

Automatic Variogram Modeling

by

Luis Antonio Davila Saavedra

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

Master of Science

in

Mining Engineering

Department of Civil and Environmental Engineering  
University of Alberta

© Luis Antonio Davila Saavedra, 2023

# Abstract

---

The variogram characterizes the spatial variability of a regionalized variable. Variogram calculation and modeling require a significant amount of professional time. The variogram has a significant impact on estimation and simulation, and it has a considerable impact in the mining industry since ordinary kriging is one of the most used tools for resource estimation. This thesis proposes novel tools and techniques related to automatic variogram calculation.

The first contribution of this thesis is implementing a GSLIB-like program: *autovar*. This program calculates the variograms for disseminated and tabular deposits. *Autovar* prompts the user for some basic parameters only. *Autovar* implements all necessary steps to get the experimental variogram points and its model directly from the data. The program infers the principal directions of continuity by eigendecomposition of the inertia tensor, followed by an order of the vectors. Then, the program calculates the experimental variogram points and models them. This thesis considers standard guidelines for variogram calculation and modeling. Different search parameters are implemented based on the deposit type. An important functionality of the program is the geolocated option, which allows the identification of variability centred at points (anchor locations) selected by the practitioner.

The second contribution is a novel technique to estimate a block model using a mix of ordinary kriging models. This mixed estimation technique considers several block kriging models and mixes them at each node of the block model. Each estimated model has a different variogram as input, reflecting different continuity characteristics. The variograms used in the kriging estimations are outputs from the *autovar* program: a general variogram and geolocated variograms centred at each selected anchor location. The weights are calculated based on the distance from each node to the anchor points. The postprocessing step standardizes the weights to add up to one, ensuring that kriging is exact. Generally, this technique recognizes local variability better than a regular technique.

This thesis presents different examples to show the potential and limitations of the proposed methodologies. This research presents a comparison of the proposed workflows with regular methodologies. Finally, a summary of tests is shown at the end to compare the results of the mixed estimation methodology with a regular estimation workflow.

# Dedication

---

To my family: Celinda, Ana, Antony and Antonio.

# Acknowledgments

---

I would like to express my gratitude to my supervisor, Dr. Clayton Deutsch, for his invaluable guidance, ideas and motivation throughout the entire research project. Clayton's exceptional expertise, trustworthiness, and kindness have made this learning process outstanding. Additionally, I would like to acknowledge the Centre for Computational Geostatistics (CCG) for their financial support.

I would like to express my appreciation to my parents, siblings, and girlfriend for their love, support, and encouragement. I also want to thank my Peruvian friends for being in contact with me and understanding my silence. I am grateful to all my friends at CCG for the stimulating discussions and the fun times that we shared. And finally, thanks to all of you who make this possible.

# Table of Contents

---

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Problem motivation . . . . .	3
1.2	Literature review . . . . .	4
1.3	Thesis statement and contribution . . . . .	5
1.4	Thesis outline . . . . .	6
<b>2</b>	<b>Theoretical Background</b>	<b>7</b>
2.1	Spatial variability . . . . .	7
2.2	Tensor of inertia to find principal directions . . . . .	11
2.3	Geolocated variogram and weighted order moments . . . . .	14
2.4	Variogram modeling and optimization function . . . . .	16
2.5	Estimation with kriging . . . . .	17
2.5.1	Simple kriging . . . . .	18
2.5.2	Ordinary kriging . . . . .	20
<b>3</b>	<b>Autovar program</b>	<b>22</b>
3.1	Initial set up . . . . .	24
3.1.1	Parameter file . . . . .	24
3.1.2	Data consistency and basic statistics . . . . .	25
3.2	Identification of directions of continuity . . . . .	26
3.3	Experimental variogram calculation . . . . .	32
3.4	Automatic modeling . . . . .	36
3.5	Geolocated function . . . . .	38
3.6	Model evaluation . . . . .	39
3.7	Outputs . . . . .	41
<b>4</b>	<b>Variograms</b>	<b>42</b>
4.1	Synthetic datasets . . . . .	42
4.1.1	Regular sampling . . . . .	43
4.1.2	Irregular sampling . . . . .	46
4.2	Real datasets . . . . .	48
4.2.1	Disseminated deposits . . . . .	48
4.2.2	Tabular deposits . . . . .	51
4.3	Comparison with varcalc and varmodel . . . . .	53
4.3.1	Comparison using a synthetic example . . . . .	54

4.3.2	Comparison using a 2-D real dataset . . . . .	56
4.3.3	Comparison using a 3-D real dataset . . . . .	58
4.4	Geolocated option . . . . .	59
<b>5</b>	<b>Estimation with geolocated variograms</b>	<b>62</b>
5.1	Proposed methodology . . . . .	62
5.2	Regular methodology . . . . .	64
5.3	Comparison of estimation with autovar and regular workflow . . . . .	65
5.4	Sensitivity with respect to number of anchors . . . . .	69
5.5	Statistics of study cases . . . . .	70
<b>6</b>	<b>Conclusions and Future work</b>	<b>74</b>
6.1	Contributions . . . . .	74
6.2	Limitations . . . . .	76
6.3	Future work . . . . .	76
	<b>References</b>	<b>78</b>

# List of Tables

---

2.1	Estimation methods in selected mines . . . . .	18
3.1	Valid combinations of vectors for the GSLIB angle convention for 3-D datasets. Part 1.	31
3.2	Valid combinations of vectors for the GSLIB angle convention for 3-D datasets. Part 2.	32
3.3	Tolerance parameters for pair selection for covariance and correlogram in <i>autovar</i> . . . . .	35
3.4	Tolerance parameters for flexible pair selection for correlogram in <i>autovar</i> . . . . .	40
5.1	Summary results for synthetic datasets with <i>autovar</i> . . . . .	71
5.2	Summary results for exhaustive images datasets with <i>autovar</i> . . . . .	72

# List of Figures

---

1.1	Input dataset and <i>autovar</i> outputs . . . . .	2
2.1	Schematic perspective of the variogram parameters to define search pairs . . . . .	10
2.2	Variogram tolerance parameters in 2-D views . . . . .	10
3.1	Pair selection for omnidirectional search . . . . .	27
3.2	Schematic representation of the rotation matrices with vectors . . . . .	29
3.3	Schematic representation of two rotation matrices that define the same pair search . . . . .	30
3.4	Experimental variogram points selection for modeling . . . . .	35
3.5	Variogram models fitting selected points . . . . .	37
3.6	Schematic representation of weighting for two pairs with the same lag distance . . . . .	39
4.1	Synthetic data generation by simulation with principal directions of ‘40,40,40’ . . . . .	43
4.2	Samples regularly spaced from initial dataset with principal directions defined by ‘40,40,40’ . . . . .	44
4.3	Experimental variogram and its model for dataset of Figure 4.2 . . . . .	44
4.4	Experimental variogram and its model for initial dataset with principal directions defined by ‘20,20,20’ . . . . .	45
4.5	Experimental variogram and its model for initial dataset with principal directions defined by ‘80,80,80’ . . . . .	46
4.6	Samples irregularly spaced from initial dataset with principal directions defined by ‘40,40,40’ . . . . .	46
4.7	Experimental variogram and its model for irregularly spaced samples from original dataset defined by ‘40,40,40’. . . . .	47
4.8	Experimental variogram and its model for irregularly spaced samples from original dataset defined by ‘80,80,80’. . . . .	48
4.9	Location plot of Large dataset . . . . .	49
4.10	Variograms of the Large dataset . . . . .	49
4.11	Location plot of Training dataset . . . . .	50
4.12	Variograms of the Training dataset . . . . .	50
4.13	Location plot of Oilsands dataset . . . . .	51
4.14	Variograms of the Oilsands datasets . . . . .	51
4.15	Location plot of Conklin dataset . . . . .	52
4.16	Variograms of the Conklin dataset . . . . .	53
4.17	Location plot of the synthetic dataset generated for <i>autovar</i> and <i>varcalc – varmodel</i> comparison . . . . .	55
4.18	Varcalc-varmodel results for dataset in Figure 4.17 . . . . .	55



4.19	Autovar results for dataset in Figure 4.17 . . . . .	56
4.20	Comparison of <i>autovar</i> and <i>varcalc – varmodel</i> results . . . . .	56
4.21	Location plot of 2-D real dataset and plot of principal directions . . . . .	57
4.22	Comparison of <i>autovar</i> and <i>varcalc – varmodel</i> for real 2-D dataset in Figure 4.21 . . . . .	57
4.23	Comparison of <i>autovar</i> and <i>varcalc – varmodel</i> for Oilsands dataset. . . . .	58
4.24	Synthetic data generation of two datasets with different directions of continuity . . . . .	59
4.25	Mixed synthetic datasets generation . . . . .	60
4.26	General variogram for mixed dataset . . . . .	60
4.27	Geolocated variogram for a mixed dataset for the South anchor . . . . .	60
4.28	Geolocated variogram for a mixed dataset for the North anchor . . . . .	61
5.1	Mixed synthetic dataset generation for the mixed estimation . . . . .	66
5.2	General variogram output for the mixed dataset . . . . .	66
5.3	South anchor’s geolocated variogram output for the mixed dataset . . . . .	66
5.4	North anchor’s geolocated variogram output for the mixed dataset . . . . .	67
5.5	Plot of weights for the block estimation models using the geolocated variograms . . . . .	67
5.6	Mixed estimation results using two anchors . . . . .	68
5.7	Variogram map and experimental variograms points and their models for the regular workflow . . . . .	68
5.8	Estimation results using regular workflow . . . . .	68
5.9	Mixed estimation results considering three anchors . . . . .	69
5.10	Mixed estimation results considering four anchors . . . . .	70
5.11	Mixed estimation results considering five anchors . . . . .	70
5.12	Summary results of mixed estimation for synthetic datasets using two anchors . . . . .	72
5.13	Summary results of mixed estimation for exhaustive images datasets using two anchors . . . . .	72

# List of Symbols

---

Symbol	Description
$\rho(\mathbf{h})$	Correlogram for lag $\mathbf{h}$
$C(\mathbf{h})$	Covariance for lag $\mathbf{h}$
$C(0)$	Covariance for distance 0
$d_i$	Distance from the point $i$ to the origin of coordinates
$E\{ \quad \}$	Expected value
$\omega(\mathbf{u}_i; \mathbf{o})$	Gaussian weighting for a value at location $\mathbf{u}_i$ with respect to the anchor $\mathbf{o}$
$\omega(\mathbf{u}_i, \mathbf{u}_i + \mathbf{h}; \mathbf{o})$	Gaussian weighting for a pair of points separated by a lag vector $\mathbf{h}$ with respect to the anchor $\mathbf{o}$
$\omega'(\mathbf{u}_i, \mathbf{u}_i + \mathbf{h}; \mathbf{o})$	Standardized Gaussian weighting for a pair of points separated by a lag vector $\mathbf{h}$ with respect to the anchor $\mathbf{o}$
$\hat{\rho}(\mathbf{h}; \mathbf{o})$	Geolocated correlogram for unit lag distance $\mathbf{h}$ with respect to the anchor $\mathbf{o}$
$\hat{C}(\mathbf{h}; \mathbf{0})$	Geolocated covariance for unit lag distance $\mathbf{h}$ with respect to the anchor $\mathbf{o}$
$\hat{m}_{+\mathbf{h}}(\mathbf{o})$	Geolocated mean value of the heads of the pairs for the lag distance $\mathbf{h}$
$\hat{m}_{-\mathbf{h}}(\mathbf{o})$	Geolocated mean value of the tails of the pairs for the lag distance $\mathbf{h}$
$\hat{\sigma}_{+\mathbf{h}}^2(\mathbf{0})$	Geolocated variance of the heads of the pairs for the lag distance $\mathbf{h}$
$\hat{\sigma}_{-\mathbf{h}}^2(\mathbf{0})$	Geolocated variance of the tails of the pairs for the lag distance $\mathbf{h}$
$\hat{\gamma}(\mathbf{h}; \mathbf{o})$	Geolocated variogram for unit lag distance $\mathbf{h}$ with respect to the anchor $\mathbf{o}$
$\hat{S}(\mathbf{o})$	Geolocated variance to standardize geolocated variogram
$\mathbf{h}$	Unit lag distance
$m_i$	Mass for MOI pair using pseudo-correlogram
$m_{i-pscov}$	Mass for MOI pair using pseudo-covariance
$m_{i-sq}$	Mass for MOI pair using pseudo-variogram
$m_{+\mathbf{h}}$	Mean value of the heads of the pairs for the lag distance $\mathbf{h}$
$m_{-\mathbf{h}}$	Mean value of the tails of the pairs for the lag distance $\mathbf{h}$
$m_{\mathbf{z}}$	Mean value of the variable $Z$
$M(g_1, g_2, \phi, t)$	Generalized mixture function: $\phi$ is the volume fraction, $g_1$ and $g_2$ are the values of a property being mixed, $t$ is a real number different than zero that characterizes the mix
$I_{xy}$	Moment of inertia for axes $y$ while the point rotates around axis $x$
$Z_{OK}^*(\mathbf{u})$	Ordinary kriging estimator for location $\mathbf{u}$
$R_{ij}$	Element of rotation matrix of row $i$ and column $j$

