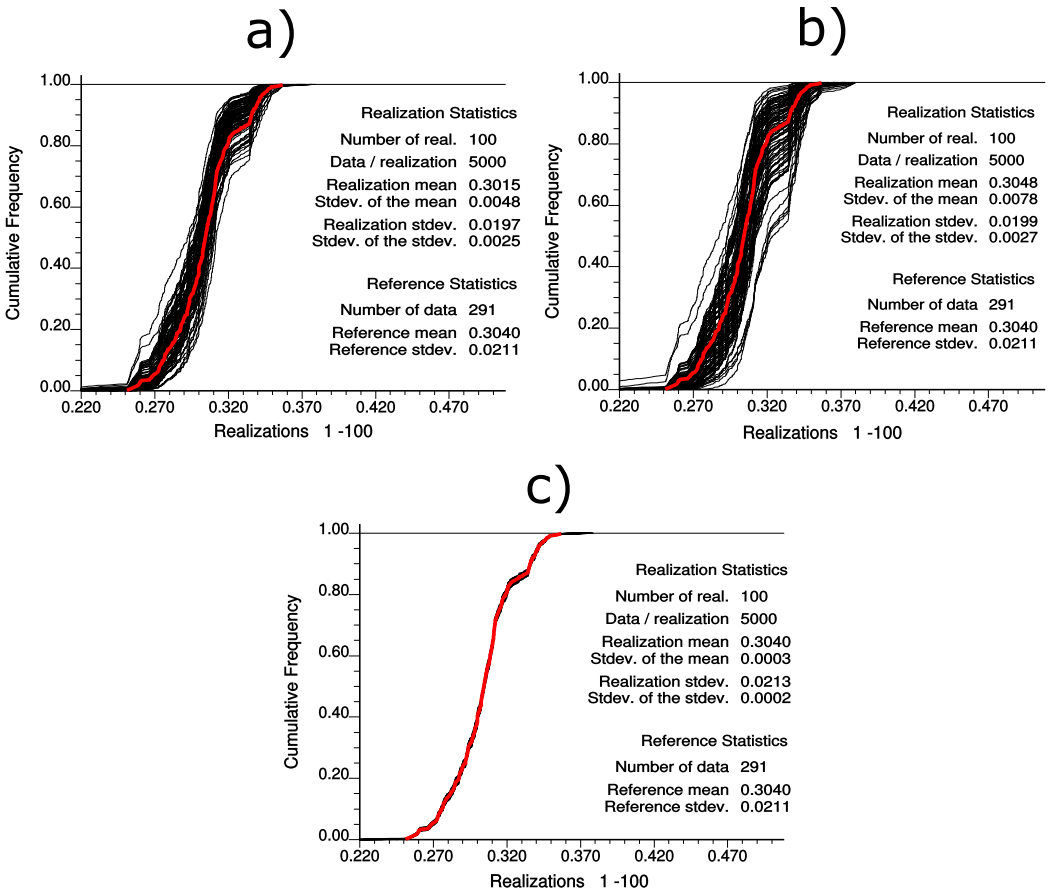

Figure 7.31: Histogram reproduction for PHIT (black lines). The red line is the reference distribution. a) Histogram reproduction with the fixed reference distribution (distribution of data) and fixed variogram model from Equation 7.7. b) Histogram reproduction for unconditional simulation. c) Histogram reproduction for a pure nugget effect variogram.

statistical technique nor with the reference histograms, the problem is only related to the trend in the well data. The shifting of histogram realizations can also be seen in multivariate geostatistical modeling with super secondary variable (see next Section).

The histogram of the well data may be declustered and improved by seismic data (Deutsch et al., 1999, 2005). However, variables PHIE and Vsh do not have a good correlation with the acoustic impedance (see Figure 7.19), thus, seismic data cannot be used to decluster the histogram of these variables. Because of multivariate geostatistical modeling, the distribution of all variables should be consistent: soft data declustering should be applied for all variables not for one of them. Therefore, the distributions of well data without soft data declustering are used in multivariate geostatistical modeling

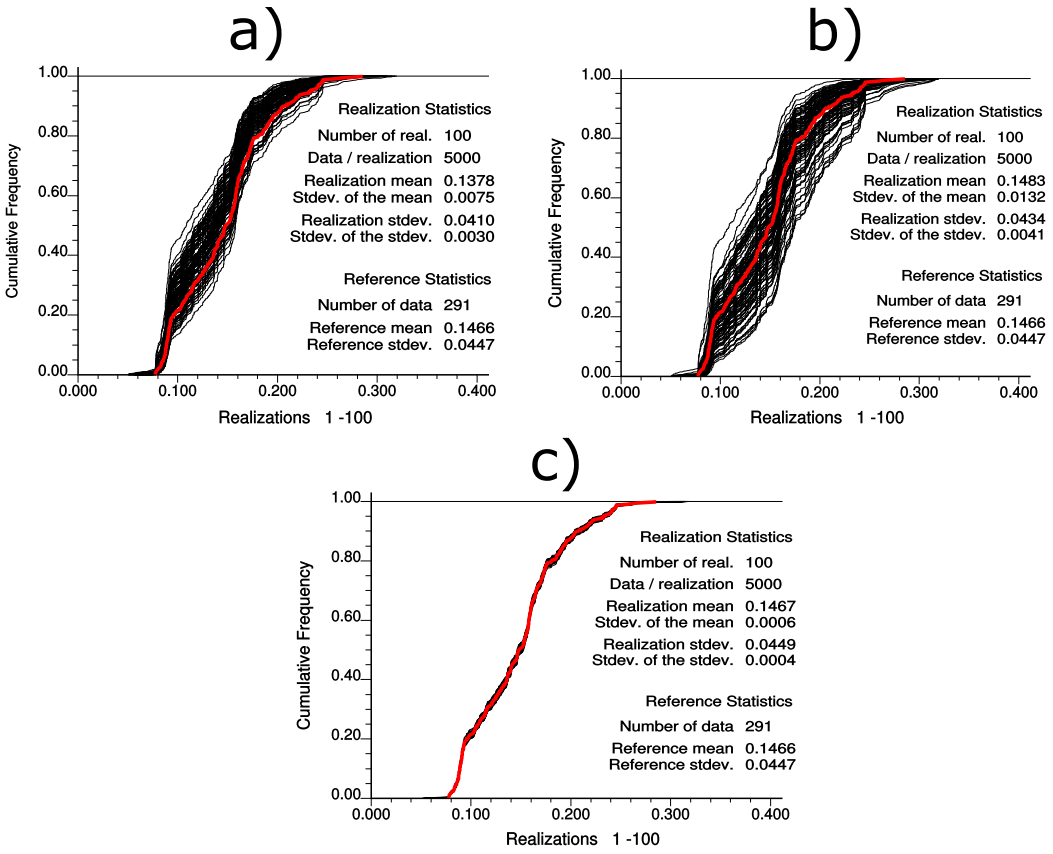

Figure 7.32: Histogram reproduction for PHIE (black lines). The red line is the reference distribution. a) Histogram reproduction with the fixed reference distribution (distribution of data) and fixed variogram model from Equation 7.7. b) Histogram reproduction for unconditional simulation. c) Histogram reproduction for a pure nugget effect variogram.

although seismic data (acoustic impedance) is used as secondary variable for modeling of each variable.

7.5 Multivariate Geostatistical Modeling with Parameter Uncertainty

Multivariate geostatistcial modeling with parameter uncertainty (histogram and variogram uncertainty) is applied to simulate each variable for the entire 3D acoustic model volume (see Figure 7.18). The joint prior histogram uncertainty is computed by the multivariate spatial bootstrap (see Chapter 6) for PHIT, PHIE, Vsh variables using the linear model of coregionalization (LMC) (see Equation 7.7). The correlated histogram

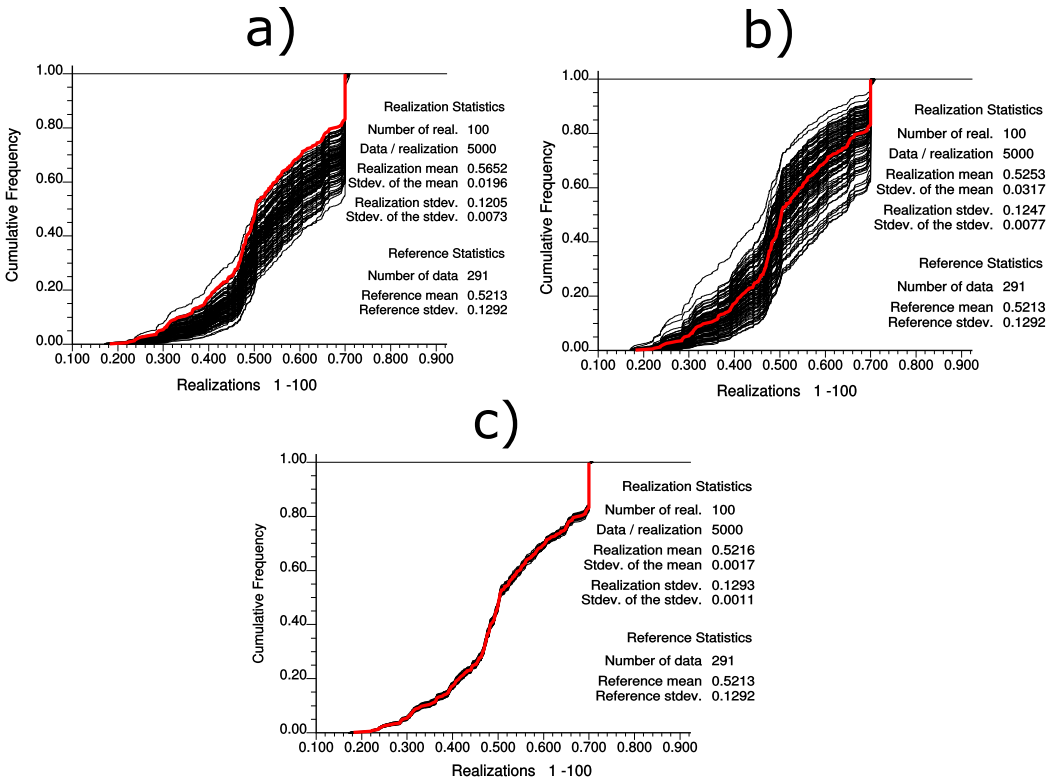

Figure 7.33: Histogram reproduction for Vsh (black lines). The red line is the reference distribution. a) Histogram reproduction with the fixed reference distribution (distribution of data) and fixed variogram model from Equation 7.7. b) Histogram reproduction for unconditional simulation. c) Histogram reproduction for a pure nugget effect variogram.

realizations of PHIT, PHIE and Vsh are used as reference distributions in conditional SGS for normal score transformation of data and back transforming the conditional Gaussian realizations to the original units for each variable. The improved variogram realizations of each variable (Section 7.3) are used in SGS to incorporate variogram uncertainty.

The histogram and variogram uncertainties are applied in cosimulation work flow: SGS with intrinsic collocated cokriging (ICC) and using super secondary variable (Babak & Deutsch, 2007b; Manchuk & Deutsch, 2012) is applied for multivariate geostatistical modeling; collocated cokriging with intrinsic model is used to avoid variance inflation (Babak & Deutsch, 2007a). 100 realizations of each variable with histogram and variogram uncertainties are simulated by a hierarchical modeling work flow:

1. Since PHIT has the highest correlation with 3D seismic data (acoustic impedance)

(see Figure 7.19), it is first simulated with one realization of histogram uncertainty and one realization of variogram uncertainty using ICC by normal score 3D acoustic impedance. Then, it is clipped to the top and bottom surfaces of the 3D acoustic impedance model. Figure 7.34 shows one SGS realization of PHIT.

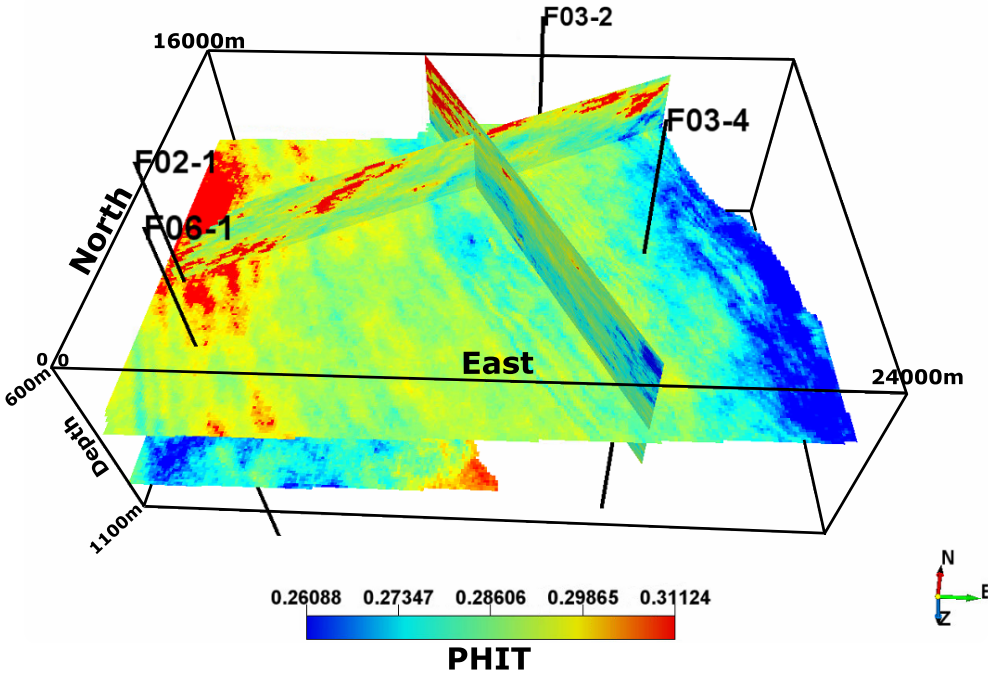

Figure 7.34: One SGS realization of PHIT with ICC using 3D acoustic impedance.