<b>Symbol</b>	<b>Description</b>
$\sigma^2$	Variance
$Z^*(\mathbf{u})$	Simple kriging estimator for location $\mathbf{u}$
$\lambda_j$	Kriging weight
$\sigma_E^2$	Estimation variance for simple kriging
$Y^*(\mathbf{u})$	Residual estimator for location $\mathbf{u}$
$\sigma$	Standard deviation
$\sigma_{+\mathbf{h}}$	Standard deviation of the heads of the pairs for the lag distance $\mathbf{h}$
$\sigma_{-\mathbf{h}}$	Standard deviation of the tails of the pairs for the lag distance $\mathbf{h}$
$\mathbf{u}$	Location in a domain
$\gamma(\mathbf{h})$	Variogram for lag $\mathbf{h}$
$z(\mathbf{u}_i)$	$z$ value for a location $\mathbf{u}_i$
$z(\mathbf{u}_i + \mathbf{h})$	$z$ value for a location $\mathbf{u}_i + \mathbf{h}$

# List of Abbreviations

---

<b>Abbreviation</b>	<b>Description</b>
2-D	Two-dimensional
3-D	Three-dimensional
azm	Azimuth to define the principal directions following the GSLIB conventions
cc	Variogram model contribution
CCG	Centre of Computational Geostatistics
CNN	Convolutional neural network
DHID	Drillhole identification
EDA	Exploratory data analysis
FIM	Fourier integral method
GSLIB	Geostatistical software library
ID	Inverse distance weighting to the power of one
ID2	Inverse distance weighting to the power of two
IDW	Inverse distance weighting
it	Variogram model type
LMC	Linear model of correlogram
LVA	Local varying anisotropy
MDS	Multidimensional scaling
MIK	Multiple indicator kriging
MOI	Moment of inertia
NE	Nugget Effect
OK	Ordinary kriging
RF	Random Function
ReV	Regionalized Variable
MSE	Mean square error
RMSE	Root mean square error
RV	Random Variable
SGS	Sequential Gaussian simulation
SK	Simple kriging

## Chapter 1

# Introduction

---

Mining is an important economic activity with the primary goal of mineral resource extraction. The final goal of resource estimation is to characterize in terms of tonnage and grade all economic mineral deposits. Kriging is one of the most used techniques to estimate mining deposits. Simulation is another technique that allows to predict the resources with a measure of uncertainty. Both techniques depend on the characterization of spatial variability.

There are different tools to characterize the spatial variability of a regionalized variable; the most common ones are the two-point statistics measures, which include the variogram, covariance, and correlogram. Another option is to use multipoint statistics, where the difficulty lies in getting a training image that represents the variable under study. This thesis focuses on the characterization of spatial variability through the variogram.

The percentage of samples extracted from a deposit for chemical analysis is tiny compared to the total volume to estimate; this scarcity of data makes two-point statistics a challenging but widely used tool. For an oil industry example, the volume of data investigated from the Brent field (North Sea) for cores is 0.000 000 001, cuttings (drilling debris) is 0.000 000 007, and for logging (geophysical measurement in the wellbore) is 0.000 001 (Chilès & Delfiner, 2012). Rossi and Deutsch (2013) mention that for mining, it is common to drill less than one billionth of the mass (Rossi & Deutsch, 2013). It could be challenging to find a variogram model that reasonably represents the spatial variability of the regionalized variable.

The variogram characterizes the spatial variability of a regionalized variable; it accounts for the available data and geology interpretation of the deposit (Gringarten & Deutsch, 2001). It is one of the most used tools in geostatistics, and the quality and consistency of it highly influence the results of any estimation or simulation process.

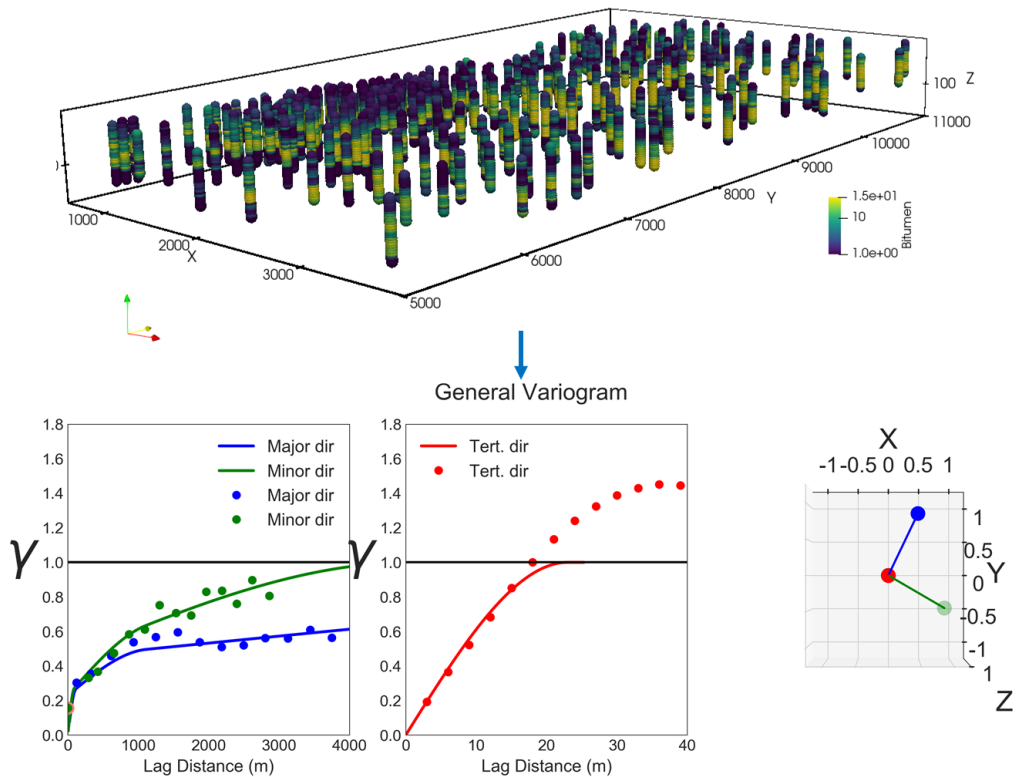
The variogram works under the assumption of the first and second order of stationarity, and under these assumptions, the variogram, covariance, and correlogram are related (Pyrcz & Deutsch, 2014). A definite positive model fits the experimental variogram. The model should fairly represent the regionalized variable's experimental points and variability. Usually, a trade off between precision and stability is needed to have a model that is as close to reality as possible and coherent with all the mathematical assumptions regarding stationarity.

The quality of a variogram depends on the data characteristics, the experience of the geostatistician and knowledge of the deposit's characteristics. Parameters and tolerances must be carefully chosen to produce a realistic experimental variogram based on the geological characteristics of the deposits.

Knowledge of the geology of the deposit gives insight into the interpretation of the calculated variogram (e.g. the trend in the deposit matches the trend in the variogram, or a zone of high mineralization explains the hole effect). The geostatistician evaluates the experimental variogram and its model to decide if it is representative of the spatial variability in the domain.

The automation of processes saves time, and it could give a reference starting point for manual tuning. The variogram is widely used in geostatistics, and its calculation significantly impacts the final estimates. After preparing the dataset, three steps can be considered: 1) finding the principal directions of continuity, 2) calculating experimental variogram points, and 3) fitting a model to the experimental variogram points.

Different steps or programs must be set to perform the three mentioned steps. *Autovar* executes all of them, inferring parameters directly from the dataset, and its final result is a variogram model. Nevertheless, a manually tuned variogram could be better, but it is not automatic and demands expertise and time from the practitioner. Figure 1.1 shows the input dataset and the program outputs.



**Figure 1.1:** Input dataset and *autovar* outputs. Vertical exaggeration on the Z axis for visualization purposes.

Using the geolocated option, the practitioner can declare  $n$  anchors and obtain  $n + 1$  variograms. The presented workflow for estimation utilizes each geolocated variogram for each kriging estimation. The workflow combines a weighted mix of each kriging estimation, with each node being influenced by different estimation models that resemble the recognized spatial characteristics of the variogram.

The suggested workflow standardizes the weight depending on the distance from the node to the anchor.

Geostatistical Software Library (GSLIB) is a set of programs developed for geostatistics in Fortran language (C. V. Deutsch & Journel, 1997). This library has been constantly improved and expanded by many researchers. *Autovar* modifies code from several previous programs of this library. The GSLIB *autovar* aims to produce a reasonable variogram for two kinds of deposits: disseminated and tabular, prompting the user to some basic parameters (e.g. the data file, the grid to consider and the deposit type). *Autovar* implements the common decisions in defining parameters according to the deposit type. The thesis explains the constraints and conditions and presents many examples to show the results.

Although the goal is to automatically produce an experimental variogram and its model for every deposit, there will be cases where this is not possible, and it could be related to the position of the samples or the poor continuity of the dataset. The automatic variogram output is a fast, reasonable reference; however, it must be validated.

### 1.1 Problem motivation

The variogram is a key tool for determining spatial variability for estimation and simulation techniques. The variogram's stability, precision, and representativeness will significantly impact the final results of resource modeling; thus, it is a step that demands significant time and effort to produce in a resource estimation workflow.

There are many deposit types with different geological characteristics. Each deposit has its unique spatial variability characteristics. Each variable potentially has different spatial characteristics, including differences in anisotropy, directions of continuity, and magnitude of continuity. The challenge is to characterize the deposit as good as possible with the available data.

Although each deposit is different, they have some characteristics in common. Similar geological processes will produce similar deposits, which translates to similar spatial variability characteristics. For each deposit type, there are general guidelines for selecting parameters and their tolerances to find the best variogram. A good parameter selection will produce a variogram that shows a structure and is consistent with the deposit; otherwise, the variogram could be noisy or non-representative.

In summary, the time and effort dedicated to generating the variogram is considerable, and an adequate parameter selection is challenging for a new practitioner. The *autovar* program looks to help with these challenges by automatically providing the experimental variogram and its model, where the practitioner sets only a few parameters. *Autovar* includes the most common configurations for the disseminated and tabular deposit types; it also evaluates the final model and relaxes constraints to look for a more stable variogram if required by the user. *Autovar* provides a good reference for the spatial characterization in a short time because it infers all parameters automatically.

The geolocated option is another functionality of *autovar* to identify local variability on datasets. A common practice when a deposit exhibits local variability is to divide it into stationary domains and model each domain separately; *autovar* analyzes the dataset without division.

The geolocated function allows the user to pick anchor points in areas with local variability. For this function, each pair analyzed for the variogram calculation will be weighted differently based on the distance from the pair to the anchor. The geolocated function could identify local variability better.

The practitioner can pick as many anchor points as necessary since the calculation is automatic and fast. A visual evaluation of the principal directions of continuity and the characteristics of the variogram (i.e. contributions, ranges) will help to identify an optimal number of anchors.

A domain is usually estimated with a unique variogram; this could generate an improper recognition of the characteristics of different areas of the domain. *Autovar* allows the calculation of geolocated variograms that can be used to generate different models that could better recognize local variability. These different models can be mixed to produce a unique kriging estimate.

*Autovar* could be a valuable tool to quickly characterize a deposit or help a new practitioner with a reference variogram; however, the practitioner's responsibility will always be the final evaluation and consideration of how good the variogram model is.