2. For simulating one realization of PHIE, super secondary variable is built by the 3D acoustic impedance and one realization of previously simulated PHIT. SGS with one realization of histogram uncertainty and one realization of variogram uncertainty using ICC by the super secondary variable is applied, and then it is clipped to the top and bottom surfaces of the 3D acoustic impedance. Figure 7.35 shows one SGS realization of PHIE. The correlation between variables PHIT and PHIE is preserved after modeling.

3. Finally, one realization of Vsh is simulated with super secondary variable built by the 3D acoustic impedance, one realization of previously simulated PHIT and PHIE. SGS with one realization of histogram and variogram uncertainties using ICC by super secondary variable is applied and clipped to the top and bottom surfaces of the 3D acoustic impedance. Figure 7.36 shows one SGS realization

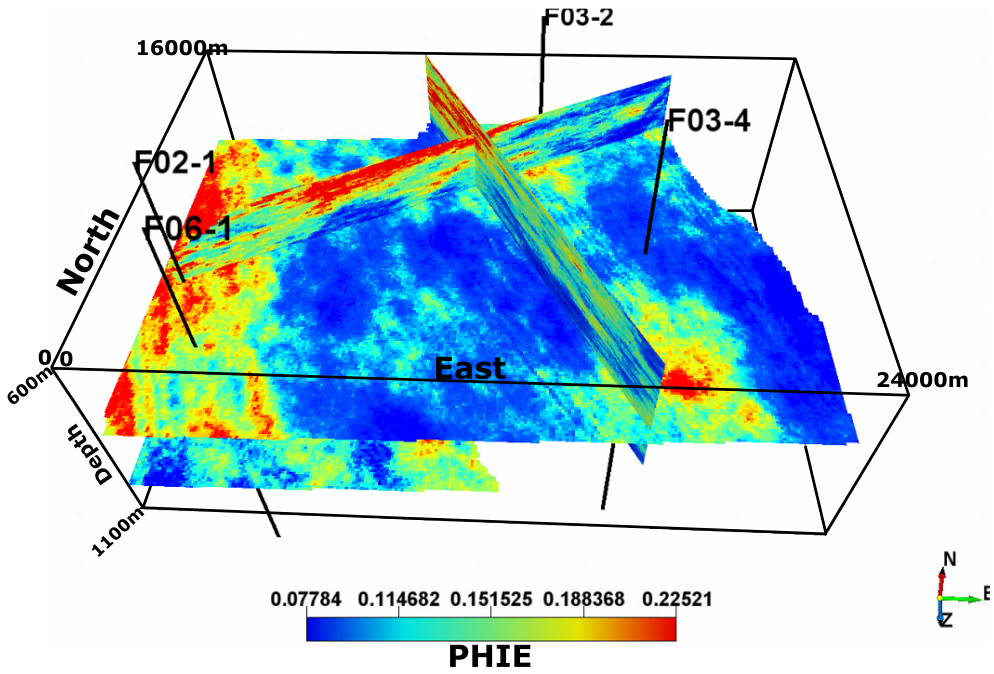

Figure 7.35: SGS realization of PHIE with ICC using super secondary variable built by the 3D acoustic impedance and realization of PHIT.

of Vsh. The correlation between variables PHIT and Vsh, PHIE and Vsh are preserved after modeling.

Steps 1 to 3 are repeated to simulate 100 realizations of PHIT, PHIE and Vsh incorporating histogram and variogram uncertainties.

The multivariate spatial bootstrap respects the correlations between PHIT, PHIE and Vsh for each histogram realization. This leads to uncertainty in correlation between variables. Figure 7.37 shows one histogram realization of the multivariate spatial bootstrap for PHIT, PHIE and Vsh. Figure 7.38 shows the uncertainty in the correlations between PHIT and PHIE (a), PHIT and Vsh (b) and PHIE and Vsh (c). There are very low and very high correlations in the histogram of correlations. The corresponding histogram realizations of these correlations may be rejected, and new histograms are simulated to satisfy the proper correlations. By incorporating the joint prior histogram uncertainty in geostatistical simulation, the uncertainty in correlations between variables will be incorporated in the final model.

The hierarchical geostatistical modeling by ICC with super secondary variable re-

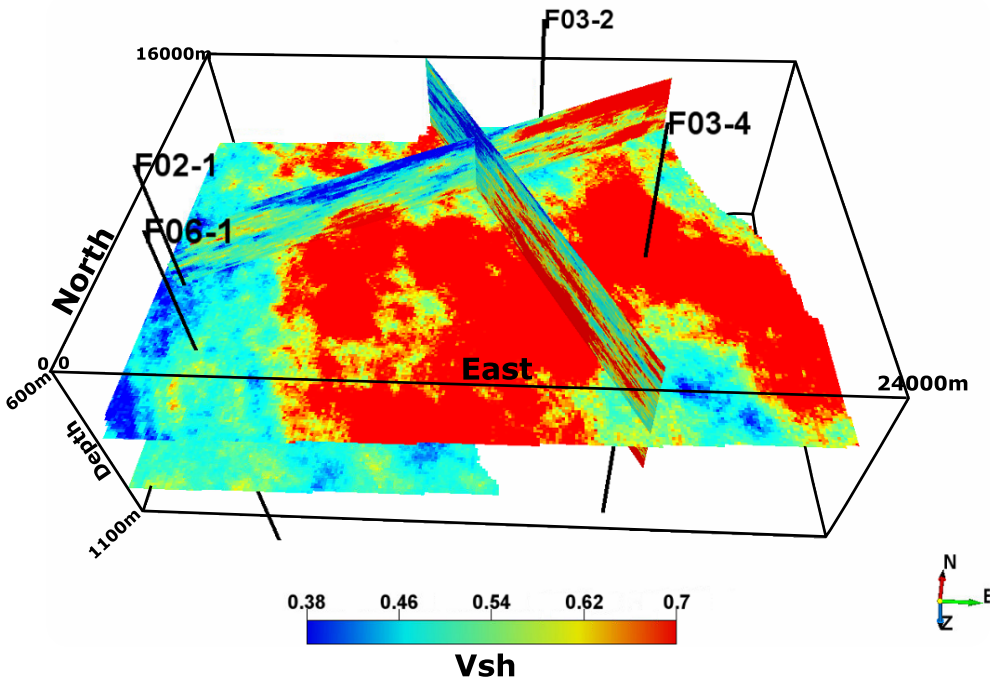

Figure 7.36: SGS realization of Vsh with ICC using super secondary variable built by the 3D acoustic impedance, realization of PHIT and PHIE.

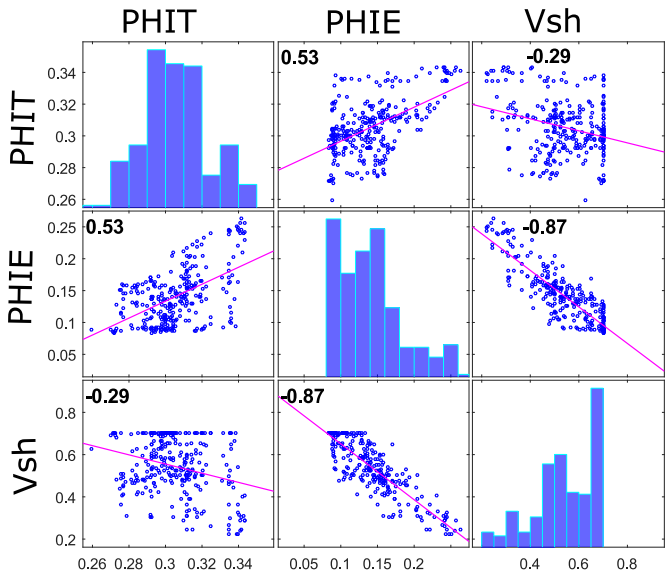

Figure 7.37: One histogram realization of the multivariate spatial bootstrap for PHIT, PHIE and Vsh.

spects the correlations between PHIT, PHIE and Vsh and the 3D acoustic impedance model; compared Figures 7.35 and 7.36, there is a high negative correlation between PHIE and Vsh after modeling; these variables are highly correlated before modeling.

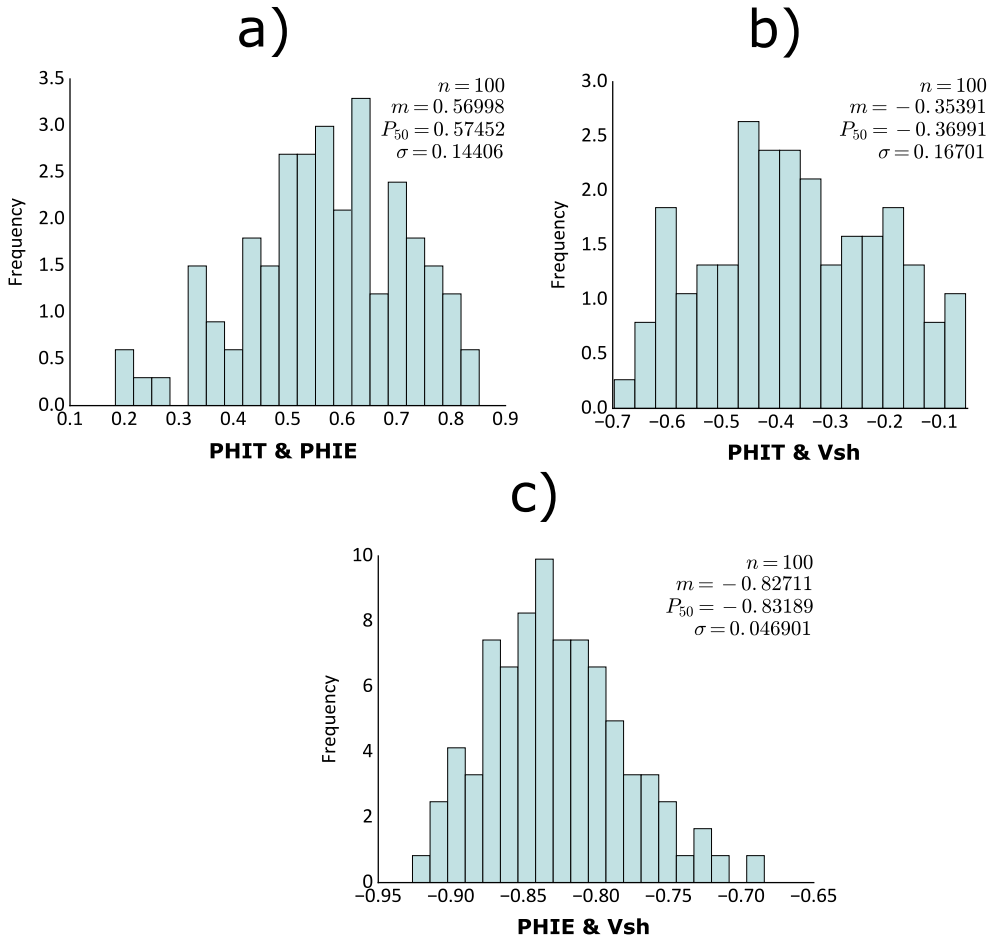

Figure 7.38: Uncertainty in correlations between PHIT, PHIE and Vsh by the multivariate spatial bootstrap. a) PHIT versus PHIE. b) PHIT versus Vsh. c) PHIE versus Vsh.

Figure 7.39 and Figure 7.40 show the crossplots between PHIT versus seismic (acoustic impedance) and PHIT versus PHIE before (a) and after modeling (b) for one realization. These variables have high correlations before modeling. The high positive and negative correlations between variables are reproduced after modeling. The values of variables are changed after modeling because of using different minimum and maximum values for tail extrapolation in SGS than the minimum and maximum of data (lower and higher than the minimum and maximum of data for each variable). The difference between correlation coefficients before and after modeling denotes the updated correlation coefficient between variables; this approach incorporates the uncertainty in correlation coefficient between primary (PHIT, PHIE and Vsh) and secondary variables

(the acoustic impedance model) in geostatistical modeling.

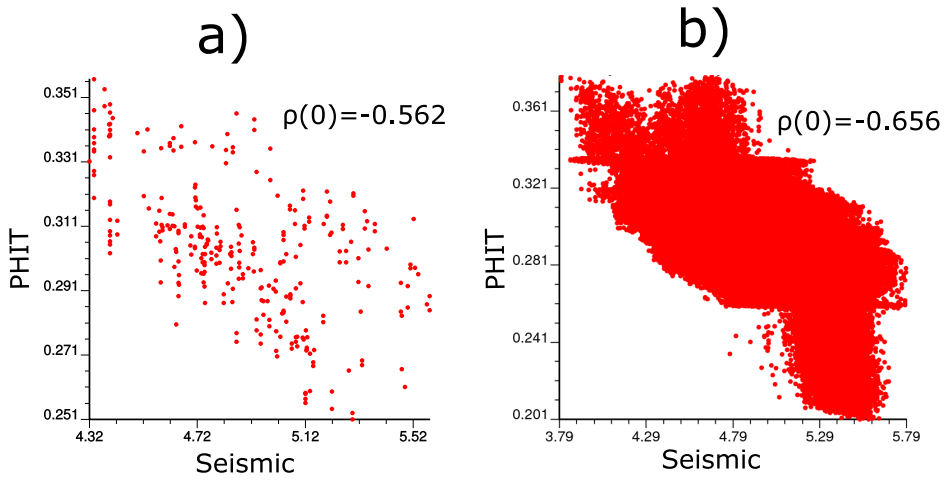

Figure 7.39: Crossplots between PHIT versus seismic before modeling (a) and after modeling (b) for one realization.

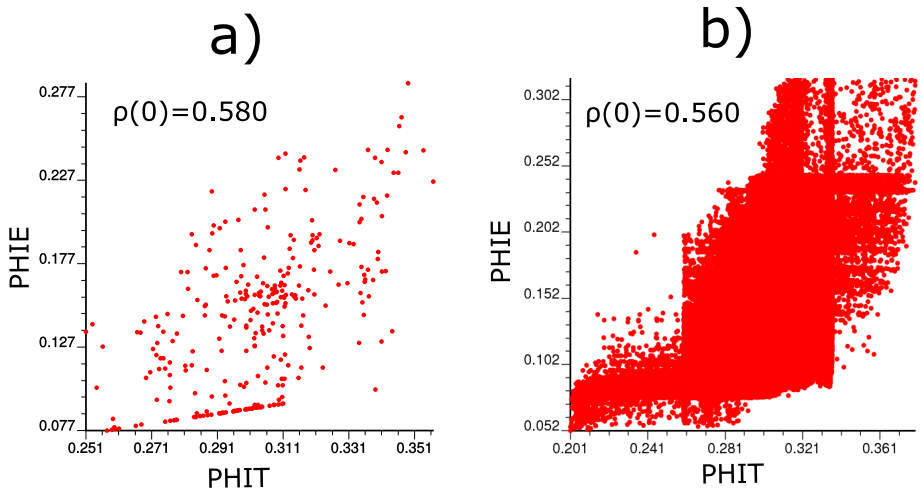

Figure 7.40: Crossplots between PHIT versus PHIE before modeling (a) and after modeling (b) for one realization.

The joint prior histogram uncertainty calculated by the multivariate spatial bootstrap is compared with histogram uncertainty by only fluctuations of SGS realizations after multivariate geostatistical modeling (ICC with super secondary), and posterior histogram uncertainty by multivariate geostatistical modeling using parameter uncertainty (histogram and variogram uncertainties). Figures 7.41, 7.42 and 7.43 show the prior histogram uncertainty (a), histogram uncertainty by only SGS realizations (b) and posterior histogram uncertainty (c) for PHIT, PHIE and Vsh, respectively. For all

variables, the prior histogram uncertainty is very high because it is calculated based on unconditional LU simulation. The histogram uncertainty by only SGS realization is low since it does not account for parameter uncertainty. Incorporating the high prior histogram uncertainty in geostatistical simulation by conditioning and clipping leads to more accurate histogram uncertainty (Chapter 6): it is lower than the prior uncertainty and higher than the histogram uncertainty by only SGS realizations. Since there are only four well data (conditioning data), the posterior histogram uncertainty does not decrease noticeably in comparison with prior uncertainty. The problem of histogram reproduction due to trend in well data discussed in univariate modeling (see Figures 7.31-a, 7.32-a and 7.33-a) also exists in multivariate geostatistical modeling without parameter uncertainty (see Figures 7.41-b, 7.42-b and 7.43-b). By incorporating the prior histogram uncertainty in geostatistical simulation, the shifting of the histogram realizations of the posterior uncertainty is not noticeable especially for Vsh. This is because of increasing the uncertainty in posterior histogram realizations.

The posterior variogram uncertainty could also be calculated by the variogram of each SGS realization to assess the results. This uncertainty is computed without any parameter uncertainty: no variogram uncertainty (fixed 3D variogram) and no histogram uncertainty (fixed reference distribution), and with parameter uncertainties: prior variogram and histogram uncertainties. Figures 7.44-a shows the posterior variogram uncertainty (100 realizations) for azimuths 0° and 90° of PHIE without parameter uncertainty, and Figures 7.44-b shows the posterior variogram uncertainties with both prior variogram and histogram uncertainties. Due to using the seismic data in multivariate geostatistical modeling, the posterior variogram uncertainty without parameter uncertainty and with parameter uncertainty are approximately the same: using parameter uncertainty leads to a little higher posterior variogram uncertainty. The posterior variogram uncertainty is influenced by the variogram of well and seismic data because of conditioning. If there is a very high prior variogram uncertainty but enough well data, conditioning of the well data in geostatistical modeling leads to low posterior variogram uncertainty (see Chapter 3). Moreover, if there are sparse well data but high correlation between well and seismic, conditioning of the seismic data

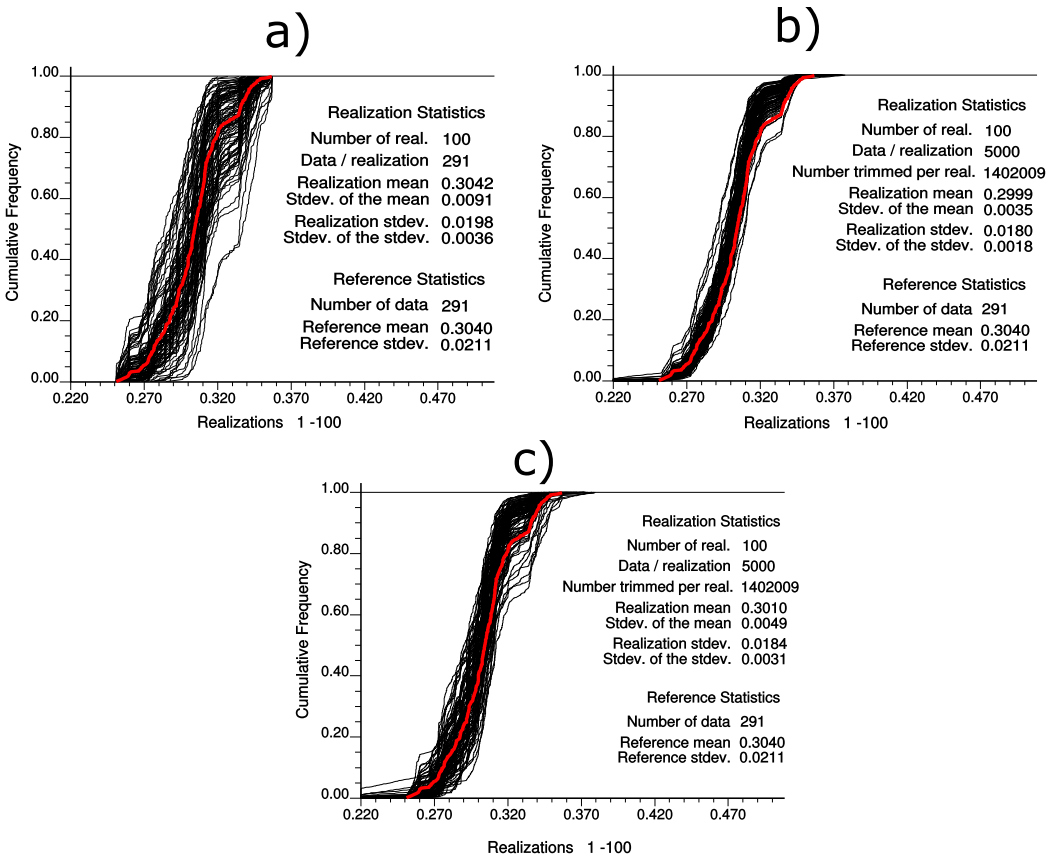

Figure 7.41: The prior histogram uncertainty (a), histogram uncertainty by only SGS realizations (fixed histogram and variogram) (b) and posterior histogram uncertainty (c) for PHIT. The red line is the reference distribution and black lines are histogram realizations.

in geostatistical modeling leads to low posterior variogram uncertainty. For this case study, there is a relatively high correlation between PHIT and seismic data. This leads to a super secondary (for PHIE and Vsh) which has high correlation with primary variable (primary variables are also highly correlated). Therefore, the posterior variogram uncertainties for all variables are influenced by the variogram of the acoustic impedance: very high continuity for azimuth 0° and very low continuity for azimuth 90° for both Figures 7.44-a and b.

7.6 Uncertainty in HIIP

Quantifying HIIP with parameter uncertainty is the main objective of this study. The results of multivariate geostatistical modeling for reservoir properties and GRV with

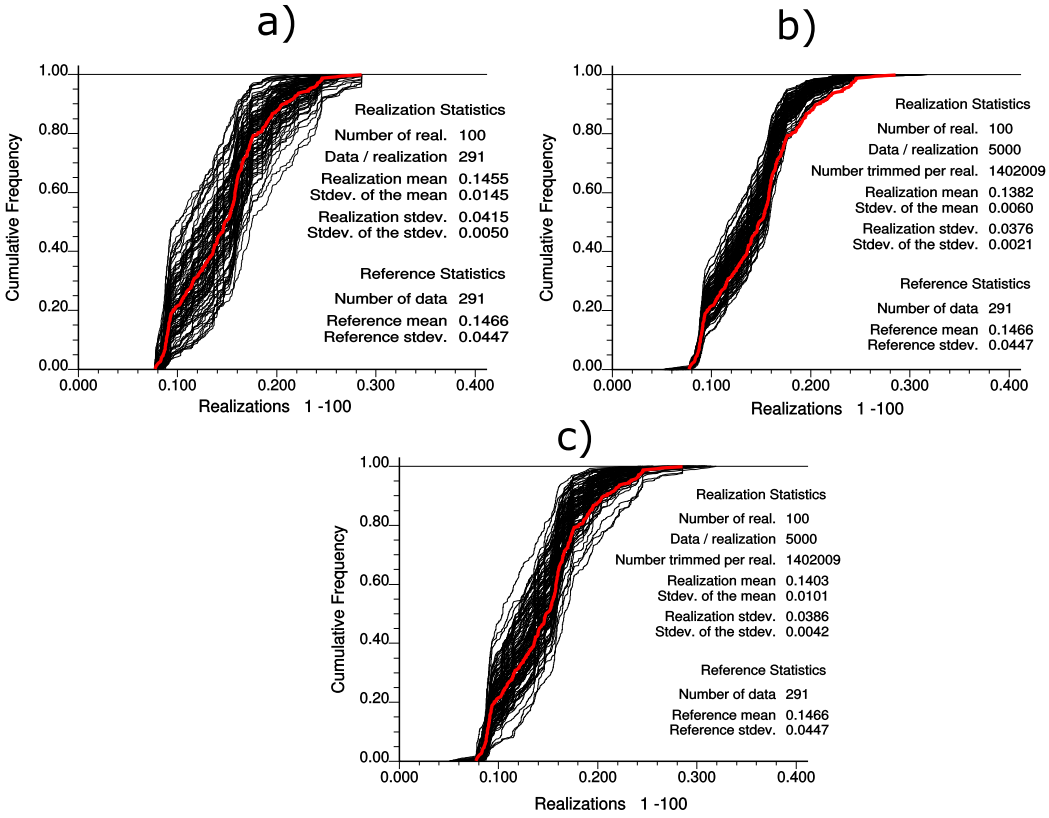

Figure 7.42: The prior histogram uncertainty (a), histogram uncertainty by only SGS realizations (fixed histogram and variogram) (b) and posterior histogram uncertainty (c) for PHIE. The red line is the reference distribution and black lines are histogram realizations.

parameter uncertainty are used to quantify uncertainty in HIIP. There are no water saturation data; Vsh is assumed as water saturation in this study. Since shale includes impermeable water, this impermeable water is likely to be the irreducible water saturation. One realization of HIIP is calculated by sum of Vl$\times$ PHIE $\times(1-Sw)$, (Vl= $100m \times 100m \times 5m$) for those grids satisfied in GRV realization (realizations of top, thickness and HWC): one realization of all variables are used to calculate one realization of HIIP. Vl is the volume of each sugar cube grid ($x = 100m$, $y = 100m$, $z = 5m$) and Sw is water saturation replaced with Vsh. This ensures that the parameter uncertainties for reservoir properties (histogram and variogram uncertainties), surfaces (uncertainties in top, thickness) and HWC are incorporated in the final model of HIIP. To define the importance of each source of uncertainty, HIIP is calculated for seven cases by incorporating the following parameter uncertainties:

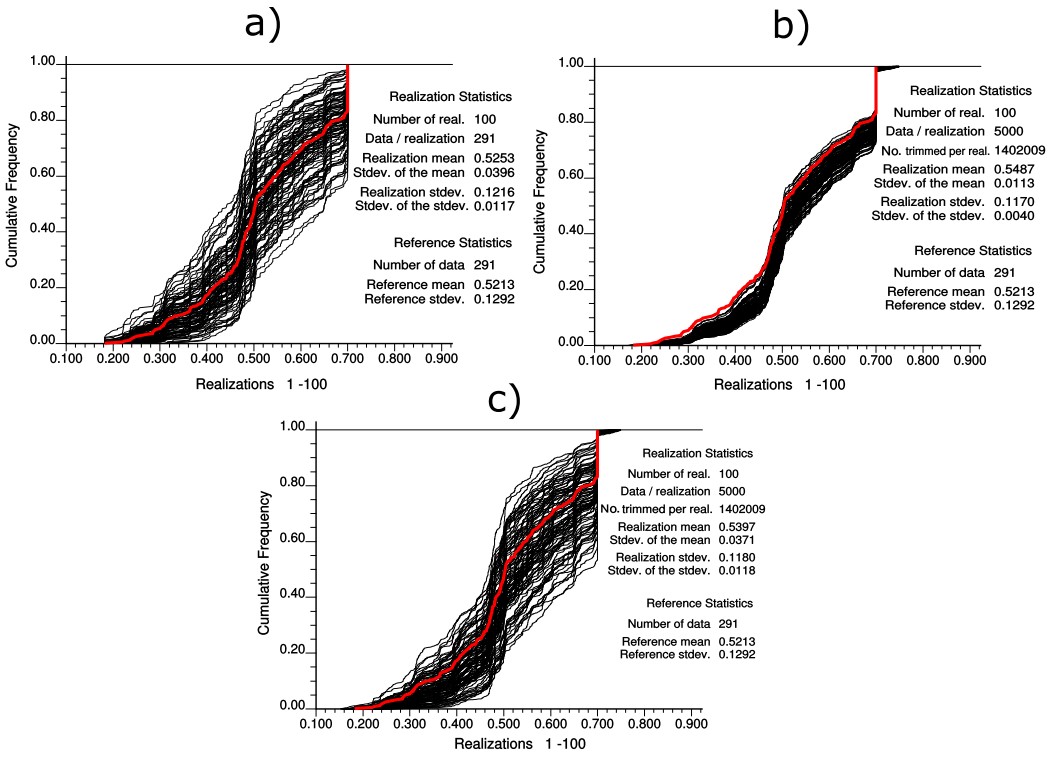

Figure 7.43: The prior histogram uncertainty (a), histogram uncertainty by only SGS realizations (fixed histogram and variogram) (b) and posterior histogram uncertainty (c) for Vsh. The red line is the reference distribution and black lines are histogram realizations.