### 1.2 Literature review

There is extensive research with respect to automatic variogram modeling; however, there are few related to automatic experimental calculation. Some authors prefer to skip the presentation of the experimental variogram and get the model directly. This section will mention some of the research done mainly on automatic variogram fitting.

The most common methodologies for fitting are least squares (fit the experimental variogram), maximum likelihood (optimization on the parameters directly) and machine learning methods, such as using a convolutional neural network (CNN).

Pardo-Igúzquiza (1999) presents an automatic fitting using squared differences; the program needs the experimental variogram as input, and the user picks a weighting function. The program uses the simplex method for minimization of the objective function. The program considers one variable, and it is for 2-D datasets.

Larrondo, Neufeld, and Deutsch (2003) shows the methodology for a semiautomatic variogram modeling where the user provides the experimental variogram and could set some parameters to restrict the modeling or let the program calculate everything. The *varfit* program calculates an initial model. Then, it performs random changes in one parameter at a time to find the model that provides the lowest objective function value (best fit). The objective function could consider



more than one variable and allows a Linear Model of coregionalization (LMC). It also supports 3-D datasets.

Emery (2010) presents three different algorithms to fit an LMC, all of which aim to reduce the squared differences between the experimental variogram and the fitted model; one includes the use of simulated annealing.

Wilde and Deutsch (2012) show a method to directly fit the variogram volume (3-D variogram map) using an optimization function where simulated annealing perturbs the initial model. The program integrates the *varfit* algorithm.

Li, Zhang, Clarke, Liu, and Zhu (2018) mention a fitting technique that considers the genetic algorithm. This method considers an objective function that includes the approach of fitting the experimental points and the approach to validate the results of kriging estimation using cross-validation.

de Carvalho and da Costa (2021) mention a fitting for 2D datasets based on the Fourier integral method (FIM). The objective function compares the squared differences between the original map and a map produced using a theoretical variogram model using FIM. They use a genetic algorithm to find the parameters that produce a good fit.

The maximum likelihood method gives the variogram model directly. It gives uncertainty in variogram model parameters such as NE, sill, and range. It assumes that the variables are Gaussian. Different authors presented research about it, including Mardia and Marshall (1984), Pardo-Igúzquiza, Mardia, and Chica-Olmo (2009).

Jo and Pyrcz (2022) present an automatic semivariogram model using convolutional Neural Networks (CNN). This methodology involves two CNNs: (1) generate exhaustive data with sequential Gaussian simulation (SGS) conditioned to the original data, and (2) get the variogram parameters (ranges, directions, anisotropy) from the exhaustive data. The results will provide more than one possible variogram model the practitioner must evaluate.

Mokdad, Koushavand, and Boisvert (2022) present automatic variogram inference using CNN. This work presents the use of two CNNs, where the first one directly calculates the variogram model parameters, and the second one predicts the experimental variogram for the major directions of continuity. The user only needs to provide the data after the CNN is trained.

### 1.3 Thesis statement and contribution

Variogram calculation and modeling can be automated for disseminated and tabular deposits since the parameter selection is generalized based on the common characteristics of each deposit type, inferring parameters directly from the dataset. The mixed estimation technique will better characterize a domain that exhibits local scale variability because it combines models with different spatial continuity.

The first contribution of this work is the automation of the spatial variability characterization using the variogram and a proposed methodology for mixed estimation. The *autovar* program infers the principal directions of continuity, calculates an experimental variogram, and then fits it with an optimization by minimization algorithm. The geolocated option in the program allows the identification of local variability and the production of several estimation models for the same domain. The second contribution is the mixed estimation, a novel methodology for getting a final estimation considering local variation. The final mixed results combine all the models from the general and geolocated variograms.

## 1.4 Thesis outline

This thesis explains the automatic calculation and modeling of the variogram through the GSLIB-like *autovar* program. It explains the logic behind the program and gives the workflow to achieve the automatic variogram model. It also compares this program with previous programs used in CCG for several years. Finally, it gives an alternative methodology for estimation. The outline is detailed as follows.

The second chapter covers all the theoretical background related to the program implementation. It includes geostatistical concepts related to spatial variability, variogram calculation and modeling, tensor of inertia, the geolocated concept and weighted order moments, and the optimization function. This chapter also includes theory related to simple and ordinary kriging.

The third chapter describes the program thoroughly. It presents a detailed workflow, the parameter file and some practical considerations implemented on the algorithms. It also describes constraints, hard coding, and the logic behind these decisions. It presents both program functionalities, the general variogram and the geolocated version.

The fourth chapter shows case studies for the use of *autovar* from both synthetic and real datasets, including disseminated and tabular deposits. There are examples of regular and irregular sampling patterns. It also compares *autovar* with *varcalc* and *varmodel* to highlight the differences and similarities with the different methodologies.

The fifth chapter shows the workflow and application of *autovar* for estimation and compares a regular with the *autovar* workflow. The last chapter presents the conclusions of the thesis and gives some paths for future work.

## Chapter 2

# Theoretical Background

---

This chapter presents the theoretical background for the concepts presented in this thesis. The first section presents definitions and formulas related to spatial variability. The second section presents the theory related to the tensor of inertia for calculating the principal directions of continuity. The third section presents the geolocated concept to understand local spatial variability better. The fourth section presents variogram modeling. Finally, the last section presents content related to the estimation with simple and ordinary kriging.

### 2.1 Spatial variability

The variogram characterizes the spatial variability of a Regionalized Variable (ReV). Some concepts and definitions are described in this chapter that help to understand it. A Random Variable (RV) can take any outcome according to a probability distribution (C. V. Deutsch & Journel, 1997), and the outcome is randomly generated according to some random probabilistic mechanism (Isaaks & Srivastava, 1989). The first order moment of a RV or the mathematical expectation is written as  $E\{Z(\mathbf{u})\} = m(Z)$  (A. G. Journel & Huijbregts, 1978). The second-order moments are the variance, covariance, and the variogram. A Random Function (RF) is a set of Random Variables.

A ReV exhibits a particular spatial structure and has two characteristics: a random aspect related to the unpredictability at a short scale and a structured aspect related to larger scale trends (Matheron, 2019). A RF is used to characterize a ReV.

The variogram is the expected value of the squared differences between pairs of points for given lags and directions. The strict definition will imply many realizations for each location; however, there is only one value per location for mining since only one physical sample exists. Therefore, the intrinsic hypothesis is assumed, stating that the variogram value only depends on the vector's separation  $\mathbf{h}$  and not the location itself (A. G. Journel & Huijbregts, 1978). The principal features of the variogram are the sill, the range, and the nugget effect; many geostatistical textbooks explain their significance in detail.

Equation 2.1 shows the variogram for a domain A. Equation 2.2 shows the stationary covariance. The semivariogram (divide equation 2.1 by 2) is commonly used and referred to as a variogram. Equation 2.3 shows the stationary covariance related to the data variance when  $\mathbf{h} = 0$ . Equation 2.4 shows the stationary correlogram. Equation 2.5 shows the relations between covariance, correlogram, and variogram when the domain is deemed stationary. Nevertheless, these relationships (2.5) are

approximate unless the domain is really stationary.

$$2\gamma(\mathbf{h}) = E\left\{[Z(\mathbf{u} + \mathbf{h}) - Z(\mathbf{u})]^2\right\} \quad \forall \mathbf{u}, \mathbf{u} + \mathbf{h} \in A \quad (2.1)$$

$$C(\mathbf{h}) = E\{Z(\mathbf{u} + \mathbf{h})Z(\mathbf{u})\} - E\{Z(\mathbf{u})Z(\mathbf{u} + \mathbf{h})\} \quad \forall \mathbf{u}, \mathbf{u} + \mathbf{h} \in A \quad (2.2)$$

$$C(0) = \text{Var}\{Z(\mathbf{u})\} = \sigma^2 \quad (2.3)$$

$$\rho(\mathbf{h}) = \frac{C(\mathbf{h})}{C(0)} \quad (2.4)$$

$$C(\mathbf{h}) = C(0) \cdot \rho(\mathbf{h}) = C(0) - \gamma(\mathbf{h}) \quad (2.5)$$

The variance of the data values is considered the sill of the variogram. Even though geostatistics needs to make stationarity assumptions, usually regionalized variables are not. Stationarity is a decision, and it is related to a RF model property, not a characteristic of the physical phenomenon (Pyrcz & Deutsch, 2014).

The previous equations show the variogram and covariance in probabilistic notation; the following equations are the experimental versions of the variogram (Equation 2.6) and covariance (Equation 2.7):

$$\widehat{\gamma}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [z(\mathbf{u}_i) - z(\mathbf{u}_i + \mathbf{h})]^2 \quad (2.6)$$

$$\widehat{C}(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [z(\mathbf{u}_i) \cdot z(\mathbf{u}_i + \mathbf{h})] - m_z^2 \quad (2.7)$$

Where  $z(\mathbf{u}_i)$  and  $z(\mathbf{u}_i + \mathbf{h})$  are the pair of sampled values separated by a lag distance  $\mathbf{h}$ , and  $N(\mathbf{h})$  is the number of pairs for lag  $\mathbf{h}$ .

Knowledge of the geology of the deposit gives valuable insight into calculating and evaluating the variogram, and it allows to enrich the interpretation of the computed variogram (e.g., the trend in the variogram is coherent with the trend on the deposit, a zone of high mineralization explains the hole effect). The geostatistician needs to evaluate the experimental variogram and its model to decide if it is representative of the spatial variability in the domain. Journal mentions that samples from different deposit types could produce a similar variogram and encourages multipoint statistics from an image related to the site's geology (A. Journal, 2018). The challenge is getting the right image, which also points out the importance of the geology for the variogram calculation and interpretation.

Although the previous equations assume stationarity, non-ergodic versions could be considered. The non-ergodic covariance does not assume that the mean of the head and tails are the same, and it is considered a more general spatial covariance; the non-ergodic correlogram is considered a robust alternative to the variogram (Rossi & Deutsch, 2013). The following equations show the non-ergodic

covariance (Equation 2.8) and non-ergodic correlogram (Equation 2.9).

$$C(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [z(\mathbf{u}_i) \cdot z(\mathbf{u}_i + \mathbf{h})] - m_{-\mathbf{h}} \cdot m_{+\mathbf{h}} \quad (2.8)$$

$$\rho(\mathbf{h}) = C(\mathbf{h}) / (\sigma_{-\mathbf{h}} \cdot \sigma_{+\mathbf{h}}) \quad (2.9)$$

Where  $m_{-\mathbf{h}}$  is the mean for the tails and  $m_{+\mathbf{h}}$  is the mean for the heads, while  $\sigma_{-\mathbf{h}}$  is the standard deviation for the tails and  $\sigma_{+\mathbf{h}}$  is the standard deviation for the heads, the following equations show the calculations for these terms:

$$m_{-\mathbf{h}} = \frac{1}{N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} z(\mathbf{u}_i) \quad (2.10)$$

$$m_{+\mathbf{h}} = \frac{1}{N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} z(\mathbf{u}_i + \mathbf{h}) \quad (2.11)$$

$$\sigma_{-\mathbf{h}} = \sqrt{\frac{1}{N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [z(\mathbf{u}_i) - m_{-\mathbf{h}}]^2} \quad (2.12)$$

$$\sigma_{+\mathbf{h}} = \sqrt{\frac{1}{N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [z(\mathbf{u}_i + \mathbf{h}) - m_{+\mathbf{h}}]^2} \quad (2.13)$$

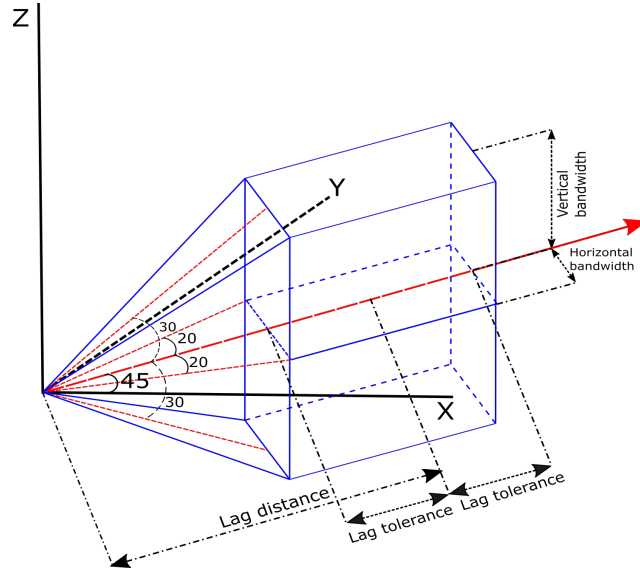
This thesis uses the non-ergodic versions of covariance and correlogram to obtain the variogram indirectly. The following section explains practical considerations for the parameter selection and its tolerances to define the search per lag distances.

The selection of the variable to calculate the variogram is an essential step for the future techniques that would be applied. Generally, the variable does not need to be transformed for kriging, but if Gaussian or indicator techniques are used, the data must be transformed accordingly (Pyrzcz & Deutsch, 2014).

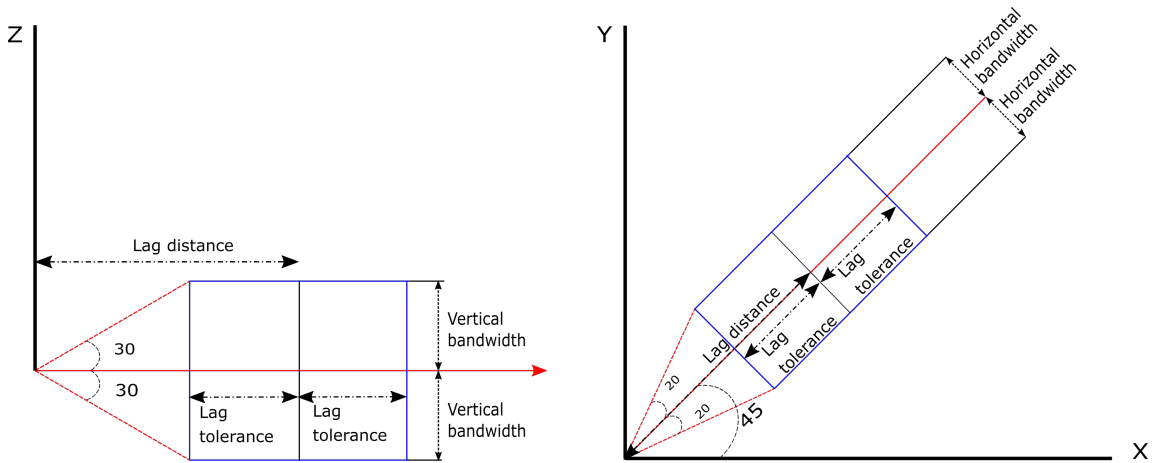
In mining, the samples are not a regular grid for various reasons, such as accessibility, the hardness of the rock, and more interest in certain areas. A variogram calculation must include tolerance parameters to group data pairs for different lag distances to get enough pairs to be representative. At least seven parameters have to be defined by the user to calculate an experimental variogram in a three dimensional dataset: three angles (azimuth, dip, and tilt) that characterize the principal directions of continuity, three lag distances for each direction of continuity, and the number of lags. Additionally, the pair selection considers direction tolerance parameters: azimuth and dip tolerance and vertical and horizontal bandwidths. A good selection of these parameters defines the quality of the experimental variogram.

The unit lag distance ( $\mathbf{h}$ ) is different per each direction of continuity (i.e. often larger for the major and shorter for the minor direction). The lag distance tolerance is usually half the unit lag distance; however, it could be larger to include more pairs per lag and get a more stable variogram.

Angle tolerances are also important for short distances. The azimuth tolerance is usually larger than the dip tolerance when working on the horizontal directions (major and minor); this is especially important for tabular deposits (J. Deutsch, 2015a). In contrast, the tolerances are similar and smaller for the tertiary direction, especially in tabular deposits where the goal is to keep the pairing of samples in the same drillhole (J. Deutsch, 2015b).



**Figure 2.1:** Schematic perspective of the variogram parameters, the red arrow represents the principal direction. The figure shows the following tolerance parameters for experimental variogram calculation: bandwidth, angles, and lag tolerances.



**Figure 2.2:** Sections from Figure 1. Left: section at the principal direction and Z axis to show the vertical bandwidth and dip tolerance. Right: Plan view (XY) to show the horizontal bandwidth and azimuth tolerance.

The horizontal and vertical bandwidth parameters have different importance for each deposit type. For tabular deposits, the definition of vertical tolerance is essential to get a representative model. For disseminated deposits, there could be more flexibility in selecting this parameter; in both cases, adequate tolerance parameters facilitate the detection of anisotropy.

Figure 2.1 shows the parameters for the major direction considering a configuration for azimuth of  $45^\circ$ , dip of  $0^\circ$ , and tilt of  $0^\circ$ . Consider the origin of the coordinate system as the location of a sample. X, Y and Z denotes a reference coordinate system. The solid red line shows the lag direction, the azimuth tolerance is  $20^\circ$ , and the dip tolerance is  $30^\circ$ . The vertical tolerance is bigger than the horizontal tolerance for visual effects; all the points inside this blue polyhedron will be compared with the point on the origin of the coordinates using the squared difference. The search compares all points for different lag distances.

Figure 2.2 shows two cuts of Figure 2.1 to show better the parameters and their tolerances. Since the dip is zero, the XY plane contains the lag vector. The dip tolerance is  $30^\circ$ . The azimuth of the vector is  $45^\circ$ , and the azimuth tolerance is  $20^\circ$ .

## 2.2 Tensor of inertia to find principal directions

The correct identification of the principal directions will ensure a good characterization of the spatial variability with the variogram. Different methods can be applied to identify the principal directions, such as variogram map, variogram sphere, neutral model, and Moment of inertia (MOI) or tensor of inertia (Kim & Deutsch, 2022). Nevertheless, MOI is preferred for this thesis because it can be automated.

A rotating rigid body with angular velocity  $\omega_v$  can be characterized using the tensor of inertia to calculate its angular momentum (Greiner, 2010). The tensor of inertia is a three by three matrix where the diagonal elements are known as moments of inertia while off the diagonal elements are the deviation moments (Greiner, 2010). One important property is that this tensor is real and symmetric; thus, only six elements of the matrix need to be calculated. The tensor of inertia can be understood as the mass distribution for a body with a gravity center in the origin of a coordinate system (Greiner, 2010).

The principal axis of inertia is defined when the deviation moments are zero, and there are only values for the diagonal of the matrix (moments of inertia). In this scenario, the body rotates around one of the principal axes of inertia, and the angular momentum and angular velocity will have the same orientation; this property allows to calculate the principal axis by eigendecomposition (Greiner, 2010). The principal axis can be considered the principal direction for a dataset. Therefore, this property helps determine a dataset's principal directions of continuity.

Equations for a rigid body are shown in different textbooks using an integral over the mass; however, since the analysis here considers discrete points, the integral is replaced with a summation. A point represents each pair, and the relative coordinates and mass (pseudo-correlation) can be calculated.

Equation 2.17 are modified from (Blundell & Damian, 2015). The original equations do not consider the inverse distance weighting. The relative distances from the points to the axis could be

of big influence. For instance, a small mass could have a big impact if it is far away from the origin of the coordinates. An inverse distance weighting (power of 2.5) is applied to reduce this effect:  $d_i^{2.5}$ . The 2.5 exponent reduces the effect of distance for every point.

Different ways to calculate the mass were tested: (1) a squared difference of values as in Equation 2.14 (pseudo-variogram), (2) a modified version of covariance (pseudo-covariance) as in Equation 2.15, and (3) a modified version of the correlation (pseudo-correlation) as in Equation 2.16. Equation 2.14 has the advantage that all mass values will be positive, while Equations 2.15 and 2.16 could produce negative values. All negative mass values will be set to zero, and the information for that pair will not inform the tensor. After testing, the results were inconsistent or inadequate in the first two cases, while the pseudo-correlation gives consistent and coherent results.

$$m_{i-sq} = [z(\mathbf{u}_i) - z(\mathbf{u}_i + \mathbf{h})]^2 \quad (2.14)$$

$$m_{i-pscov} = z(\mathbf{u}_i) * z(\mathbf{u}_i + \mathbf{h}) - m_z \quad (2.15)$$

$$m_i = \frac{[z(\mathbf{u}_i) * z(\mathbf{u}_i + \mathbf{h})] - m_z}{\sigma^2} \quad (2.16)$$

Where  $m_z$  and  $\sigma^2$  are the mean and variance of the selected point for the tensor of inertia analysis.

$$\begin{aligned} I_{xx} &= \sum_{i=1}^N m_i \frac{(y_i^2 + z_i^2)}{d_i^{2.5}} & I_{xy} &= I_{yx} = - \sum_{i=1}^N m_i \frac{x_i y_i}{d_i^{2.5}} \\ I_{yy} &= \sum_{i=1}^N m_i \frac{(x_i^2 + z_i^2)}{d_i^{2.5}} & I_{xz} &= I_{zx} = - \sum_{i=1}^N m_i \frac{x_i z_i}{d_i^{2.5}} \\ I_{zz} &= \sum_{i=1}^N m_i \frac{(x_i^2 + y_i^2)}{d_i^{2.5}} & I_{yz} &= I_{zy} = - \sum_{i=1}^N m_i \frac{y_i z_i}{d_i^{2.5}} \end{aligned} \quad (2.17)$$

Where the  $I_{xy}$  is the moment of inertia around the  $y$  axis when  $x$  is the referential axis of rotation;  $m_i$  is the mass of the point (pseudo-correlation);  $x_i$ ,  $y_i$  and  $z_i$  are the distance in the  $X$ ,  $Y$  and  $Z$  direction respectively, from the point to the rotation axis; and,  $d_i$  is the distance from the point to the center of the rotation axis.

The final result of these calculations is a tensor; then, the eigenvectors and eigenvalues of the matrix can be calculated by eigendecomposition. Each vector is sorted using the eigenvalues, and finally, the transpose of them is the rotation matrix.



$$\begin{array}{ccc}
 \begin{bmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{bmatrix} & \begin{array}{c} \rightarrow \\ \text{Eigendecomposition, ordering} \\ \text{and transpose} \\ \rightarrow \end{array} & \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \\
 \text{Tensor of inertia} & & \text{Rotation matrix}
 \end{array}$$

Different angle conventions are potential sources of confusion. This thesis uses the GSLIB convention to avoid this. In summary, the GSLIB convention considers Y as the major direction, X the minor, and Z the tertiary; the geostatistical lesson about GSLIB angles (M. Deutsch, 2015) provides a complete explanation of the rotation order. The angles can be calculated analytically using the equations to generate the rotation matrix following the GSLIB convention or using optimization by minimization to get a similar rotation matrix (Hadavand & Deutsch, 2022).

Whether the dip angle is  $90^\circ$  or not determines how the equations are solved; Slabaugh (1999) presents a detailed explanation of calculating the angles for any convention. The equations used in the program follow the GSLIB convention adapted from *varcalc*, and are specifically set for this thesis. When the dip is  $90^\circ$ , the value of the element  $R_{13}$  is one; also, the azimuth and tilt will be related, and there will be infinity solutions (Gimbal lock). The tilt is set to zero (Slabaugh, 1999) for convenience; the following equations show the solution system. In Fortran, the calculations are in radians; here, they will be presented as if all the calculations are done in degrees:

$$azm = 90 - \arctan\left(\frac{-R_{21}}{R_{31}}\right) \quad (2.18)$$

$$dip = 90^\circ \quad (2.19)$$

$$tilt = 0^\circ \quad (2.20)$$

When the dip is different than  $90^\circ$ , the following equations are used:

$$dip = -\arcsin(-R_{13}) \quad (2.21)$$

$$azm = 90^\circ - \arctan\left(\frac{R_{12}}{R_{11}}\right) \quad (2.22)$$

$$tilt = -\arctan\left(\frac{-R_{23}}{R_{33}}\right) \quad (2.23)$$

The previous equations show the general case and could be used for 2-D or 3-D cases. However, for a 2-D dataset only the azimuth is inferred because the dip and tilt are zero. Then, only Equation 2.22 is used for 2-D datasets.

These equations are the analytical solution to define the principal directions of continuity using three angles (GSLIB convention). Additionally, each vector of the principal direction can be defined by calculating two angles.