1. All uncertainties such as variogram, histogram, surface and HWC.
2. Histogram and variogram uncertainties (freeze surface and HWC uncertainties).
3. Histogram uncertainty (freeze variogram, surface and HWC uncertainties).
4. Surface uncertainties: uncertainty in top surface and thickness (freeze variogram, histogram and HWC uncertainties).
5. HWC uncertainty (freeze variogram, histogram and surface uncertainties).
6. Variogram uncertainty (freeze histogram, surface and HWC uncertainties).
7. Only SGS realizations of PHIE and Vsh (freeze all parameter uncertainties: histogram, variogram, surface and HWC)

Figure 7.45 shows uncertainty in HIIP for these seven cases (values are in million cubic meter (MCM)). Figure 7.46 shows a visually summary of Figure 7.45 in form of a tornado chart. The horizontal axis shows the range of P10 and P90 for each seven

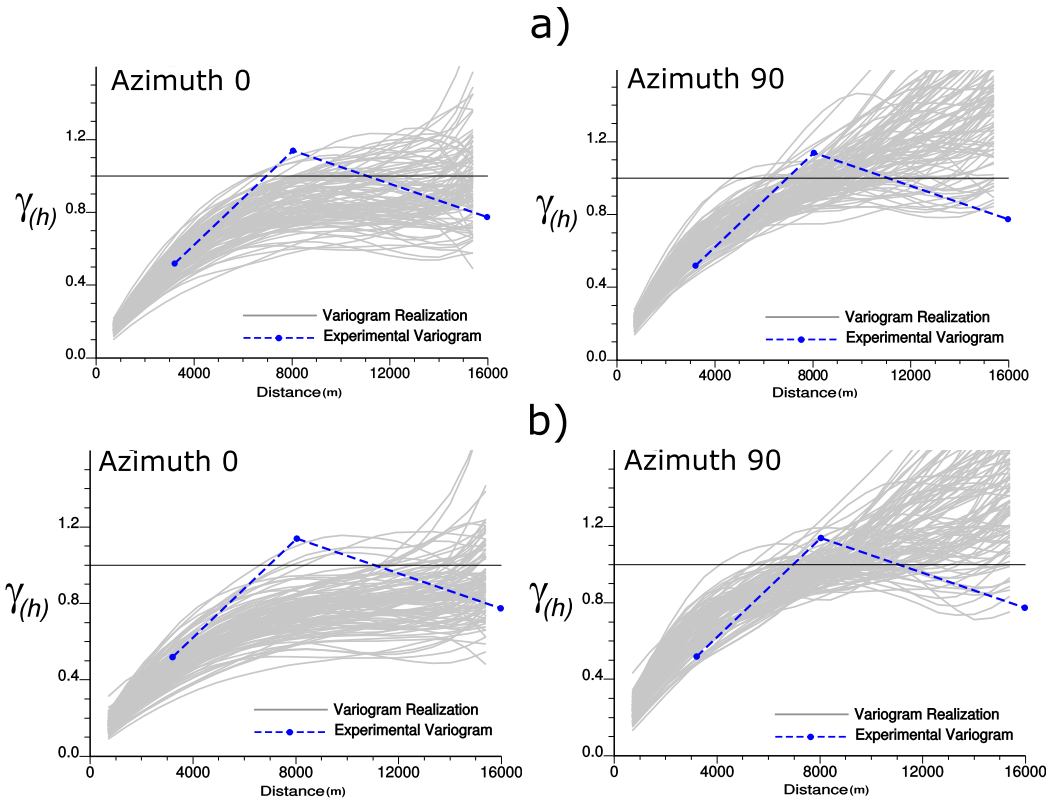

Figure 7.44: a) The posterior variogram uncertainty (100 realizations) for azimuths 0° and 90° of PHIE using the fixed histogram and variogram. b) The posterior variogram uncertainty with parameter uncertainty.

cases. The vertical axis is P10-P90. The seven cases are sorted from the highest to the lowest P10-P90. Using all uncertainties leads to the highest uncertainty in HIIP (P10-P90=797.0 MCM), and only realizations leads to the lowest uncertainty (P10-P90=390.7 MCM). Histogram plus variogram uncertainties leads to the second highest uncertainty, followed by histogram, surface, HWC and variogram uncertainties, respectively. Histogram uncertainty is the most influential parameter uncertainty in the uncertainty of HIIP, and variogram uncertainty is the least influential parameter uncertainty: variogram uncertainty is the second lowest uncertainty in HIIP, just higher than only realizations (P10-P90=422.3 MCM). Variogram uncertainty has a little impact on resource/reserve estimation due to the fact that the frequency of both high and low simulated values are increased or decreased for high and low variogram ranges in geostatistical modeling; this leads to canceling out the high and low values and the mean of the realizations does not change noticeably in terms of variogram uncertainty.

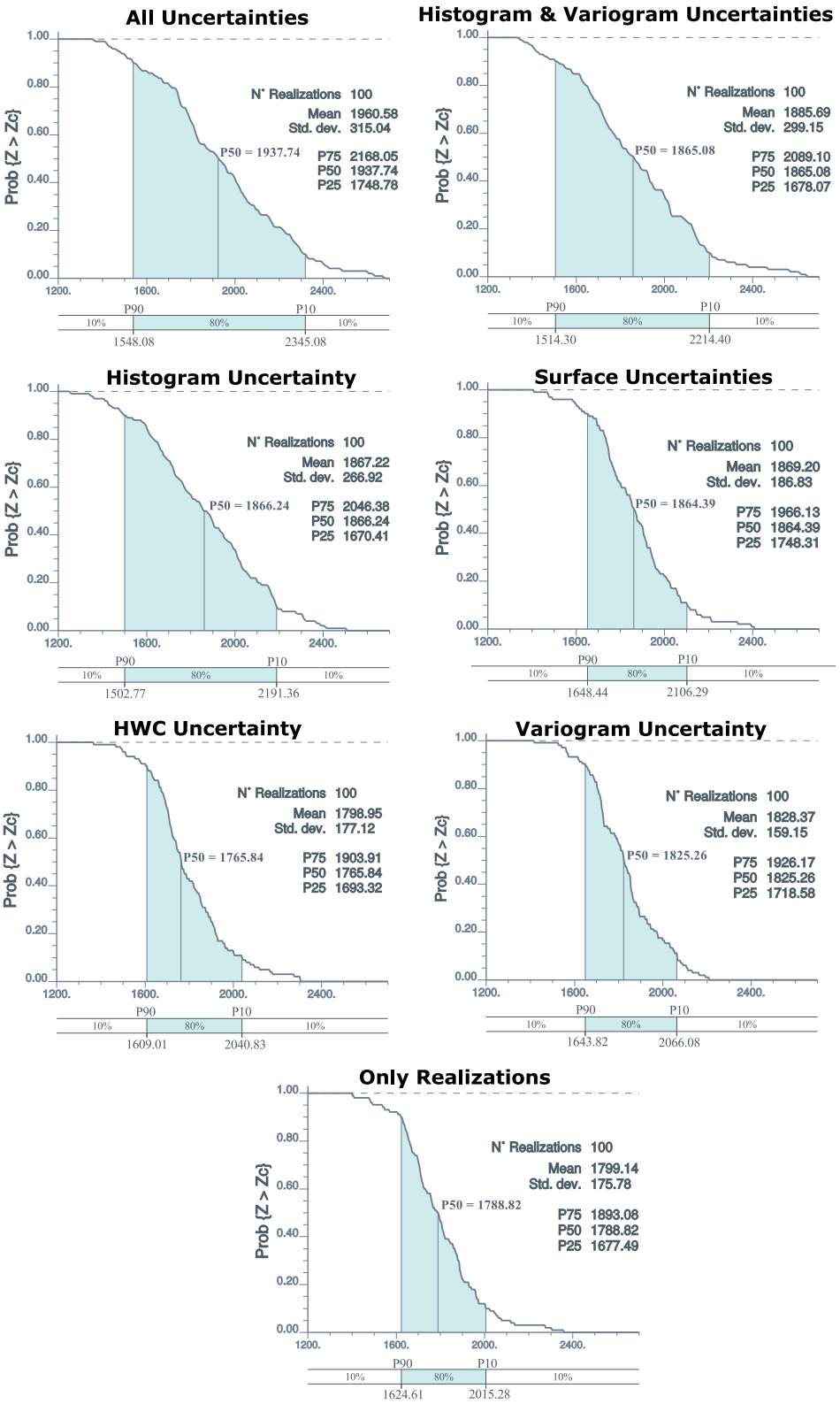

Figure 7.45: Uncertainty in HIIP for seven cases (values are in MCM).

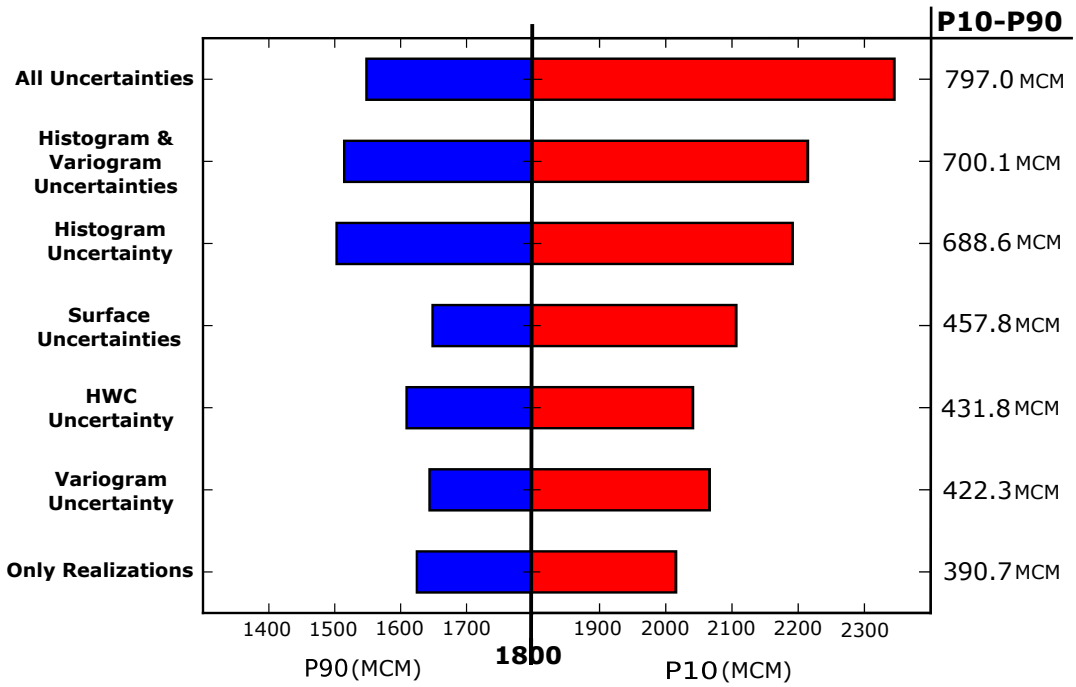

Figure 7.46: Visually summary of Figure 7.45 in the form of a tornado chart. The horizontal axis shows the range of P10 and P90 for each seven cases. The vertical axis is P10-P90.

Variogram uncertainty must have high impact on flow simulation such as oil production rate or water cut.

For this case study, uncertainty in reservoir volume (surfaces and HWC uncertainties) do not contribute significantly for HIIP uncertainty. However, when they are combined with the histogram and variogram uncertainties, it leads to higher P10 and P90 for HIIP uncertainty because of higher GRV: higher P10 and P90 for HIIP in case of all uncertainties compared with histogram and variogram uncertainties (see Figure 7.46). Since only surface uncertainties lead to higher P10 and P90, HWC uncertainty similar to variogram uncertainty has a little influence on the final model; so, they could be fixed without uncertainty for this reservoir.

7.7 Sensitivity Analysis

Sensitivity analysis in reservoir modeling enables investigating the relationship between input parameters and response variables, to detect changes in values of responses caused

by changes in the values of input parameters. The results indicate the important input parameters in the model and those of no contribution (Saltelli et al., 2008). There are several approaches of sensitivity analysis. This study applies the partial derivative of a model response with respect to input variable for calculating the standardized sensitivity coefficients (Azmy & Sartori, 2010; Balakrishnan, Castillo, & Sarabia, 2007; Saltelli et al., 2008). First order sensitivity coefficients are known as linear regression coefficients and used for sensitivity analysis of this reservoir (for more details see (Zagayevskiy & Deutsch, 2011)).

The input variables PHIE, Vsh, top, thickness with all uncertainties (see Figure 7.46) are assessed by sensitivity analysis. HWC is not included in the sensitivity analysis because according to Figure 7.46, it does not contribute to HIIP uncertainty. The variable PHIT is not also considered for sensitivity analysis because it is not used directly to calculate HIIP. The average of each realization of Vsh, top, thickness are assumed as predictors, and HIIP for each realization is the response variable. The standardized sensitivity coefficients are calculated (see (Zagayevskiy & Deutsch, 2011) for more details). Figure 7.47 shows sensitivity plot based on the values of the standardized sensitivity coefficients for linear model. R^2 is the goodness of the fitted

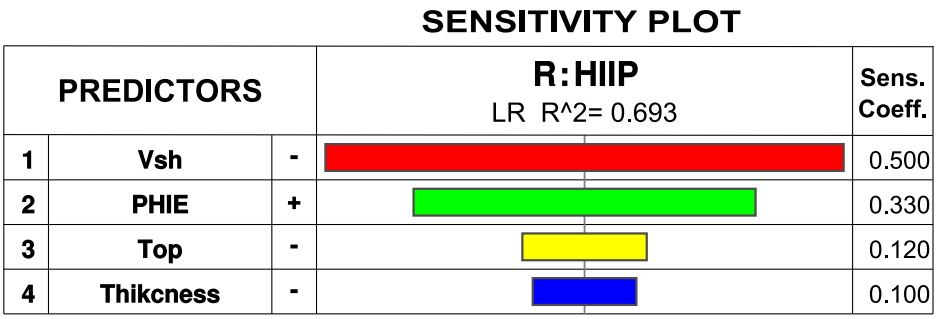

Figure 7.47: Sensitivity plot based on the values of the standardized sensitivity coefficients for a linear model.

model, which is 0.693 out of 1.0. Vsh and PHIE are the most influential variables: the standardized sensitivity coefficient for Vsh (-0.5) is higher than PHIE (0.33). This is justifiable by higher uncertainty in the histogram of Vsh than PHIE (see Figure 7.43) since histogram uncertainty has more impact on HIIP uncertainty (see Figure 7.46). Although the standardized sensitivity coefficients of the top and thickness are similar

(-0.12 vs -0.1), the uncertainty in top surface has more influence on HIIP uncertainty than thickness: similar to variogram and HWC uncertainties, thickness could also be fixed without uncertainty for this reservoir.