### 2.3 Geolocated variogram and weighted order moments

A domain could have local variations; there are two approaches to face this: (1) separate the domain in subdomains that are deemed stationary or (2) consider different influences from data based on the distance to a reference point in the domain. The reference point will be referred to as an anchor from this point on the thesis. This thesis allows the practitioner to declare anchors to identify local variations in the data without making subdomains. Avoiding the domain division can be helpful if there is local variability or if there is not enough data in the domain, and it also could be considered to get the most information from the data available.

The geolocated variogram assigns a different weight to each pair based on the distance from the points being evaluated for that pair to the anchor. A unique weight for each pair is calculated using the geometric mean of the two Gaussian weights of the points. There are different ways to combine two different components that are derived from the general mixture rule (Korvin, 1982) as in Equation 2.24. Machuca (2010) details that a  $t$  value that approaches to 0, which generates the geometric average (consider  $\phi = 0.5$ ), disregards effects from large lag distances, and when the lag distance is zero, the two point weight is a one point weight.

$$M(g_1, g_2, \phi, t) = [\phi g_1^t + (1 - \phi)g_2^t]^{1/t}, \quad t \neq 0 \quad (2.24)$$

Where  $\phi$  is the volume fraction,  $g_1$  and  $g_2$  are the values of a property being mixed,  $t$  is a real number different than zero that characterizes the mix.

A Gaussian weighting function is applied to have a smooth transition of weights; inverse distance weighting does not produce smooth transitions. Equation 2.25 shows the calculation of weight for each point, and Equation 2.26 shows the mixture of two weights using the geometric average.

$$\omega(\mathbf{u}_i; \mathbf{o}) = \exp\left(\frac{-dist(\mathbf{u}_i, \mathbf{o})^2}{2 \cdot sp^2}\right) + \epsilon \quad (2.25)$$

Where  $dist(\mathbf{u}_i, \mathbf{o})$  is the distance from the point  $i$  to the anchor  $\mathbf{o}$ ,  $sp$  is the smoothing factor, and  $\epsilon$  is a small number that ensures that the weight is greater than zero. The smoothing parameter ( $sp$ ) is calculated for each dataset and comes from the maximum extension of the dataset in the  $X$  or  $Y$  direction (the code assumes that the biggest extension would be on the horizontal coordinates rather than the vertical extension).

$$\omega(\mathbf{u}_i, \mathbf{u}_i + \mathbf{h}; \mathbf{o}) = \sqrt{\omega(\mathbf{u}_i; \mathbf{o}) \cdot \omega(\mathbf{u}_i + \mathbf{h}; \mathbf{o})} \quad (2.26)$$

This concept assumes local stationarity at each anchor location. Machuca presents a detailed explanation of the local stationarity decision (Machuca, 2010); following this logic, first and second moment orders can be weighted to have the respective local statistics of the first and second order moment for each anchor point. Machuca (2010) provides the formulas for calculating the weighted first and second order moments. The following formulas (from 2.27 to 2.35) are adapted or taken from Machuca's Ph.D. thesis. Equation 2.27 shows the standardized weight. Equation 2.28 shows

the standardized variogram; it uses the standardized weights. Equation 2.29 shows the weighted variance used to standardize the geolocated variogram.

$$\omega'(\mathbf{u}_i, \mathbf{u}_i + \mathbf{h}; \mathbf{o}) = \frac{\omega(\mathbf{u}_i, \mathbf{u}_i + \mathbf{h}; \mathbf{o})}{\sum_{i=1}^{N(\mathbf{h})} \omega(\mathbf{u}_i, \mathbf{u}_i + \mathbf{h}; \mathbf{o})} \quad (2.27)$$

$$\hat{\gamma}(\mathbf{h}; \mathbf{o}) = \frac{1}{2} \sum_{i=1}^{N(\mathbf{h})} \omega'(\mathbf{u}_i, \mathbf{u}_i + \mathbf{h}; \mathbf{o}) \cdot [z(\mathbf{u}_i) - z(\mathbf{u}_i + \mathbf{h})]^2 \quad (2.28)$$

$$\hat{S}(\mathbf{o}) \approx \frac{1}{2 \sum_{i=1}^N \sum_{j=1}^N \omega(\mathbf{u}_i, \mathbf{u}_j; \mathbf{o})} \sum_{j=1}^N \sum_{i=1}^N \omega(\mathbf{u}_i, \mathbf{u}_j; \mathbf{o}) \cdot [z(\mathbf{u}_i) - z(\mathbf{u}_j)]^2 \quad (2.29)$$

Where  $\omega'(\mathbf{u}_i, \mathbf{u}_i + \mathbf{h}; \mathbf{o})$  and  $\omega(\mathbf{u}_i, \mathbf{u}_j; \mathbf{o})$  are the standardize and non standardize weights for the points  $\mathbf{u}_i$  and  $\mathbf{u}_j$  with respect to anchor  $\mathbf{o}$ , respectively;  $z(\mathbf{u}_i)$  and  $z(\mathbf{u}_j)$  are the  $i$  and  $j$  sample values.  $N$  is the total number of data being analyzed for anchor  $\mathbf{o}$ .

Equation 2.30 shows the weighted version of the non-ergodic covariance, Equations 2.31 and 2.32 shows the local weighted means used on Equation 2.30. Equation 2.33 shows the local weighted correlogram, and Equations 2.34 and 2.35 show the local weighted standard deviations used on Equation 2.33.

$$\begin{aligned} \hat{C}(\mathbf{h}; \mathbf{o}) &= \sum_{i=1}^{N(\mathbf{h})} \omega'(\mathbf{u}_i, \mathbf{u}_i + \mathbf{h}; \mathbf{o}) \cdot z(\mathbf{u}_i) \cdot z(\mathbf{u}_i + \mathbf{h}) - \hat{m}_{-\mathbf{h}}(\mathbf{o}) \cdot \hat{m}_{+\mathbf{h}}(\mathbf{o}) \\ &= \sum_{i=1}^{N(\mathbf{h})} \omega'(\mathbf{u}_i, \mathbf{u}_i + \mathbf{h}; \mathbf{o}) [z(\mathbf{u}_i) - \hat{m}_{-\mathbf{h}}(\mathbf{o})] [z(\mathbf{u}_i + \mathbf{h}) - \hat{m}_{+\mathbf{h}}(\mathbf{o})] \end{aligned} \quad (2.30)$$

$$\hat{m}_{-\mathbf{h}}(\mathbf{o}) = \sum_{i=1}^{N(\mathbf{h})} \omega'(\mathbf{u}_i, \mathbf{u}_i + \mathbf{h}; \mathbf{o}) \cdot z(\mathbf{u}_i) \quad (2.31)$$

$$\hat{m}_{+\mathbf{h}}(\mathbf{o}) = \sum_{i=1}^{N(\mathbf{h})} \omega'(\mathbf{u}_i, \mathbf{u}_i + \mathbf{h}; \mathbf{o}) \cdot z(\mathbf{u}_i + \mathbf{h}) \quad (2.32)$$

$$\hat{\rho}(\mathbf{h}; \mathbf{o}) = \frac{\hat{C}(\mathbf{h}; \mathbf{o})}{\sqrt{\hat{\sigma}_{-\mathbf{h}}^2(\mathbf{o}) \cdot \hat{\sigma}_{+\mathbf{h}}^2(\mathbf{o})}} \in [-1, +1] \quad (2.33)$$

$$\hat{\sigma}_{-\mathbf{h}}^2(\mathbf{o}) = \sum_{i=1}^{N(\mathbf{h})} \omega'(\mathbf{u}_i, \mathbf{u}_i + \mathbf{h}; \mathbf{o}) \cdot [z(\mathbf{u}_i) - \hat{m}_{-\mathbf{h}}(\mathbf{o})]^2 \quad (2.34)$$

$$\hat{\sigma}_{+\mathbf{h}}^2(\mathbf{o}) = \sum_{i=1}^{N(\mathbf{h})} \omega'(\mathbf{u}_i, \mathbf{u}_i + \mathbf{h}; \mathbf{o}) \cdot [z(\mathbf{u}_i + \mathbf{h}) - \hat{m}_{+\mathbf{h}}(\mathbf{o})]^2 \quad (2.35)$$

Where:  $\omega'(\mathbf{u}_i, \mathbf{u}_i + \mathbf{h}; \mathbf{o})$  is the standardized weight for the pair of points  $\mathbf{u}_i$  and  $\mathbf{u}_i + \mathbf{h}$  with respect to the anchor  $\mathbf{o}$ .  $\hat{\gamma}(\mathbf{h}; \mathbf{o})$  is the weighted variogram for lag  $\mathbf{h}$ ;  $\hat{S}(\mathbf{o})$  is the weighted variance;  $\hat{C}(\mathbf{h}; \mathbf{o})$  is the weighted covariance;  $\hat{\rho}(\mathbf{h}; \mathbf{o})$  is the weighted correlogram;  $\hat{m}_{-\mathbf{h}}(\mathbf{o})$  and  $\hat{m}_{+\mathbf{h}}(\mathbf{o})$  are the weighted means for tails and heads;  $\hat{\sigma}_{-\mathbf{h}}^2(\mathbf{o})$  and  $\hat{\sigma}_{+\mathbf{h}}^2(\mathbf{o})$  are the weighted tail and head variances,

all of the previous first and second order moments are with respect to the anchor  $\mathbf{o}$ .

Equations 2.36 and 2.37 show the relationships between the geolocated variogram, geolocated covariance and geolocated correlation.

$$\hat{\gamma}(\mathbf{h}; \mathbf{o}) = \frac{(\hat{S}(\mathbf{o}) - \hat{C}(\mathbf{h}; \mathbf{o}))}{\hat{S}(\mathbf{o})} \quad (2.36)$$

$$\hat{\gamma}(\mathbf{h}; \mathbf{o}) = (1 - \hat{\rho}(\mathbf{h})) \quad (2.37)$$

Where  $\hat{\gamma}(\mathbf{h})$  is the geolocated variogram (Equation 2.28),  $\hat{S}(\mathbf{o})$  is the geolocated variance (Equation 2.30), and  $\hat{\rho}(\mathbf{h})$  is the geolocated correlogram (Equation 2.33). This thesis calculates the experimental variogram points using Equations 2.36 and 2.37.

## 2.4 Variogram modeling and optimization function

Variogram modeling allows having the covariance value for all distances and directions, and not only at the experimental points; this is required to perform estimation or simulation since every possible distance can be obtained from the model. A variogram model includes geological interpretation such as anisotropy, nugget effect and trends (Rossi & Deutsch, 2013).

A positive definite model is required to solve the kriging equations with a unique solution and positive kriging variance (Rossi & Deutsch, 2013). The practitioner can use any model that satisfies the positive definite condition or use the most common positive definite models. The spherical, exponential, Gaussian and power-law models are the most used, especially the first two.

The variogram model can combine different licit models known as nested structures; Rossi and Deutsch (2013) mentions that three structures will usually satisfy most fitting needs. Different proportions of the model can be assigned to different structures, which usually have different ranges. For instance, a variogram model can combine three different structures of a spherical model with different ranges.

Pyrcz and Deutsch (2014) describe three steps to model a variogram: (1) find the minimum necessary nested structures, (2) for each nested structure, select the model type, variance contribution and anisotropy parameters (angles and ranges for each principal direction), and (3) refine the model to get a ‘good’ fit. Thus, seven parameters are required to fit each structure: three angles, three ranges, and the model. The angles and ranges are related to the three principal directions of continuity. The fit can be manual or automatic, using optimization to minimize a measure of mismatch.

Some practical considerations are evaluated when modeling the experimental points: (1) a representative and stable fit are preferred, (2) some points should not be considered for modeling if they are adding noise (especially near the origin), (3) how to model the variogram with respect to the sill. The sill is the variance of the values for this thesis.

There are three approaches to model the experimental points: (1) Above the sill, (2) to the sill, and (3) considering the sum of each structure as sill (Samson & Deutsch, 2021). Another important consideration is to calculate the nugget effect (NE). Rossi and Deutsch (2013) mentions that a common practice is to calculate the downhole variogram to get its value; Chilès and Delfiner (2012) mentions that the NE value could be calculated by extrapolating the first points of the variogram, but it could be adjusted based on the knowledge of the physical property being studied.

An optimization model calculates values of decision variables that maximize or minimize an objective function considering given constraints (Winston & Goldberg, 2004). The experimental points are used to fit a function that minimizes the mean square error (MSE), comparing them with the values of the model at those distances. The decision variables for variogram modeling are the contributions and ranges of each structure. The experimental points are in the principal directions, and there is no need to evaluate angles.