7.8 Remarks

The multivariate geostatistical modeling (ICC with super secondary) with parameter uncertainty is used to calculate full uncertainty in HIIP for a real case study called Netherlands Offshore F3 Block with only four wells and a 3D seismic data. The calculated parameter uncertainties are: 1- Uncertainty in GRV: uncertainty in top and thickness of the seismic data and uncertainty in HWC. 2- Variogram uncertainty: high uncertainty in the omnidirectional horizontal variogram is reduced by the seismic-derived and vertical-derived variograms. 3- Histogram uncertainty: the multivariate spatial bootstrap is used to calculate the joint prior histogram uncertainty and incorporate this uncertainty to geostatistical simulation to arrive at a final posterior histogram uncertainty.

Since there are only four wells, the experimental variogram and variogram uncertainty of top and thickness cannot be calculated; so, variogram uncertainty is not incorporated in the top and thickness models. However, a reference variogram model is required to model the top surface and thickness and quantify parameter uncertainty. This variogram could be attained from the variogram of the detrended seismic: trends in the top surface and thickness of the seismic data are modeled and removed from the seismic data ($detrend = seismic - trend$, in depth domain). This variogram is assumed as the variogram of $\Delta = actual - seismic$ required to tie the seismic surfaces (top and thickness) to the well observations and quantify parameter uncertainty.

Parameter uncertainty in the top and thickness are calculated by: 1- Tie seismic to well to achieve reference models: simple kriging is applied by the data value of $\Delta = actual - seismic$ for each well location with the variogram model of the detrend seismic ($detrend = seismic - trend$), and added to the seismic top and thickness (in depth domain) to achieve the reference structures. This eliminates the mismatch

between well and seismic for each well location; however, the mismatch could be used to quantify parameter uncertainty. 2- Calculate histogram uncertainty of Δ and use as reference distribution for SGS with conditioning data of zero value in original units. Since there are only four wells and the reference distribution of Δ is not applicable for the spatial bootstrap, a parametric approach is applied. 3- SGS realizations are back-transformed to the histogram realizations of Δ and added to the reference structures. This ensures that the realizations (top and thickness) honor the actual observations at well locations.

Due to using seismic data in multivariate geostatistical modeling, the posterior variogram uncertainty without parameter uncertainty and with parameter uncertainty are approximately the same. The posterior variogram uncertainty is influenced by the seismic data and reduced considerably if there is a high correlation between well and seismic data.

The posterior histogram realizations without using parameter uncertainty are shifted to the left or right of the reference distributions. This is resolved by unconditional SGS and conditional SGS with a pure nugget effect; hence, the shifting problem is likely related to the strong trend in well data. By incorporating the prior histogram uncertainty and increasing the uncertainty in the posterior histogram, shifting of the histogram realizations is not noticeable. Unlike posterior variogram uncertainty, using seismic data does not lead to low posterior histogram uncertainty. Only conditioning of the well data can improve the high prior histogram uncertainty. Although the posterior histogram uncertainty is lower than the prior histogram uncertainty for this case study, it is still high due to the sparse well data.

In general, histogram uncertainty is the most influential parameter uncertainty in the uncertainty of HIIP, and variogram uncertainty is the least influential one. Variogram uncertainty should have major influence on flow simulation where the connectivity of rock is critical. For example, variogram uncertainty for permeability leads to high impact on oil production rate or water cut. Since there is no permeability data, the influence of variogram uncertainty on flow simulation is not included in the case study of this Chapter.

Chapter 8

Conclusions

Numerical reservoir modeling is very important in the early stage of field development when there are limited well data. A realistic numerical model that reasonably represents the inevitable uncertainty is required at this stage for decision making. Major uncertainties in presence of limited data include the variogram and histogram. Uncertainty in these parameters should be quantified correctly and incorporated in the final reservoir model.

8.1 Summary of Contributions

The main contribution of this thesis is methodology to improve uncertainty quantification in reservoir modeling. A novel approach of variogram declustering technique is developed to remove artifacts from the experimental variogram. A new technique of variogram uncertainty is implemented. High variogram uncertainty is improved by the variogram of seismic data and the vertical variogram of the well data. A new approach of quantifying histogram uncertainty is proposed and verified by true histogram uncertainty.

8.1.1 Variogram declustering

An accurate variogram model is very important for quantifying parameter uncertainty as well as geostatistical model construction. The preferential placement of wells can lead to an unreliable experimental variogram. Weighting the pairs entering variogram

calculation by ordinary global kriging (OGK) is proposed for declustering the variogram. This is a useful tool to improve the determination of the spatial structure of a data set in the presence of preferential sampling. Global kriging is applied on the variogram pairs of each lag distance considering the fourth order covariance between pairs. The fourth order covariances comes from a preliminary variogram model that is updated once the declustering is performed. According to the synthetic and real examples, the declustered variogram is not necessarily close to the preliminary variogram; the preliminary variogram is not used directly as a target for variogram declustering. The new fitted variogram model to the declustered sample variogram is used for geostatistical modeling and quantifying uncertainty.

8.1.2 Variogram Uncertainty

There is unavoidable uncertainty associated with the experimental variogram although variogram declustering can remove some noise and artifacts. A new approach of quantifying variogram uncertainty is proposed that is independent of the domain size. The fourth order covariance is applied to directly calculate the degrees of freedom (DoF) for each lag. The Chi-square distribution of each lag distance is built by the calculated DoF. Variogram realizations are drawn from these distributions. The variogram realizations reproduce the correlation between lag distances by considering LU unconditional simulation with the correlation matrix between lag distances. These variogram realizations are used in geostatistical modeling to incorporate variogram uncertainty in the final model.

The calculated variogram uncertainty is not conditioned to the well data. By incorporating variogram uncertainty in geostatistical modeling, the initial prior uncertainty is reduced due to conditioning of the data.

8.1.3 Improve Horizontal Variogram Uncertainty by Secondary Data

In case of sparse well data, the high uncertainty in the horizontal variogram is not reduced by the conditioning data. A novel technique is developed to improve the horizontal variogram of well data by seismic data. The vertical variogram from wells

and the horizontal variogram from seismic data are well-defined in most cases, and these can be used to improve the horizontal variogram inference from well data. An approach is proposed to combine information from a noisy horizontal variogram from wells, a horizontal seismic variogram, and analogue anisotropy ratios combined with the vertical variogram from well data.

The acceptable range of the unknown horizontal variogram of the well data is attained from the horizontal seismic variogram considering the cross covariance to the well data. A rejection sampling is proposed to ensure variogram realizations of the well data fall within the limits. The limits are narrower if there is a high correlation between well and seismic data. Well and seismic data should be at the same scale to apply this methodology. In case of 2D data set, it is assumed they are scale consistent. However, for 3D data, the vertical scale of the seismic data is always larger than the vertical scale of the well data. Thus, the seismic variogram should be down scaled before applying this approach, and the correlation between well and seismic data at the scale of the well data should be achieved.

The vertical variogram of the well data can be scaled to scenarios of the horizontal variogram by considering horizontal-to-vertical anisotropy ratios (H:V). Uncertainty in the H:V comes from geological analogues. Samples are drawn from the distribution of the H:V and the sill value. The updated variogram distribution for each lag distance is calculated considering the seismic limits.

The noisy horizontal variogram and the vertical-derived variogram (after applying seismic limits) are merged by the method of combining error ellipses for independent Gaussian distributions by Blachman (1989). A transformation table is considered since the distribution of each lag distance is Chi-square and not Gaussian. The final distribution of uncertainty for each lag is reduced because it is achieved by integrating all secondary data (the vertical variogram of the well data and the horizontal variogram of the seismic data).

After merging variogram distributions, new horizontal variogram realizations are drawn from the uncertainty intervals of lag distances considering the correlation matrix of the variogram lags. The horizontal variogram realizations can be autofitted with a

fixed vertical variogram to obtain the final 3D variogram realizations.

8.1.4 Histogram Uncertainty

Quantifying histogram uncertainty is an important task for geostatistical modeling since it has a large affect on resource uncertainty. A new simulation-based approach is developed. The histogram uncertainty from the spatial bootstrap is taken as the prior uncertainty. This prior uncertainty could be conditioned and trimmed to the study area to achieve more accurate posterior histogram uncertainty. The histogram uncertainty techniques are checked against the true uncertainty in some special cases.

The true variance of the mean could be calculated by a scan-based approach. This approach looks for similar patterns of a data configuration within a large image, and calculates the mean of the similar patterns of the data configuration. In order to make a reliable distribution of the mean, the data configurations should have enough replicates within the image. By scanning many data configurations and several templates, the true uncertainty in the mean is estimated. This supports selection of the best approach of quantifying histogram uncertainty that is the posterior approach. The spatial bootstrap and global kriging are found to overestimate and underestimate the histogram uncertainty, respectively.

The multivariate spatial bootstrap is used to calculate the joint prior parameter uncertainty. The posterior histogram uncertainty for multiple variables is calculated by transferring the joint prior parameter uncertainty in multivariate geostatistical modeling. Since the multivariate spatial bootstrap respects the correlation between the variables, the uncertainty in the correlation between variables as well as statistics are incorporated in the final model. The prior histogram uncertainty in the proportion of each facies for categorical variables could also be computed and used in conditional simulation to achieve the posterior uncertainty in the proportion of each variable.

According to the case study in this thesis, histogram uncertainty is the most influential parameter uncertainty in resource estimation because it affects the mean of the realizations. Variogram uncertainty is the least influential due to the fact that both low and high simulated values average out. Variogram uncertainty has a large impact

on flow simulation because connectivity is crucial.

8.2 Limitations

Although the developed methodologies have been checked, there are some computational and conceptual limitations.

- Variogram declustering by global kriging with the fourth order moments is efficient for less than a few thousand variogram pairs of each lag distance because of CPU time. This is due to solving the covariance matrix of the global kriging for each lag distance and using the fourth order covariance. This methodology is limited to the early stages of field development with sparse sampling and less than 10000 pairs for each lag distance.
- The experimental variogram depends on input parameters including lag distance, lag tolerance, azimuth angle, azimuth tolerance, dip angle, dip tolerance and bandwidth. The calculated experimental variogram is sensitive to these parameters. This might lead to unreliable declustered sample variogram due to using the variogram model fitted to the calculated experimental variogram by choosing these parameters.
- The DoF approach of quantifying variogram uncertainty is not conditioned to well data. This leads to a high variogram uncertainty. Although conditioning of data reduces the high variogram uncertainty if there are enough wells, there is no proof that the results are close to the real variogram uncertainty. A similar check as applied to the histogram uncertainty is required.
- The CPU time is also an issue for variogram uncertainty due to the fact that the DoF technique uses the fourth order covariance of variogram pairs for calculating the degrees of freedom.
- The proposed methodology of the seismic derived variogram requires the well and seismic variogram to be scale consistent. The variogram downscaling approach assumes the variables average linearly and the shape and number of variogram

structures are unchanged after variogram downscaling. This assumption may not be correct.

- Variogram uncertainty is only calculated for direct variograms: variogram uncertainty is not considered for cross variograms. Although the uncertainty in cross variogram could be calculated by an intrinsic model considering the uncertainty in the direct variogram, it would be advisable to apply multivariate geostatistical modeling techniques that do not require the cross variogram such as the projection pursuit multivariate transformation (PPMT) (Barnett, Manchuk, & Deutsch, 2014).

There are several assumptions for the methodologies proposed in this thesis. The first assumption is that the normal score variables should be multivariate Gaussian. This is required for variogram declustering and variogram uncertainty due to using the fourth order covariance, and for histogram uncertainty because of LU simulation for the spatial bootstrap. A stationary domain is another assumption for all techniques: the variogram model fitted to the experimental variogram should reach to stationary sill. Assumption of representative variogram and histogram models are also required for sampling variogram and histogram realizations. If these representative models are not accurate after declustering (variogram declustering by fourth order moments and histogram declustering by secondary data), the drawn variogram and histogram realizations may be unreliable.

8.3 Future Work

Additional work may be considered for geostatistical modeling in presence of limited data. Some ideas for future research are presented as:

- Variogram declustering could be applied to improve the indicator variogram of each facies. The validity of the fourth order moment approach would have to be established for categorical variables.
- The methodology for improving the horizontal variogram uncertainty of contin-

uous variables could also be extended to categorical variables. The horizontal indicator variogram uncertainty for each facies may be calculated and improved by the vertical indicator variogram of the well data and horizontal variogram of the seismic data.

- The variogram downscaling technique could be improved to address nonlinear averaging and the change in variogram shape.
- The multivariate variogram realizations in this thesis is achieved by applying LU simulation with the positive correlation matrix of the variables. However, multiple variable variograms (variogram realizations) could be sampled at the same time.
- Posterior histogram uncertainty for multi variables and proportions of the categorical variable could be evaluated with true histogram uncertainties.

Only variogram and histogram uncertainty are considered in this thesis. This work could be extended to calculate uncertainty in other parameters such as formation volume factor, recovery factor and dynamic uncertainty, and incorporate them in geostatistical modeling for more accurate reservoir forecast.

References

AlRuwaili, S. B., & AlWaheed, H. H. (2004). Improved Petrophysical Methods and Techniques for Shaly Sands Evaluation. In *SPE (Society of Petroleum Engineers) Annual Technical Conference and Exhibition.*

Alshehri, N. S. (2010). *Quantification of Reservoir Uncertainty for Optimal Decision Making*. PhD Thesis. University of Alberta.

Amudo, C., Graf, T., Dandekar, R., & Randle, J. M. (2009). The Pains and Gains of Experimental Design and Response Surface Applications in Reservoir Simulation Studies. In *SPE (Society of Petroleum Engineers) Reservoir Simulation Symposium.*

Asquith, G., Krygowski, D., Henderson, S., & Hurley, N. (2004). *Basic Well Log Analysis* (Vol. 16). American Association of Petroleum Geologists.