### 2.5 Estimation with kriging

Kriging is an estimation technique developed in 1960 and has been extensively used since, initially focused on mining; nowadays, there are applications in different areas, including other types of natural resources and earth sciences (Chilès & Desassis, 2018). A comprehensive explanation of the Kriging equations and assumptions can be found in many geostatistical references (Chilès & Delfiner, 2012; Isaaks & Srivastava, 1989; A. G. Journel & Huijbregts, 1978; Pyrcz & Deutsch, 2014; Rossi & Deutsch, 2013). The kriging weights can be calculated using the normal equations, where the covariances are calculated from the variogram model.

Kriging can be used on many deposit types of simple or moderate complexity (e.g. tabular or porphyry deposits with simple faulting or folding); a more detailed reference of kriging and other estimation methods can be reviewed in (Darling, 2011). Kriging techniques are widely used in the mining industry; the block models for mineral deposits are generally estimated using ordinary kriging (OK). A few reports were reviewed to get an idea of the use of kriging in the industry; eight mines mention the use of ordinary kriging out of ten mines. Table 2.1 shows a summary of these technical reports.

The references for table 2.1 are: Krolak et al. (2017)<sup>33</sup>, Lehouiller et al. (2020)<sup>37</sup>, Gray et al. (2022)<sup>23</sup>, Ellis et al. (2023)<sup>19</sup>, Graden (2022)<sup>22</sup>, Lozada and Espinoza (2011)<sup>39</sup>, Schalekamp and McAllister (2022)<sup>53</sup>, Seward et al. (2022)<sup>54</sup>, Glencore (2022)<sup>21</sup>.

Bazania and Boisvert (2023) present an extensive analysis of techniques for estimation in the mining industry. They revised 150 reports and showed that ordinary kriging is used 47% of the time. The classification of the resources according to the industry standards shows that mainly geometric methods are used. Silva and Boisvert (2013) mention that 6% of the reports using kriging variance from a total of 120 revised technical reports that had enough information related to resource

**Table 2.1:** Estimation methods in selected mines. IDW, ID2 and ID stand for inverse distance weighting, inverse distance weighting to the power of 2, and inverse distance weighting to the power of 1, respectively. OK stands for ordinary kriging. MIK stands for Multiple indicator kriging.

Mine	Metal/Mineral	Estimation method	Year
Red Dog Mine <sup>33</sup>	Zn, Pb, Fe, Ba, Ag	OK	2017
Canadian Malartic Mine <sup>37</sup>	Au	OK and ID2	2021
Ravensthorpe Nickel Operations <sup>23</sup>	Ni	OK and MIK	2022
The Neves-Corvo Mine <sup>19</sup>	Cu,Zn,Sn,Pb	OK and IDW	2023
Highland Valley Copper <sup>22</sup>	Cu, Mo	OK	2013
Antamina <sup>39</sup>	Cu, Zn, Ag, Mo, Fe	OK and SK	2011
Elkview Coal Operation <sup>53</sup>	Coal	ID and ID2	2022
Greenhills Coal Operation <sup>54</sup>	Coal	ID2	2022
Lomas Bayas (II) <sup>21</sup>	Cu	OK	2022
Antappaccay <sup>21</sup>	Au,Cu,Ag	OK	2022

classification. Owusu (2019) reviewed 45 NI 43-101 technical reports related to the gold industry; only two reports include the use of kriging variance for resource classification. Bazania and Boisvert (2023) show that Drillhole Spacing (DHS) and Search Neighbourhood (SN) are used more than 80% of the time for resource classification.

Before the application of kriging, some conditions have to be reasonably ensured: data quality, outlier management, definition of estimation domains, manage of contacts, and others (J. Deutsch & Deutsch, 2015). The kriging search plan is defined based on the purpose of the estimate, which could be interim estimates, visualization and trend model estimates, final estimates, and probabilistic predictions estimates (J. Deutsch & Deutsch, 2015).

### 2.5.1 Simple kriging

Kriging estimates a value at an unsampled location using a linear weighted combination of the available samples. A strong assumption in simple kriging is that the mean is known for a domain deemed stationary. The practitioner can calculate the cumulative distribution function, mean, variance ( $\sigma^2$ ), variogram ( $\gamma(\mathbf{h})$ ), and the covariance ( $C(\mathbf{h})$ ) from the available data (C. Deutsch, 2021). Equation 2.38 shows the form of a linear estimator; the left part of the equation is also known as the residual of the estimated value, and it is usually represented as  $Y^*(\mathbf{u})$ .

$$Z^*(\mathbf{u}) - m = \sum_{i=1}^n \lambda_i [Z(\mathbf{u}_i) - m] \quad (2.38)$$

Where  $Z^*(\mathbf{u})$  is the estimated value at an unsampled location,  $m$  is the mean, and  $\lambda_i$  is the weight applied to the data value  $Z(\mathbf{u}_i)$ . A more consistent notation for the weights is  $\lambda(\mathbf{u}_i)$ ; however,  $\lambda_i$  is displayed to keep the notation simple.

The error of the estimation is the difference between the estimated value and the true value at that location; the problem is that we do not have access to the true value, so it needs to be probabilistically approached using the Random Function (Isaaks & Srivastava, 1989). Since the domain is deemed stationary, the sampled locations are considered outcomes of a random variable;

also, the unsampled location is a random variable because it is a linear estimation of the sampled values.

The criteria for best estimate used in kriging is the one that minimizes the error variance (C. Deutsch, 2021). The idea is to minimize the error variance (MSE) because there is no access to the true values at the locations being estimated. Equation 2.43 will be minimized to get the kriging equations, also known as normal equations.

Equation 2.38 can be expressed in terms of residuals ( $Y^*(\mathbf{u})$ ), as in Equation 2.39; this is implemented for easy calculations and takes advantage of the stationarity assumptions. Equations 2.40, 2.41, and 2.42 are the first and second order moments for the residuals assuming stationarity on the domain.

$$Y^*(\mathbf{u}) = \sum_{i=1}^n \lambda_i Y(\mathbf{u}_i) \quad (2.39)$$

$$E \{Y(\mathbf{u})\} = 0 \quad (2.40)$$

$$C(\mathbf{u}_i, \mathbf{u}_j) = E \{Y(\mathbf{u}_i)Y(\mathbf{u}_j)\} \quad (2.41)$$

$$Var[Y] = E \{Y^2(\mathbf{u})\} - (E \{Y(\mathbf{u})\})^2 = E \{Y^2(\mathbf{u})\} = C[Y(\mathbf{u}), Y(\mathbf{u})] \quad (2.42)$$

Where  $Y^*(\mathbf{u})$  is the estimated residual at the unsampled location,  $\lambda_i$  is the weight for the sampled (residual) value  $Y(\mathbf{u}_i)$ . The expected value of the residuals is zero as in Equation 2.40. The covariance of sampled values is expressed as the expected value of the product of the sampled values as in Equation 2.41. The variance of the residuals is the covariance of sampled values at the same locations as in Equation 2.42.

The following equations show the expansion of this expression to reach the kriging equations. Equation 2.42 is used in Equation 2.45 to express the last term as a function of the covariance. Equation 2.41 is used in Equation 2.46 to express the first two terms in function of the covariance.

Equation 2.47 shows the estimation variance ( $\sigma_E^2$ ) in terms of covariance ( $C$ ). The first term accounts for the redundancy of the data, the second for the closeness of the data, and the third for the variance (Pyrzc & Deutsch, 2014). Equation 2.47 can be derived with respect to  $\lambda_i$  to get the kriging equations. The last term of the equation will be zero after the derivation since there is no  $\lambda_i$  in it, giving Equation 2.48.

Equation 2.49 shows the simple kriging equations, where  $\lambda_j$  are the weights that will be calculated using the covariances between the data ( $C(\mathbf{u}_i, \mathbf{u}_j)$ ) and the covariance between the data and unsampled location ( $C(\mathbf{u}_i, \mathbf{u})$ ). The covariance values could be calculated from the variogram.

$$\sigma_E^2 = E \{[Y^* - Y]^2\} \quad (2.43)$$

$$\sigma_E^2 = E \left\{ \left[ \left( \sum_{i=1}^n \lambda_i Y(\mathbf{u}_i) \right)^2 - 2 \left( \sum_{i=1}^n \lambda_i Y(\mathbf{u}_i) \right) (Y(\mathbf{u})) + (Y^2(\mathbf{u})) \right] \right\} \quad (2.44)$$

$$\sigma_E^2 = E \left\{ \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j Y(\mathbf{u}_i) Y(\mathbf{u}_j) \right\} - E \left\{ 2 \left( \sum_{i=1}^n \lambda_i Y(\mathbf{u}_i) \right) (Y(\mathbf{u})) \right\} + E \left\{ (Y^2(\mathbf{u})) \right\} \quad (2.45)$$

$$\sigma_E^2 = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j E \left\{ Y(\mathbf{u}_i) Y(\mathbf{u}_j) \right\} - 2 \sum_{i=1}^n \lambda_i E \left\{ Y(\mathbf{u}_i) (Y(\mathbf{u})) \right\} + C[Y(\mathbf{u}), Y(\mathbf{u})] \quad (2.46)$$

$$\sigma_E^2 = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C[Y(\mathbf{u}_i), Y(\mathbf{u}_j)] - 2 \sum_{i=1}^n \lambda_i C[Y(\mathbf{u}_i), Y(\mathbf{u})] + C[Y(\mathbf{u}), Y(\mathbf{u})] \quad (2.47)$$

$$\frac{\partial \sigma_E^2}{\partial \lambda_i} = 2 \sum_{j=1}^n \lambda_j C(\mathbf{u}_i, \mathbf{u}_j) - 2C(\mathbf{u}_i, \mathbf{u}) = 0 \quad (2.48)$$

$$\sum_{j=1}^n \lambda_j C(\mathbf{u}_i, \mathbf{u}_j) = C(\mathbf{u}_i, \mathbf{u}), \quad \text{for } i = 1, \dots, n \quad (2.49)$$

Kriging has remarkable properties, these properties are explained by many authors in different geostatistical books (C. Deutsch, 2021; A. G. Journel & Huijbregts, 1978; Pyrcz & Deutsch, 2014; Rossi & Deutsch, 2013). Some properties are: (1) kriging minimizes the error variance by taking the second derivative of the error variance respect to each weight (C. Deutsch, 2021), (2) it has a unique solution since the variance is positive definite and there is only one data at each sample location (C. Deutsch, 2021), (3) kriging is exact since the value estimated at the sampled locations is the actual sample value (Rossi & Deutsch, 2013), (4) the kriging variance depends on the covariance values, not on the data values (Rossi & Deutsch, 2013), (5) the weights assigned to each sampled value depends on the sampled geometry rather than the actual values (Rossi & Deutsch, 2013), (6) the distance used to estimate in kriging is a structural distance (considers the structural continuity of the variable), not a Euclidean one (Rossi & Deutsch, 2013), (7) kriging accounts for clustering because of the covariances between data values (Rossi & Deutsch, 2013), and (8) kriging produces smoother estimates which are suitable for some mining applications (Rossi & Deutsch, 2013).

### 2.5.2 Ordinary kriging

There are different kriging types, such as simple kriging, ordinary kriging, kriging with a trend or universal kriging, indicator kriging and multi-Gaussian kriging; however, in this thesis, only simple and ordinary kriging will be used, thus only they are covered on this chapter.

The main difference between simple and ordinary kriging is the assumptions related to the mean. In ordinary kriging, the mean is not assumed to be known; this makes ordinary kriging widely used in the mining industry.

Equation 2.38 can be rewritten as in Equation 2.50. Then, the sum of weights is set as 1 to ensure the unbiased condition, and this constraint will disregard the mean value. Equation 2.51 shows the kriging estimator. The Lagrange parameter is necessary to minimize and solve the estimation variance. The ordinary kriging system is expressed in Equations 2.52 and 2.53. The



ordinary kriging system is not shown as the equations in the simple kriging section (see Equations 2.43 to 2.49); however, the procedure is similar and only needs to consider the Lagrange parameter.

$$Z^*(\mathbf{u}) = \sum_{i=1}^n \lambda_i [Z(\mathbf{u}_i)] + \left(1 - \sum_{i=1}^n \lambda_i\right) m \quad (2.50)$$

$$Z_{OK}^*(\mathbf{u}) = \sum_{i=1}^n \lambda_i [Z(\mathbf{u}_i)] \quad (2.51)$$

$$\sum_{j=1}^n \lambda_j C(\mathbf{u}_i, \mathbf{u}_j) + \mu = C(\mathbf{u}_i, \mathbf{u}), \quad \text{for } i = 1, \dots, n \quad (2.52)$$

$$\sum_{i=1}^n \lambda_i = 1 \quad (2.53)$$

Where  $Z^*(\mathbf{u})$  and  $Z_{OK}^*$  are simple and ordinary kriging estimators, respectively,  $n$  is the number of data used to estimate that point,  $m$  is the mean of the dataset,  $\mu$  is the Lagrange parameter,  $\lambda_i$  are the weights,  $Z(\mathbf{u}_i)$  are the sampled values,  $C(\mathbf{u}_i, \mathbf{u}_j)$  is the covariances between the data,  $C(\mathbf{u}_i, \mathbf{u})$  is the covariance between the data and unsampled location.

This chapter gives the theoretical background related to this thesis. It includes important concepts and presents the equations that are implemented in the *autovar* program (Chapter 3) and the kriging theory necessary for Chapter 5. The next chapter describes in detail the workflow and assumptions for the program.

## Chapter 3

# Autovar program

---

Variogram calculation and modeling are key tools to estimate or simulate a variable in a domain. The GSLIB-like *autovar* program aims for automatic calculation and modeling of variograms. The main goal is to get a model that represents the spatial distribution of the variable. This thesis considers a set of parameters to get representative experimental variogram points and common modeling decisions to fit them properly.

This chapter explains the logic and the assumptions of the program. *Autovar* considers the data and standard parameters according to the geology of disseminated and tabular deposits. The following workflow gives the order in which the program infers parameters and a general overview of the implemented algorithms. Specific details and reasons for the selected parameters of the workflow are presented after this overview.

### 1. Initial set up:

- a) Read the parameter file, perform initial calculations, and check the data file (trimming limits, data inside grid definition, count valid data and allocate memory, basic statistics).
- b) Read anchor file (if the practitioner uses the geolocated option).
- c) Calculate weights per anchor (if the practitioner uses the geolocated option).

### 2. Identification of directions of continuity:

- a) Select points with omnidirectional search: get points inside a selected radius (`max_dist`) based on the domain size.
  - i. Calculate horizontal spacing.
  - ii. Select pairs with different algorithms based on the deposit type and if the dataset is 2-D or 3-D.
  - iii. Calculate the vertical spacing.
- b) Calculate the rotation matrix from the tensor of inertia calculated from the pseudo-correlation between pairs. Calculate GSLIB and referential angles of the vectors that characterize the principal directions.

### 3. Experimental variogram calculation:

- a) Calculate the lag size per direction based on data spacing. The program uses the rotated coordinates.

### 3. Autovar program

---

- b) Calculate the extension of the variogram per direction using the rotated coordinates (set to  $0.5 \times \text{max\_dist}$  for the major and minor in disseminated deposits;  $0.75 \times \text{max\_dist}$  and  $0.5 \times \text{max\_dist}$  for the major and minor of the tabular deposits). If the geolocated option is selected, the maximum extension permitted is  $0.3 \times \text{max\_dist}$ .
  - c) Check if extensions and lag sizes are reasonable based on data spacing and domain extension. The program enforces extensions from the original coordinates if the rotation matrix is inadequate.
  - d) Check order relations and enforce them if needed.
  - e) Call *varexp* subroutine to calculate the experimental variogram and weighted experimental variogram if the program uses the geolocated option.
4. Automatic modeling: Get the experimental variogram points from 3 and fit them using the spherical model and an optimization algorithm.
- a) Set the angles according to the GSLIB angle definition.
  - b) Set parameters general parameters and array for modeling.
  - c) Model the experimental variograms using an objective function. Only the spherical variogram model is available.
5. Evaluate the variogram if asked by user:
- a) Evaluate based on the anisotropy ratio between the major and minor.
  - b) Recalculate the experimental variogram with larger tolerance parameters if the current ratio is smaller than the anisotropy ratio.
6. Write outputs (GSLIB format):
- a) Variogram model.
  - b) Experimental variogram.
  - c) Rotation matrix (for plotting and quick check).
  - d) Variogram model in text format for GSLIB program.
  - e) Weights for each anchor.

When the user selects the geolocated option, the steps from two to six are repeated in a loop to calculate the variograms per anchor. Equation 3.1 shows the maximum distance (*max\_dist*) parameter calculation. Step 5 is only recommended for disseminated deposits. The user can optionally declare general anisotropy ratios between major-minor and major-tertiary; otherwise, the program uses the default ratio (1.1). The next sections give a detailed explanation of these steps.

### 3.1 Initial set up

The *autovar* program infers all the parameters from the data. The user needs to specify the deposit type and if the geolocated option will be used. The practitioner can optionally declare anisotropy relationships. For the geolocated option, it is necessary to add a file with the location of the anchor points.

#### 3.1.1 Parameter file

Step one in *autovar* relates to the parameter file and datafile management. The program creates a default parameter file if one does not exist. When a complete and adequate parameter file is provided, the program reads the information from it. The user declares the data file in Line 5. Line 6 informs the columns for coordinates, variable and drillhole identification (DHID). Line 7 specifies the variable's range values.

```
1 Parameters for AUTOVAR
2 *****
3
4 START OF PARAMETERS:
5 oilsands.dat           - file with data
6 2 3 4 5 1             - columns for X, Y, Z, var, DHID
7 -1.0e21 1.0e21        - trimming limits
8 0 1                   - Geolocated variogram, number of anchors
9 anchor.out            - file with anchors
10 2 3 4                - X Y Z column of anchor file
11 0                    - Attempt to stabilize variogram?
12 0 2 4                - Use anisotropy ratio ,maj/min,maj/ter
13 67 587.0 50.0       - nx,xmn,xsiz
14 120 5030.0 50.0     - ny,ymn,ysiz
15 47 142.5 3.0        - nz,zmn,zsiz
16 autovarmod_oil.out   - file for variogram model output
17 autovar_oil.out      - file for experimental pairs output
18 debug_oil.out        - file for directions
19 varfit_oil.out       - variogram model text GSLIB format
20 weights_oil.out      - weights for each anchor
21 1                    - Debug (1=yes)
22 2                    - Disseminated(1) or tabular(2) deposit type?
```

The user specifies 1 to use the geolocated function and specifies the number of anchors in Line 8. Line 9 gives the anchor file, and Line 10 shows the columns for the coordinates of each anchor; the

program automatically assigns numbers to each anchor in ascending order (e.g., the first line with coordinates will be the first anchor).

If the user declares Line 11 as 1, the program will attempt to get a more stable variogram using fewer restrictions for pair selection in each lag. Line 12 is optional, and it informs anisotropy ratios. Lines 13, 14 and 15 inform the grid limits following the GSLIB format. The user can declare the name of the outputs on Lines 16 to 20. Line 21 selects the debug option (the user specifies 1), which allows to display some extra calculations on the screen. Line 22 declares the deposit type, where 1 stands for disseminated and 2 for tabular.

The previous parameter file shows the correct configuration without using anchors (Line 8 option 0) and with no stabilization attempt selected (Line 11 option 0) for a tabular deposit type (Line 22 option 2). In this thesis, a variogram where all the pairs have constant weight will be referred to as a general variogram (see Equation 2.6). In contrast, variograms that consider different weights per pair will be referred to as geolocated variograms (see Equation 2.28).

The user does not have to declare names for the geolocated variograms when the geolocated option is used. The program will generate output files that follow the format of the general variogram files, including the anchor number at the end. For instance, “varfit\_oil.out” will be the file for the general variogram, while “varfit\_oil1.out” will be the file for the first anchor.

### 3.1.2 Data consistency and basic statistics

The program reads the data and analyzes it. It selects points inside the grid and in the trimming limits range. *Autovar* calculates a reference parameter (`max_dist`) for consistency checks in the program’s next steps. It takes the maximum and minimum coordinates in X, Y, and Z. Equation 3.1 illustrates the reference parameter (`max_dist`) calculation.

$$\text{max\_dist} = \sqrt{(x_{max} - x_{min})^2 + (y_{max} - y_{min})^2 + (z_{max} - z_{min})^2} \quad (3.1)$$

Where  $x_{max}$  and  $x_{min}$  are the maximum and minimum value for the X coordinates, respectively;  $y_{max}$  and  $y_{min}$  are the maximum and minimum value for the Y coordinates respectively,  $z_{max}$  and  $z_{min}$  are the maximum and minimum value for the Z coordinates respectively.

This thesis implements workflows for 2-D or 3-D datasets only, and there is no omnidirectional calculation of variograms. The program uses the DHID column for the dataset to calculate the sample spacing in the vertical direction. The vertical spacing determines the vertical bandwidth parameters for the horizontal directions (major and minor).

The program displays the selected options in the parameter file at the beginning. It also displays if the dataset is 2-D or 3-D and some basic statistics, such as the number of data, and the variable average and variance. *Autovar* also shows the minimum, average and maximum spacing in the horizontal direction for 2-D and 3-D datasets; the program is set to compare the spacing between the four closest samples. For 3-D datasets, the program displays the total number of drillholes and

the averages of Z spacing, drillhole length, and number of samples per drillhole. The initial spacing information is in the units of the dataset (e.g. meters or feet).

If the option debug on screen is selected, the program displays more information on the screen, including the inertia tensor, eigenvectors and eigenvalues, and referential information of the original coordinates: the 90% extension in each direction and a referential lag distance. At this point, the program rotates the original coordinates using the rotation matrix to align them with the principal directions. The program displays information on the spacing in the principal directions, such as the maximum search distances and all the spacing information, but for the rotated coordinates. *Autovar* will select a maximum search and lag size according to the deposit type selected by the user. The program will calculate the experimental variogram according to all the previous inferred parameters. All the variogram experimental points, the variogram model, and the initial and final value of the objective function will be displayed on the screen if the debug option is selected.

### 3.2 Identification of directions of continuity

The second part identifies the directions of continuity using the eigendecomposition of the inertia tensor. The first step is to calculate the horizontal spacing of the points using the *datspac\_horz* subroutine, which is an adaptation of the GSLIB *datspac* program. It calculates a dataset's minimum, average, and maximum average spacing, assuming that all drillholes are vertical; otherwise, it will find distances between the projections of the last point of each drillhole onto a 2-D horizontal plane. The algorithm compares the closest four points, and the program shows the spacing results on the screen.

*Autovar* could consider all pairs; however, the distance from a mass point to the centre of reference highly influences the moment of inertia, as explained in Section 2.2. *Autovar* implements some additional constraints according to the dataset characteristics. For 2-D, the average spacing in sampling usually tends to be similar, and pre-processing is generally unnecessary. For 3-D datasets, the influences of the samples (spacing and number of samples per drillhole) in the vertical direction will have a big impact (i.e. the sampling spacing in the downhole could be 3 m, and the average spacing in the horizontal could be over 100 m).

For a 3-D dataset, the program calculates the average spacing between adjacent downhole samples using *datspac\_vert* subroutine. For a disseminated deposit type, *autovar* scales the Z coordinates by a ratio (see Equation 3.3) that allows a proper selection of points for the tensor of inertia since all points will be at a similar distance from each other.

$$F_{sc} = \frac{horz_{spac}}{vert_{spc}} \quad (3.2)$$

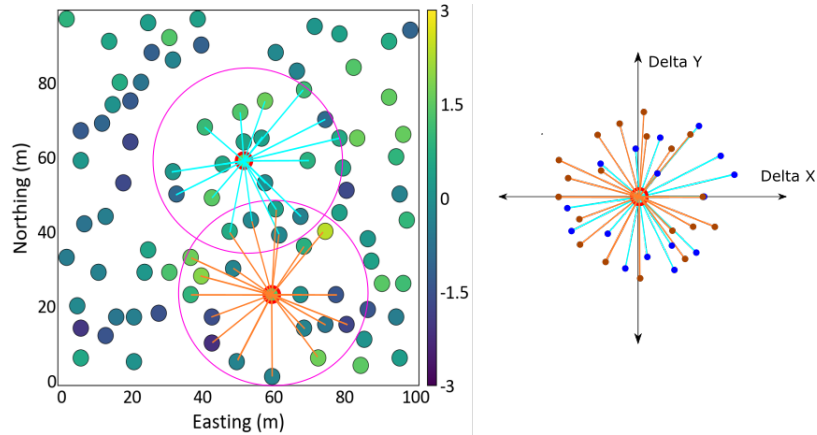
Where  $F_{sc}$  is the factor to scale the coordinates,  $horz_{spac}$  and  $vert_{spc}$  are the average horizontal and vertical spacing, respectively.