Azmy, Y., & Sartori, E. (2010). *Nuclear Computational Science*. Springer.

Babak, O., & Deutsch, C. V. (2007a). *An Intrinsic Model of Coregionalization that Solves Variance Inflation in Colocated Cokriging*. CCG Annual Report 9-107, Centre for Computational Geostatistics, University of Alberta.

Babak, O., & Deutsch, C. V. (2007b). *Merging Multiple Secondary Data for Collocated Cokriging (recall of super secondary variable)*. CCG Annual Report 9-109, Centre for Computational Geostatistics, University of Alberta.

Babak, O., & Deutsch, C. V. (2009). Accounting for Parameter Uncertainty in Reservoir Uncertainty Assessment: The Conditional Finite-Domain Approach. *Natural resources research*, *18*(1), 7–17.

Babak, O., & Leuangthong, O. (2008). *Direct Upscaling of Variograms and Cross Var-*

iograms for Scale Consistent Geomodeling. CCG Annual Report 10-118, Centre for Computational Geostatistics, University of Alberta.

Balakrishnan, N., Castillo, E., & Sarabia, J. M. (2007). *Advances in Distribution Theory, Order Statistics, and Inference*. Springer Science & Business Media.

Barnett, R. M. (2015). *Managing Complex Multivariate Relations in the Presence of Incomplete Spatial Data*. PhD Thesis. University of Alberta.

Barnett, R. M., Manchuk, J. G., & Deutsch, C. V. (2014). Projection Pursuit Multivariate Transform. *Mathematical Geosciences*, *46*(3), 337–359.

Blachman, N. M. (1989). On Combining Target-Location Ellipses. *IEEE Transactions on Aerospace and Electronic Systems*, *25*(2), 284–287.

Bogaert, P., & Russo, D. (1999). Optimal spatial sampling design for the estimation of the variogram based on a least squares approach. *Water Resources Research*, *35*(4), 1275–1289.

Bretherton, C. S., Widmann, M., Dymnikov, V. P., Wallace, J. M., & Bladé, I. (1999). The Effective Number of Spatial Degrees of Freedom of a Time-Varying Field. *Journal of Climate*, *12*(7), 1990–2009.

Cebastiant, A., & Osbon, R. A. (2011). Experimental Design for Resource Estimation: A Comparison with the Probabilistic Method. In *SPE (Society of Petroleum Engineers) Reservoir Characterisation and Simulation Conference and Exhibition.*

Chambers, R. L., & Yarus, J. M. (2002). Quantitative Use of Seismic Attributes for Reservoir Characterization. *CSEG Recorder*, *27*(6), 14–25.

Chiles, J. P., & Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons.

Chu, J., Xu, W., & Journel, A. G. (1994). 3-D Implementation of Geostatistical Analyses–The Amoco Case Study. *AAPG Special Volumes*, *16*, 201–216.

Cressie, N. (1985). Fitting Variogram Models by Weighted Least Squares. *Mathematical Geology*, *17*(5), 563–586.

Davis, J. E. (2007). *Combining Error Ellipses.* Retrieved from http://lists.m.jedsoft.org/physics/notes/ellipse.pdf

Davis, M. W. (1987). Production of Conditional Simulations via the LU Triangular

Decomposition of the Covariance Matrix. *Mathematical Geology*, *19*(2), 91–98.

Derakhshan, S. H., & Deutsch, C. V. (2008). *Generalized Linear Distributions to Account for Parameter Uncertainty in Sequential Gaussian Simulation*. CCG Annual Report 10-104, Centre for Computational Geostatistics, University of Alberta.

Deutsch, C. V. (1989). DECLUS: A Fortran 77 Program for Determining Optimum Spatial Declustering Weights. *Computers & Geosciences*, *15*(3), 325–332.

Deutsch, C. V. (1992). *Annealing Techniques Applied to Reservoir Modeling and the Integration of Geological and Engineering (Well Test) Data*. PhD Thesis. Stanford University.

Deutsch, C. V. (2004). *A Statistical Resampling Program for Correlated Data: Spatial_Bootstrap*. CCG Annual Report 6-401, Centre for Computational Geostatistics, University of Alberta.

Deutsch, C. V. (2005). *A New TRANS Program for Histogram and Trend Reproduction*. CCG Annual Report 7-306, Centre for Computational Geostatistics, University of Alberta.

Deutsch, C. V., Frykman, P., & Xie, Y. (1999). *Declustering with Seismic or 'Soft' Geological Data*. CCG Annual Report 1-101, Centre for Computational Geostatistics, University of Alberta.

Deutsch, C. V., Frykman, P., & Xie, Y. (2005). *Programs for Debiasing and Cloud Transformation*. CCG Annual Report 7-404, Centre for Computational Geostatistics, University of Alberta.

Deutsch, C. V., & Journel, A. G. (1998). *GSLIB: Geostatistical Software Library and User's Guide*. Oxford University Press.

Deutsch, J. L., & Deutsch, C. V. (2009). *Multiple Bivariate Gaussian Plotting and Checking*. CCG Annual Report 11-403, Centre for Computational Geostatistics, University of Alberta.

Deutsch, J. L., & Deutsch, C. V. (2010). *Some Geostatistical Software Implementation Details*. CCG Annual Report 12-412, Centre for Computational Geostatistics, University of Alberta.

dGB. (n.d.). Retrieved from https://www.dGBes.com (dGB Earth Sciences B.V.)

Doyen, P. M., Den Boer, L. D., & Pillet, W. R. (1996). Seismic Porosity Mapping in the Ekofisk Field Using a New Form of Collocated Cokriging. In *SPE (Society of Petroleum Engineers) Annual Technical Conference and Exhibition* (pp. 21–30).

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, *7*(1), 1–26.

Ehinola, O. A., & Akinbodewa, O. S. (2014). Uncertainty in Quantification and Ranking of Stock Tank Oil Initially in Place (STOIIP) in OA Field, Niger Delta. In *AAPG Unconventional Resources Technology Conference (URTEC).*

Emery, X., & Ortiz, J. M. (2005). Histogram and variogram inference in the multigaussian model. *Stochastic Environmental Research and Risk Assessment*, *19*(1), 48–58.

Emery, X., & Ortiz, J. M. (2007). Weighted sample variograms as a tool to better assess the spatial variability of soil properties. *Geoderma*, *140*(1), 81–89.

Etherington, J., Pollen, T., & Zuccolo, L. (2005). Comparison of Selected Reserves and Resource Classifications and Associated Definitions. *Society of Petroleum Engineers, Final Report-December 2005*.

Frykman, P., & Deutsch, C. V. (2002). Practical Application Of Geostatistical Scaling Laws For Data Integration. *Petrophysics*, *43*(03).

Gantmacher, F., & Krein, N. (1950). *Oscillation Matrices and Small Vibrations of Mechanical Systems.* Gostekhizdat, Moscow (in Russian).

Genton, M. G. (1998). Variogram Fitting by Generalized Least Squares Using an Explicit Formula for the Covariance Structure. *Mathematical Geology*, *30*(4), 323–345.

Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation.* Oxford University Press.

Goovaerts, P. (1998). Ordinary Cokriging Revisited. *Mathematical Geology*, *30*(1), 21–42.

Gringarten, E., & Deutsch, C. V. (2001). Teacher's Aide Variogram Interpretation and Modeling. *Mathematical Geology*, *33*(4), 507–534.

Hall, M. (2006). Predicting bed thickness with cepstral decomposition. *The Leading Edge*, *25*(2), 199–204.

Hall, M., & Trouillot, E. (2004). Predicting stratigraphy with spectral decomposition. In *CSEG National Convention.*

Higham, N. J. (1988). Computing a Nearest Symmetric Positive Semidefinite Matrix. *Linear Algebra and its Applications*, *103*, 103 - 118. Retrieved from http://www.sciencedirect.com/science/article/pii/0024379588902236

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous Univariate Distributions. Fifth Edition.* John Wiley and Sons.

Johnson, R. A., & Wichern, D. W. (2002). *Applied Multivariate Statistical Analysis.* Prentice-Hall, Inc.

Journel, A. G. (1983). Nonparametric Estimation of Spatial Distributions. *Mathematical Geology*, *15*(3), 445–468.

Journel, A. G., & Bitanov, A. (2004). Uncertainty in N/G ratio in early reservoir development. *Journal of Petroleum Science and Engineering*, *44*(1), 115–130.

Journel, A. G., & Xu, W. (1994). Posterior Identification of Histograms Conditional to Local Data. *Mathematical Geology*, *26*(3), 323–359.

Kaleta, M. P., Bennett, R., Brint, J., Van Den Hoek, P., Van Doren, J., Van Essen, G., & Woodhead, T. J. (2012). Coupled Static/Dynamic Modeling for Improved Uncertainty Handling. In *74th EAGE Conference & Exhibition.*

Kalkomey, C. T. (1997). Potential risks when using seismic attributes as predictors of reservoir properties. *The Leading Edge*, *16*(3), 247–251.

Khan, K. D., & Deutsch, C. V. (2016). Practical Incorporation of Multivariate Parameter Uncertainty in Geostatistical Resource Modeling. *Natural Resources Research*, *25*(1), 51–70.

Khan, K. D., Deutsch, J. L., & Deutsch, C. V. (2014). *A Refresher on Large Scale Uncertainty for Resource Estimation*. CCG Annual Report 16-119, Centre for Computational Geostatistics, University of Alberta.

Koushavand, B., Ortiz, J. M., & Deutsch, C. V. (2008). *A Methodology to Quantify and Transfer Variogram Uncertainty through Kriging and Simulation*. CCG Annual

Report 10-310, Centre for Computational Geostatistics, University of Alberta.

Kovitz, J., & Christakos, G. (2004). Spatial statistics of clustered data. *Stochastic Environmental Research and Risk Assessment*, *18*(3), 147–166.

Kupfersberger, H., & Deutsch, C. V. (1999). Methodology for Integrating Analog Geologic Data in 3-D Variogram Modeling. *AAPG Bulletin*, *83*(8), 1262–1278.

Kupfersberger, H., Deutsch, C. V., & Journel, A. G. (1998). Deriving Constraints on Small-scale Variograms due to Variograms of Large-Scale Data. *Mathematical Geology*, *30*(7), 837–852.

Larrondo, P., Neufeld, C., & Deutsch, C. V. (2003). *VARFIT: A Program for Semi-Automatic Variogram Modelling*. CCG Annual Report 5-122, Centre for Computational Geostatistics, University of Alberta.

Manchuk, J. G., & Deutsch, C. V. (2012). A flexible sequential Gaussian simulation program: USGSIM. *Computers & Geosciences*, *41*, 208–216.

Marchant, B. P., & Lark, R. M. (2004). Estimating Variogram Uncertainty. *Mathematical Geology*, *36*(8), 867–898.

Matheron, G. (1965). *Les variables régionalisées et leur estimation*. Masson et Cie. Editeurs, Paris.

McVay, D. A., & Dossary, M. N. (2014). The Value of Assessing Uncertainty. *SPE (Society of Petroleum Engineers) Economics & Management*, *6*(02), 100–110.

Meddaugh, W. S., Barge, D. L., Todd, W. W., & Griest, S. (2007). The Jurassic-age Marrat Reservoir at Humma Field Partitioned Neutral Zone (PNZ) Saudi Arabia and Kuwait-Utilization of a Probabilistic Two Stage Design of Experiments-based Workflow for Reservoir Characterization and Management. In *SPE (Society of Petroleum Engineers) International Petroleum Technology Conference.*

Meddaugh, W. S., Champenoy, N., Osterloh, W., & Tang, H. (2011). Reservoir Forecast Optimism–Impact of Geostatistics, Reservoir Modeling, Heterogeneity, and Uncertainty. In *SPE (Society of Petroleum Engineers) Annual Technical Conference and Exhibition.*

Meddaugh, W. S., Griest, S., & Barge, D. (2009). Quantifying Uncertainty in Carbonate Reservoirs-Humma Marrat Reservoir Partitioned Neutral Zone (PNZ) Saudi

Arabia and Kuwait. In *SPE (Society of Petroleum Engineers) Middle East Oil and Gas Show and Conference.*

Meddaugh, W. S., Gross, S. J., Griest, S. D., Todd, W., & Barge, D. (2006). Impact of Volumetric and Connectivity Uncertainty on Reservoir Management Decisions: Case Study from the Humma Marrat Reservoir, Partitioned Neutral Zone. *Society for Sedimentary Geology*.

Menezes, R., Garcia-Soidán, P., & Febrero-Bande, M. (2007). A kernel variogram estimator for clustered data. *Scandinavian Journal of Statistics*, *35*(1), 18–37.

Merrow, E. W. (2011). Oil industry megaprojects: Our recent track record. In *SPE (Society of Petroleum Engineers) Offshore Technology Conference.*

Morris, P. A. (1974). Decision Analysis Expert Use. *Management Science*, *20*(9), 1233–1241.

Morris, P. A. (1977). Combining Expert Judgments: A Bayesian Approach. *Management Science*, *23*(7), 679–693.

Murtha, J. A. (1997). Monte Carlo Simulation: Its Status and Future. *Journal of petroleum technology*, *49*(4), 361–374.

NASA. (n.d.). Retrieved from http://oceancolor.gsfc.nasa.gov/cgi/image_archive.cgi?c=CHLOROPHYLL

Neufeld, C., & Wilde, B. (2005). *A Global Kriging Program for Artifact-Free Maps*. CCG Annual Report 7-403, Centre for Computational Geostatistics, University of Alberta.

NIST. (n.d.). Retrieved from http://www.itl.nist.gov/div898/handbook/eda/section3/eda366b.htm (Engineering and Statistics Handbook)

Olea, R. A. (2007). Declustering of Clustered Preferential Sampling for Histogram and Semivariogram Inference. *Mathematical Geology*, *39*(5), 453–467.

Omre, H. (1984). The Variogram and its estimation. *Geostatistics for natural resources characterization*, 107–125.

OpendTect. (n.d.). Retrieved from https://opendtect.org/osr/pmwiki.php/Main/NetherlandsOffshoreF3BlockComplete4GB (OpendTect Software Website)

Orechovesky, J. R. (1996). *Single Source Error Ellipse Combination*. Master Thesis.

Naval Postgraduate School. United States Navy.

Ortiz, J., & Deutsch, C. V. (2002). Calculation of Uncertainty in the Variogram. *Mathematical Geology*, *34*(2), 169–183.

Pardo-Igúzquiza, E., & Dowd, P. (2001). Variance–Covariance Matrix of the Experimental Variogram: Assessing Variogram Uncertainty. *Mathematical Geology*, *33*(4), 397–419.

Partyka, G., Gridley, J., & Lopez, J. (1999). Interpretational applications of spectral decomposition in reservoir characterization. *The Leading Edge*, *18*(3), 353–360.

Peng, C. Y., & Gupta, R. (2004). Experimental Design and Analysis Methods in Multiple Deterministic Modelling for Quantifying Hydrocarbon In-Place Probability Distribution Curve. In *SPE (Society of Petroleum Engineers) Asia Pacific Conference on Integrated Modelling for Asset Management.*

Pu, X., & Tiefelsdorf, M. (2015). A variance-stabilizing transformation to mitigate biased variogram estimation in heterogeneous surfaces with clustered samples. *13th International Conference on GeoComputation Geospatial Information Sciences, University of Texas at Dallas, Richardson, Texas, USA*.

Pyrcz, M. J., & Deutsch, C. V. (2014). *Geostatistical Reservoir Modeling. Second Edition.* Oxford University Press.

Qu, J., & Deutsch, C. V. (2014). *Modeling Non-stationary Mean and Standard Deviation on Continuous Variable*. CCG Annual Report 16-110, Centre for Computational Geostatistics, University of Alberta.

Rezvandehy, M. (2014). Logical depth modeling of a reservoir layer with the minimum available data-integration geostatistical methods and seismic attributes. *Journal of Unconventional Oil and Gas Resources*, *7*, 11–21.

Rezvandehy, M., Aghababaei, H., & Raissi, S. (2011). Integrating seismic attributes in the accurate modeling of geological structures and determining the storage of the gas reservoir in Gorgan Plain (North of Iran). *Journal of Applied Geophysics*, *73*(3), 187–195.

Rezvandehy, M., & Deutsch, C. V. (2014a). *Combining Multivariate Gaussian Distributions from Different Sources*. CCG Annual Report 16-127, Centre for Compu-

tational Geostatistics, University of Alberta.

Rezvandehy, M., & Deutsch, C. V. (2014b). *A New Program for Calculating Variogram of Gridded Data*. CCG Annual Report 16-408, Centre for Computational Geostatistics, University of Alberta.

Rezvandehy, M., & Deutsch, C. V. (2014c). *Theoretical and Numerical Framework for Improving Variogram of High Resolution Well data by Different Sources*. CCG Annual Report 16-206, Centre for Computational Geostatistics, University of Alberta.

Rezvandehy, M., & Deutsch, C. V. (2016). *"VarUn", A Fortran code for Estimation Correct Variogram Uncertainty and Transfer to Geostatistical Modeling*. CCG Annual Report 18-404, Centre for Computational Geostatistics, University of Alberta.

Richmond, A. (2002). Two-point declustering for weighting data pairs in experimental variogram calculations. *Computers & geosciences*, *28*(2), 231–241.

Robinson, J. G., & Elliott, D. (2004). National Instrument 51-101 (NI 51-101) Reserves Reconciliation-Part 1. *Journal of Canadian Petroleum Technology*, *43*(11), 6–12.

Roecker, J. (1991). On Combining Multidimensional Target Location Ellipsoids. *IEEE Transactions on Aerospace Electronic Systems*, *27*, 175–177.

Rose, P. R. (2004). Delivering on our E&P promises. *The Leading Edge*, *23*(2), 165–168.

Rose, P. R. (2007). Measuring what we think we have found: Advantages of probabilistic over deterministic methods for estimating oil and gas reserves and resources in exploration and production. *AAPG Bulletin*, *91*(1), 21–29.

Salinas, R., Di Nezio, A., & Huerta Petroperu, V. (2014). Uncertainty Analysis and Risk Assessment Methodology in Early Development Fields. In *SPE (Society of Petroleum Engineers) Latin America and Caribbean Petroleum Engineering Conference.*

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., & Saisana, M. (2008). *Global sensitivity analysis: The primer.* John Wiley & Sons.

Samson, P., Dubrule, O., & Euler, N. (1996). Quantifying the impact of structural uncertainties on gross-rock volume estimates. In *SPE (Society of Petroleum Engineers) European 3-D reservoir modelling conference* (pp. 381–392).

Singh, V., Hegazy, M., & Fontanelli, L. (2009). Assessment of reservoir uncertainties for development evaluation and risk analysis. *The Leading Edge*, *28*(3), 272–282.

Singh, V., Yemez, I., & Sotomayor, J. (2013). Key Factors Affecting 3D Reservoir Interpretation and Modelling Outcomes: Industry Perspectives. *British Journal of Applied Science & Technology*, *3*(3), 376–405.

Solow, A. R. (1985). Bootstrapping correlated data. *Mathematical Geology*, *17*(7), 769–775.

Vandebril, R., Van Barel, M., & Mastronardi, N. (2008). *Matrix computations and semiseparable matrices: Linear systems (volume 1).* The Johns Hopkins University Press.

Van Elk, J. F., Guerrera, L., Vijayan, K., & Gupta, R. (2000). Improved Uncertainty Management in Field Development Studies through the Application of the Experimental Design Method to the Multiple Realisations Approach. In *SPE (Society of Petroleum Engineers) Asia Pacific Oil and Gas Conference and Exhibition.*

Vis, G. J., Van Gessel, S. F., MijnLieff, H. F., Pluymaekers, M. P. D., Hettelaar, J. M. M., & Stegers, D. P. M. (2010). *Lower Cretaceous Rijnland Group aquifers in the West Netherlands Basin: suitability for geothermal energy. TNO-034-UT-2009-02410.* Retrieved from http://docplayer.net/17795608-Lower-cretaceous-rijnland-group-aquifers-in-the-west-netherlands-basin-suitability-for-geothermal-energy.html

Von Eye, A., & Bogat, G. A. (2004). Testing the assumption of multivariate normality. *Psychology Science*, *46*, 243–258.

Wang, F., & Wall, M. M. (2003). Incorporating parameter uncertainty into prediction intervals for spatial data modeled via a parametric variogram. *Journal of agricultural, biological, and environmental statistics*, *8*(3), 296–309.

Wang, J., & Dou, Q. (2010). Integration of 3D Seismic Attributes into Stochastic Reservoir Models Using Iterative Vertical Resolution Modeling Methodology. In

SPE (Society of Petroleum Engineers) Western Regional Meeting.

Webster, R., & Oliver, M. A. (1992). Sample adequately to estimate variograms of soil properties. *European Journal of Soil Science*, *43*(1), 177–192.

Winkler, R. L. (1968). The Consensus of Subjective Probability Distributions. *Management Science*, *15*(2), 61–75.

Wolff, M. (2010). Probabilistic Subsurface Forecasting-What Do We Really Know? *Journal of Petroleum Technology*, *62*(05), 86–92.

Zagayevskiy, Y., & Deutsch, C. V. (2011). *Updated Code for Sensitivity Analisis Based on Regression*. CCG Annual Report 13-401, Centre for Computational Geostatistics, University of Alberta.

Appendix A

Programs

This section provides documentation and description of the programs written in the thesis for variogram calculation, variogram uncertainty and variogram improvement. The programs include *gamt* for calculating the experimental variogram of exhaustive data (seismic data) and the cross covariance between well data and seismic data, *Vardec* for variogram declustering, *VarUn* for variogram uncertainty, *VarSeis* for the seismic-derived variogram, *VarVtH* for the vertical-derived variogram and *VarMerg* for final improved horizontal variogram realizations by applying the seismic-derived variogram and merge with the vertical-derived variogram.

A.1 Program *gamt*

The program *gamt* provides more accurate experimental variogram of the seismic data and cross covariance between well and seismic data with lower CPU time than available alternatives. The methodology is mentioned in Figure 4.11 of Chapter 4. The experimental variogram of exhaustive seismic data and the cross covariance are very important for calculating the seismic-derived variogram (Chapter 4). Due to large number of seismic data, variogram and cross covariance calculation could lead to high CPU time. The GSLIB code *gam* for gridded data (Deutsch & Journel, 1998) only calculates variogram for certain directions and lags and cannot calculate the cross covariance between sparse well data and exhaustive seismic data. The new program *gamt* for variogram and cross covariance calculation is implemented based on a predefined

template around each location of variogram (or cross covariance) pair and choosing the nearest points to this location for variogram calculation (see Figure 4.11 of Chapter 4). This process is repeated for all lag distances to calculate the experimental variogram (For more information see Rezvandehy and Deutsch (2014b)).

Table A.1 shows the default parameter file of the program *gamt*. **Line 5** specifies the gridded data file which is usually the seismic data. This data should be a simplified Geo-EAS formated file (Deutsch & Journel, 1998). **Line 6** defines number of variables in the seismic data file (the gridded data) followed by column number for each variable. Trimming limits are specified on **Line 7**. Grid specification for the seismic data are entered on **Lines 8** to **10**. **Line 11** denotes the primary data file (well data). This file is only required for calculating the cross covariance. **Lines 12** and **13** specify x, y, z columns, and number of variables and column number for each variable, respectively.

Table A.1: Parameter file of *gamt* program.

```
 1                  Parameters for gamt
 2                  ******************
 3
 4  START OF PARAMETERS:
 5  Seismic.dat          -file with gridded data
 6  2  1  2             -  number of variables, column numbers
 7  -100   1.0e21        -  trimming limits
 8  50  0.5  1.0         -  nx, xmn, xsiz
 9  50  0.5  1.0         -  ny, ymn, ysiz
10  1   0.5  1.0         -  nz, zmn, zsiz
11  well.dat             -file with well/drillhole data (optional)
12  1  2  3              -  columns for X, Y and Z coordinates
13  1  4                 -  number of variables, column numbers
14  gamt.out             -file for variogram output
15  2                    -number of directional variograms
16  50.0   0.0           -  azimuth and dip
17  140.0  0.0           -  azimuth and dip
18  25 1.0 10            -number of lags, lag distance, tolerance number
19  1                    -standardize sill? (0=no, 1=yes)
20  7                    -number of variograms
21  1  1  1              -tail variable, head variable, variogram type
22  2  2  1              -tail variable, head variable, variogram type
23  1  2  2              -tail variable, head variable, variogram type
24  1  1  3              -tail variable, head variable, variogram type
25  2  2 -3              -tail variable, head variable, variogram type
26 -1  1  3              -tail variable, head variable, variogram type
27  1 -2 -3              -tail variable, head variable, variogram type
```

Line 14 is the output file for variogram calculation. **Line 15** is number of directional variograms followed by **Lines 16** and **17** defining the azimuth and dip of each direc-

tional variogram. Number of lags, lag distance and tolerance number are entered on **Line 18**. The tolerance number is the acceptable number of the head locations in the defined template for variogram (cross covariance) calculation, and the template is constructed based on this tolerance number. **Line 19** indicates standardization of the calculated variogram (yes=1, no=0). Number of variograms are entered on **Line 20**. The following **Lines 21** to **27** define the tail and head variable, and variogram type for each variogram number, respectively. Variogram type 1 is the traditional variogram, so there should be the same number for both head and tail variables. **Line 21** is the traditional variogram for the first variable, and **Line 22** for the second variables of the seismic data. The tail and head variables should be positive for variogram type 1. Variogram type 2 signifies the traditional cross variogram between two variables of the seismic data; the tail and head variables should be different (**Line 23**). Variogram type 3 denotes the covariance of the seismic data (**Line 24**) if the tail and head variables are the same and positive. Negative variogram type 3 (-3) means converting the calculated covariance to variogram through subtracting the covariance by the stationary variance. The cross covariance between the well data and seismic data is calculated if one of the tail or the head variables. In this case, the tail variable is well data and the head variable is the seismic data. For example, **Line 26** is the cross covariance between the first variable of the well data and the first variable of the seismic data. -3 for variogram type in case of the cross covariance calculation denotes converting the cross covariance to the cross variogram. **Line 27** shows the cross variogram between the first variable of the well data and the second variable of the seismic data. Variogram type 4 signifies the correlogram.

A.2 Program *Vardec*

The Program *Vardec* is developed for variogram declustering based on the methodology in Chapter 2; ordinary global kriging is applied on the variogram pairs of each lag distance to give each pair an appropriate weight. Table A.2 shows the default parameter file. The data file is entered on **Line 5**. This data should be a simplified Geo-EAS

formated file (Deutsch & Journel, 1998). **Line 6** defines x, y and z coordinates. **Line 7** indicates column number of the variable for variogram declustering. Trimming limits are entered on **Line 8**. **Lines 9** to **11** define grid specification for the area of interest for variogram declustering. **Line 12** specifies grid increment for discretization the study area into variogram pairs for each lag distance (see Figure 2.3 of Chapter 2). The discretization of variogram pairs are used in the left hand side covariance matrix for global kriging approach. Figure A.1 shows a schematic illustration of discretization

Table A.2: Parameter file of *Vardec* program.

```
1                    Parameters for Vardec
2                  ***************************
3
4   START OF PARAMETERS:
5   Data.dat                 - file with data
6   1   2   0                - columns for X, Y, Z coordinates
7   4                        - col number for variable
8  -100   1.0e21             - trimming limits
9   50  0.5  1.0             - nx, xmn, xsiz
10  50  0.5  1.0             - ny, ymn, ysiz
11  50  0.5  1.0             - nz, zmn, zsiz
12  3                        - grid increment , 1 all blocks
13  10                       - number of lags
14  5.0                      - lag separation distance
15  2.5                      - lag tolerance
16  1                        - number of directions
17  0 90 50 0 90  50         - azm,atol,bandh,dip,dtol,bandv
18  1                        - standardize variograms yes(1)/no(0)
19  gam.out                  - output file for declustered variogram
20  vargplt.ps               - output file for variogam ploting
21  varfit.var               - variogram model for declustered variogram
22  1  0                     - nugget effect
23  1  1  0  0  0            - it,cc,ang1,ang2,ang3
24  25 25 10                 - a_hmax, a_hmin, a_vert
```

of an area into 16 variogram pairs for lag distance **h**, azimuth approximately 35° and grid increment of 2. The study area in this Figure has $nx = 9 \times ny = 9$ grids. If grid increment is chosen to 1, all possible variogram pairs (56 pairs) are attained. In case of large grids (say bigger than $nx = 100 \times ny = 100$), grid increment 1 leads to high CPU time. **Lines 13** to **15** specify parameters for variogram calculation: number of lags (**Line 13**), lag distance (**Line 14**) and lag tolerance (**Line 15**). **Lines 16** to **17** define the direction of variogram calculation. **Line 18** indicates variogram standardization (yes=1, no=0) for both the experimental and the declustered variograms (see Section

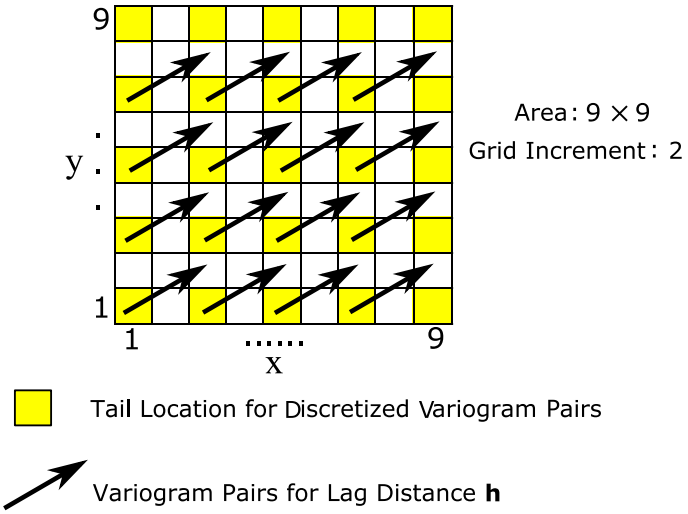

Figure A.1: A schematic illustration of discretization of an area into 16 variogram pairs for lag distance **h** and azimuth approximately 35°.

2.8 of Chapter 2). **Line 19** is the output file of variogram declustering. **Line 20** is the output file for variogram plotting. The experimental variogram, the declustered sampled variogram and the new fitted variogram to the declustered variogram are plotted. This fitted variogram model is calculated by embedding *Varfit* (Larrondo et al., 2003) software in *Vardec* program. **Line 21** is the variogram model fitted to the declustered sampled variogram, which can be used in geostatistical modeling. **Line 22** to **24** are the base case variogram model fitted to the experimental variogram before variogram declustering. This variogram model is required for estimating the fourth order covariance (see Chapter 2).

A.3 Program *VarUn*

The program *VarUn* provides variogram uncertainty and draws variogram realizations from the uncertainty interval of lag distances (see Chapter 3). Table A.3 shows the default parameter file. Although the most reliable approach of variogram uncertainty is probably the DoF approach (Chapter 3), FOM and global kriging approaches (both SGK and OGK) are also included in this code for further comparison and research. *VarUn* program by default calculates the variogram uncertainty by FOM and DoF approaches. SGK and OGK are optional (see **Line 9**). A base case variogram model

is required for all approaches. The experimental variogram realizations are modeled automatically inside the code by embedding *varfit* GSLIB code (Larrondo et al., 2003). The data file is given on **Line 5**. This data should be a simplified Geo-EAS formated file (Deutsch & Journel, 1998). **Line 6** defines x, y and z coordinates. **Line 7** indicates column number of the variable for calculating variogram uncertainty. Trimming limits are entered on **Line 8**. **Line 9** specifies variogram uncertainty by global kriging (yes=1, no=0). If variogram uncertainty by global kriging is applied, simple kriging is 1 and ordinary kriging is 0. **Lines 10** to **12** define grid specification for the area of interest for variogram uncertainty only by global kriging. **Line 13** denotes grid

Table A.3: Parameter file of *VarUn* Program.

```
1                        Parameters for VarUn
2                   **************************
3
4   START OF PARAMETERS:
5   Data.dat                 - file with data
6   1  2  3                  - columns for X, Y, Z coordinates
7   4                        - col number for variable
8   -1.0e21   1.0e21         - trimming limits
9   0    1                   - global kriging, simple(1)/ordinary(0)
10  50  0.5  1.0             - nx, xmn, xsiz
11  50  0.5  1.0             - ny, ymn, ysiz
12  50  0.5  1.0             - nz, zmn, zsiz
13  2                        - grid increment, 1 considering all blocks
14  10                       - number of lags
15  5.0                      - lag separation distance
16  2.5                      - lag tolerance
17  2                        - number of directions
18  0  10 10 0 90 50         - azm,atol,bandh,dip,dtol,bandv (major)
19  20 10 10 0 90 50         - azm,atol,bandh,dip,dtol,bandv (minor)
20  1                        - standardize experimental variogram
21  1                        - standardize variogram realizations
22  12521                    - random number seed
23  100                      - number of variogram realizations
24  1                        - range scale
25  0  1.3                   - variogram limits for variogram plot
26  corr.out                 - correlation matrix between lag distance
27  FOM.out                  - variogram realizations (FOM)
28  FOM.ps                   - file for variogam ploting (FOM)
29  DoF.out                  - variogram realizations (DOF)
30  DoF.ps                   - file for variogram ploting (DOF)
31  GK.out                   - variogram realizations (global kriging)
32  GK.ps                    - file for variogam ploting (global kriging)
33  1  0                     - nst,nugget effect
34  1  1  0  0  0            - it,cc,ang1,ang2,ang3
35  35  35  10               - a_hmax, a_hmin, a_vert
```

increment for discretization the study area into variogram pairs (only global kriging).

Lines 14 to **16** specify parameters for variogram calculation: number of lags (**Line 14**), lag distance (**Line 15**) and lag tolerance (**Line 16**). The directions of the variogram calculation are entered on **Lines 17** to **19**. Variogram uncertainty is applied for one direction in case of isotopic or omnidirectional variograms, and for two directions in case of anisotropic variograms which only major and minor directions of continuity are needed. For anisotropic variogram, the first direction should be the major direction of continuity and the second direction is minor direction of continuity. **Lines 20** and **21** indicate variogram standardization (yes=1, no=0) for the experimental variogram and variogram realizations as it is discussed in Chapter 3. **Line 22** is random number seed for simulating variogram realizations. **Line 23** defines number of variogram realizations. Autofitting software could lead to artifacts for modeling the experimental variogram realizations: in case of more than one variogram structure, the range of the first variogram structure may not be very low or very high. In order to avoid this, a_hmax (variogram range for major direction) and a_hmin (variogram range for minor direction) of the first variogram structure for all variogram realizations could be enforced to be within the scaled a_hmax and a_hmin of the first variogram structure of the base case variogram model. **Line 24** defines this scale. No correction is applied if scale is 1. **Line 25** defines variogram realization limits for the variogram plot. **Line 26** is output file for the correlation matrix between lag distances. **Lines 27** to **32** specify the output file of the variogram models and plots of the variogram realization by FOM (**Lines 27 and 28**), DoF (**Lines 29 and 30**), SGK or OGK (**Lines 31 and 32**). The variogram plots give variogram realizations before (top variogram) and after variogram fitting (bottom variogram) for each direction. The variogram models could be used in geostatistical simulation to incorporate variogram uncertainty. The base case variogram model is entered on **Line 33** to **35**. This variogram model is required for estimating the fourth order covariance of all variogram uncertainty techniques.

A.4 Program *VarSeis*

The program *VarSeis* calculates the seismic-derived variogram based on the methodology in Section 4.3 (Chapter 4). Table A.4 shows the default parameter file. This program calculates the upper and lower limits of the seismic-derived variogram. **Line 5** specifies the output file of the upper and lower limits of the seismic-derived variogram. **Line 6** indicates the plots of the upper and lower limits variograms plus the seismic variogram and the cross covariance. **Lines 7** and **8** define number of lags and lag distance. The cross covariance between well and seismic data can be calculated by *gamt* (see Section A.1). If there are insufficient well data, the cross covariance could be calculated by Markov model (Pyrcz & Deutsch, 2014). **Line 9** asks if Markov model is used for calculating cross covariance (yes=1, no=0). If so, the correlation coefficient between well and seismic data is required. **Line 10** is the percentage of the maximum positive determinant as minimum acceptable determinant (δ in Figure 4.9 of Chapter 4). **Line 11** defines the probability of the upper and lower limits taken from the cumulative distribution functions (CDF) of the positive determinants. **Lines 16** to **18** are the variogram model for the seismic data, and **Lines 23** to **25** are the cross covariance model if Markov model is not used.

A.5 Program *VarVtH*

The program *VarVtH* is implemented to calculate the vertical-derived variogram based on the methodology in Chapter 5, Section 5.3. Table A.5 shows the default parameter file. **Lines 5** to **6** specify the output files for variogram models and plot of variogram realizations, respectively. The variogram models (realizations) could be directly used in geostatistical modeling if the horizontal variogram cannot be calculated due to limited well data (less than 4 wells). **Lines 7** to **8** indicate number of lags and lag distance. **Line 9** defines the deviation from the sill 1, which is Δ_v in Equation 5.4 (Chapter 5). The minimum and maximum of the horizontal to vertical anisotropy ratio (H:V) for building a triangular distribution are entered on **Line 10**. **Line 11** is random number seed for simulating variogram realizations. **Line 12** is number of variogram realizations.

Table A.4: Parameter file of *VarSeis* Program.

```
 1                          Parameters for VarSeis
 2                        ***************************
 3
 4   START OF PARAMETERS:
 5   Seisvar.dat          - output file for seismic limits
 6   Seisvar.ps           - plot of seismic limits
 7   200                  - number of lags
 8   0.2                  - lag distance
 9   0     0.6            - Markov model 1/0, correlation coefficient
10   0.0                  - percentage of maximum positive determinant
11   10    90             - min and max probabilities
12
13   -------------------Seismic Variogram Model ---------------------
14
15   Seismic Variogram Model
16   1  0                 - nst,nugget effect
17   1   1   0  0  0      - it,cc,ang1,ang2,ang3
18   30  30 30            - a_hmax, a_hmin, a_vert
19
20   ------------------ Cross Covariance Model ---------------------
21
22   Cross Covariance Model
23   1  -0.65            - nst,nugget effect
24   1   0.65  0  0  0    - it,cc,ang1,ang2,ang3
25   28  28   28          - a_hmax, a_hmin, a_vert
```

Lines 13 to **14** are parameters for the vertical variogram such as number of variogram structures (nst), variogram type (it), variance contribution (cc) and variogram range for each structure (range).

Table A.5: Parameter file of *VarVtH* Program.

```
 1                       Parameters for VarVtH
 2                     ***************************
 3
 4   START OF PARAMETERS:
 5   Vertvar.var                - output file of variogram realizations
 6   Vertvar.ps                 - plot of variogram realizations
 7   100                        - number of lags
 8   18                         - lag distance
 9   0.1                        - deviation from the sill 1
10   75   125                   - Min and Max anisotropy ratios
11   1211                       - random number seed
12   100                        - number of simulations
13   1    0.0                   - nst,nugget effect
14   1    1.0   15.0            - it,cc,range
```

A.6 Program *VarMerg*

The final program *VarMerg* is developed to achieve the final horizontal variogram realizations based on the methodology in Chapter 5. **Line 5** specifies the output file of the horizontal variogram realizations calculated by *VarUn* program (see Section A.3). **Line 6** indicates number of directional variograms which is 1 for the omnidirectional

Table A.6: Parameter file of *VarMerg* Program.

```
 1                    Parameters for VarMerg
 2                   **************************
 3
 4   START OF PARAMETERS:
 5   HorVar.dat            - horizontal well variogram
 6   2  10  13             - number of directional variograms, nlag
 7   SeisVar1.dat          - seismic-derived for direction 1
 8   SeisVar2.dat          - seismic-derived for direction 2
 9   VerVar1.dat           - vertical-derived for direction 1
10   VerVar2.dat           - vertical-derived for direction 2
11   Corr.dat              - correlation matrix between lags
12   Final.dat             - output final variogram realizations
13   Final.ps              - plot of final variogram realizations
14   0.2                   - cross covariance to enforce seismic limits
15   12115                 - random number seed
16   100                   - number of simulations
17   1                     - standardize variogram realizations
18   1   45                - Hmajor:Hminor, azimuth
```

or isotropic variograms, and 2 in case of anisotropy variogram (for major and minor directions of continuity). Number of lags (nlag) for each direction should be given. **Lines 7** to **8** are the seismic-derived variograms achieved by *VarSeis* (see Section A.4) program for the major and minor directions of continuity. **Lines 9** to **10** are the vertical-derived variograms attained by *VarVtH* (see Section A.5) for the major and minor directions of continuity. The first direction for both the seismic-derived and the vertical-derived variograms should be the major direction of continuity. The correlation matrix between lag distances calculated by *VarUn* program (see Section A.3) is entered on **Line 11**. **Line 12** is the output file for the final horizontal variogram realizations and **Line 13** is the plot of variogram realizations. **Line 14** is the minimum value of the cross covariance between well and seismic data for applying the seismic-derived limits: the limits are applied on the variogram realizations until the lag distance corresponding to this correlation. **Line 15** is random number seed for

variogram realizations. **Line 16** specifies number of variogram realizations. **Line 17** indicates variogram standardization (yes=0, no=0): the sill of all variogram realizations could be enforced to 1 (see Chapter 3). **Line 18** specifies the ratio of the maximum to minimum variogram ranges ($h_{major} : h_{minor}$), and azimuth of major direction for variogram realizations. This is the case when only the omnidirectional variogram could be calculated because of limited data, however, there is a clear anisotropy in the area. This anisotropy could be enforced on the omnidirectional variogram realizations (see Chapter 7, Section 7.3.3). $h_{major} : h_{minor}$ may be achieved from the seismic data. If $h_{major} : h_{minor}$ is one, no correction is applied. Although this program is aimed at improving the horizontal variogram, the directional variograms for 2D data set can be improved by the seismic-derived variogram.