For 2-D datasets, the subroutine *omnipair* gets all possible pairs in every direction and calculates the relative coordinates and a pseudocorrelation value for each pair, which is the input for the *moi* subroutine. For 3-D disseminated deposit type, the points are scaled using Equation 3.3 and only points inside a cube are considered. The cube's edge is the minimum extension in the X or Y directions; this ensures that the same extension is considered in X and Y directions. For 3-D tabular deposits, all points are considered with no scaling; however, a ratio (*strat\_ratio*) compares the averages of horizontal spacing and the drillhole length to determine if a stratigraphic transformation is necessary.

$$strat_{ratio} = \frac{horz_{spac}}{vert_{ext}} \quad (3.3)$$

Where *strat\_ratio* is the stratigraphic ratio, *horz\_spac* is the average horizontal spacing, and *vert\_ext* is the vertical extension of the dataset.

The program compares each point against other points in a radius of  $0.2 \times \max\_dist$ . Different factors were tested (from 0.1 to 0.5) using 40 synthetic datasets with different 3-D orientations. Smaller factors ( $< 0.2$ ) are not representative, and bigger factors ( $> 0.2$ ) get inconsistent results. The factor 0.2 gives the correct extension for pair selection, which is key to identifying the directions of continuity. Figure 3.1 shows how the pairs are selected for a 2-D dataset. The two circles have the mentioned radius. All the points are compared to the points located in the centre of each circle. A plot of the relative coordinates of the pairs generated is shown on the right of Figure 3.1.



**Figure 3.1:** Pair selection for omnidirectional search. Left: Plan view of samples and two selected points that are compared to the closest neighbours. Right: Representation of the relative coordinates from the two selected points.

The program calculates a pseudocorrelation, and the relative coordinates are the differences of X, Y, and Z coordinates between pairs. Each pair has a delta of X ( $x_i$ ), Y ( $y_i$ ), and Z ( $z_i$ ) and a pseudocorrelation ( $m_i$ ) that acts as mass. The program sets the pseudocorrelation to zero if the value is less than zero because the mass can not be negative, and that pair is no longer analyzed. *Autovar* considers only values below the 99<sup>th</sup> percentile to minimize the influence of outliers; they will produce a huge mass that can lead to inaccurate results. The program calculates

the pseudocorrelation or ‘mass’ using the Equation 2.16.

For 3-D disseminated deposits, the *moi* subroutine uses the original coordinates after the point selection done by *omnipair*. If the geolocated option is selected, only samples up to  $0.3 \times \text{max\_dist}$  distance from the anchor are selected. The 0.3 factor was calculated by testing synthetic datasets and showed the best local characterization of principal directions. Smaller factors do not capture the local variability correctly, and bigger factors will show more regional principal directions. The selection of points near the anchors allows for better local identification of directions at each anchor.

For 3-D tabular deposits, the *strat\_ratio* is evaluated. When *strat\_ratio* is less than 10, the *omnipair* subroutine is executed. Otherwise, the dataset is transformed internally to stratigraphic coordinates using the proportional method, and then the *omnipair* subroutine is applied. The ratio of 10 was determined after testing real cases of stratigraphic deposits. The results of the variogram in the tertiary direction will be in stratigraphic units when the stratigraphic transformation is applied. Subsequently, the program performs only the inference of directions for major and minor, while the tertiary direction is vertical by default.

The user is encouraged to provide a dataset that has already been properly transformed in the desired stratigraphic units; otherwise, *autovar* will check the *strat\_ratio* and set a transformation. The proportional, truncation, and onlap methods (or a combination of them) are possible options to transform the data to stratigraphic units (Latifi & Boisvert, 2022). Only the proportional method is implemented in *autovar* to cover some common cases. The program calculates the average thickness of the drillholes to do the transformation as in Equation 3.4.

$$z_{strat} = \frac{z - z_{bot}}{z_{top} - z_{bot}} \times Th \quad (3.4)$$

Where  $z_{strat}$  is the stratigraphic coordinate for the original  $z$  coordinate,  $z_{bot}$  and  $z_{top}$  are the bottom and top coordinates of the drillholes (bottom and top surfaces), and  $Th$  is the average thickness of the drillholes.

The user is encouraged to ensure that the data provided is adequate to get a reasonable variogram (i.e. Exploratory data analysis). Exploratory data analysis (EDA) consists of a univariate or bivariate statistical analysis to understand the data (Tukey, 1977). Additionally, some of the prerequisites before variogram calculation could be coordinate transformation for tabular deposits, compositing the data, outlier management, and normal score transformation (J. Deutsch, 2015a). The variogram could be unstable and noisy if the variable has a highly skewed distributions and preferential sampling in highly valued areas (Kumara & Deutsch, 2017).

The *moi* subroutine takes the output from *omnipair* and calculates the tensor of inertia, eigenvalues and eigenvectors (see equations in Section 2.2). The eigenvectors are set in ascending order using the eigenvalues for 2-D and 3-D disseminated and tabular deposits. If the program uses the stratigraphic coordinates, the first two vectors are sorted in descending order, while the last vector remains in the same position in the matrix.

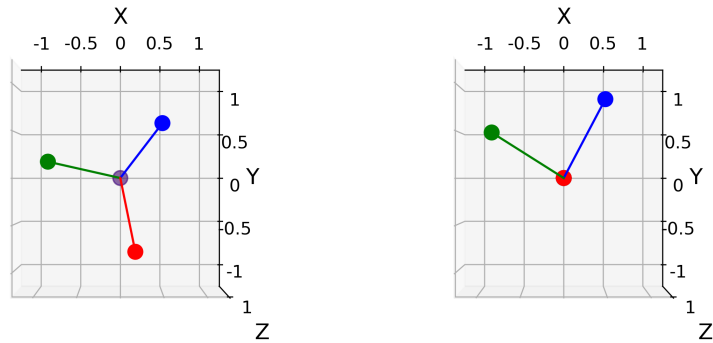


There are other methods to determine the principal directions, such as variogram maps, variogram sphere, and neutral model (Kim & Deutsch, 2022). The MOI is preferred since it can be automated. The program transposes the ordered eigenvectors to get the rotation matrix; this calculation skips the need to prompt the user with the angles of the three directions on continuity.

The rotation matrix is an output of the program to facilitate reviewing and validating the inferred directions. For a 3-D dataset, a three by three matrix is calculated; for a 2-D dataset, a two by two matrix is calculated, but a three by three matrix is used for consistency. For the 2-D datasets, only four terms out of the matrix are calculated, and  $R_{31}, R_{32}, R_{13}, R_{23}$  others are filled with zeros except for element  $R_{33}$  which is one. Figure 3.2 shows a 3-D representation of the rotation matrix as vectors, and the following matrices show the exact values for both plots.

$$\text{3-D rotation matrix: } \begin{bmatrix} 0.492 & 0.587 & 0.643 \\ -0.852 & 0.176 & 0.492 \\ 0.176 & -0.790 & 0.587 \end{bmatrix}$$

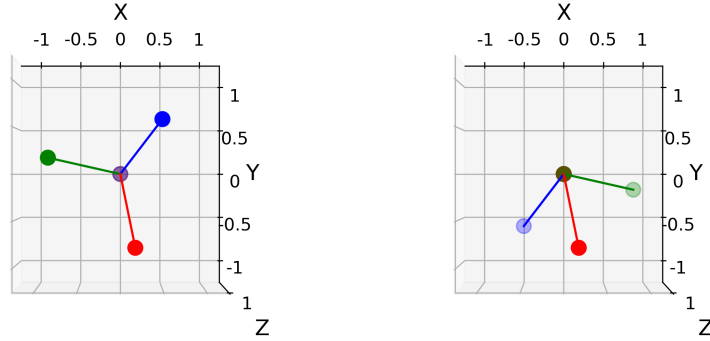
$$\text{2-D rotation matrix: } \begin{bmatrix} 0.500 & 0.866 & 0.000 \\ -0.866 & 0.500 & 0.000 \\ 0.000 & 0.000 & 1.000 \end{bmatrix}$$



**Figure 3.2:** Schematic representation of the rotation matrices with vectors. Left: Plot for 3-D dataset rotation matrix. Right: Plot for 2-D dataset rotation matrix.

There is more than one set of angles that will produce the same rotation matrix. There are two sets of angles that produce the same matrix considering the following ranges for the azimuth, dip and tilt:  $[-360,360]$ ,  $[-90,90]$  and  $[-90,90]$ . For instance, considering the GSLIB convention ('Azimuth, Dip, Tilt'), the set of angles of '40, 40, 40' produces the same rotation matrix as '-320, 40, 40'. The same pair search can be defined with two rotation matrices. Consequently, four sets of angles can define the same pair search. For instance, '40, 40, 40', '-320, 40, 40', '220, -40, -40', '-140, -40, -40'

produces the same pair search.



**Figure 3.3:** Schematic representation of two rotation matrices that define the same pair search. Left: rotation matrix defined by angles ‘40, 40, 40’. Right: rotation matrix defined by angles ‘-320, 40, 40’.

One important property of the variogram and covariance is that they are even functions (Montero, Fernández-Avilés, & Mateu, 2015). This means that  $\gamma(\mathbf{h}) = \gamma(-\mathbf{h})$  or  $C(\mathbf{h}) = C(-\mathbf{h})$ . Then, any of the four sets of angles will produce the same variogram or covariance.

All the analytic equations to find the GSLIB angles are in Section 2.2 (see Equations 2.20 and 2.23.) Depending on the orientation of the eigenvectors, the equations could generate a set of angles that is different than the GSLIB convention. There are eight combinations of eigenvectors that will produce the same pair search, and only two of them can be defined using the GSLIB angle convention. The program implements a subroutine to force the eigenvectors to be coherent with the GSLIB angles.

All the possible combinations of vectors with GSLIB angles were characterized to recognize the valid set of eigenvectors that comes from the GSLIB rotation matrix. Tables 3.1 and 3.2 show the valid position of the major, minor and tertiary in octants for 3-D datasets. The index of the octant is a unique number that identifies each possible combination (see Equation 3.5). The subroutine prioritizes that the major direction vector is in the octants where  $x \geq 0$  because this ensures a correct azimuth calculation with the analytical solution. All the combinations are compared to the valid set of eigenvectors to pick the one that follows the GSLIB angle convention.

$$oct_{idx} = oct_{maj} \times 100 + oct_{min} \times 10 + oct_{ter} \quad (3.5)$$

Where  $oct_{idx}$  is the unique octant index,  $oct_{maj}$ ,  $oct_{min}$ , and  $oct_{ter}$  are the octants where the major, minor and tertiary vectors are located, respectively.

This solves most inconsistencies in the angle calculation with analytical equations. A rotation matrix is calculated using the inferred angles. This matrix is compared against the right set of eigenvectors with an objective function shown in Equation 3.6. A last consistency check generates another set of possible angles, changing the dip and tilt for its negative values and getting a rotation

**Table 3.1:** Valid combinations of vectors for the GSLIB angle convention for 3-D datasets. Part 1.

Octant for major	Octant for minor	Octant for tertiary	Octant index
1	1	1	111
1	1	2	112
1	1	4	114
1	3	2	132
1	4	1	141
1	4	2	142
1	4	3	143
1	5	1	151
1	5	3	153
1	5	4	154
1	8	3	183
1	8	4	184
2	1	3	213
2	1	4	214
2	4	3	243
2	4	4	244
2	5	1	251
2	5	4	254
2	6	1	261
3	1	4	314
3	2	1	321
3	2	4	324
3	6	1	361
3	6	2	362
3	7	1	371
3	7	2	372
4	2	1	421
4	2	2	422
4	3	1	431
4	3	2	432
4	7	2	472
4	7	3	473
4	8	2	482
4	8	3	483
5	1	1	511
5	1	2	512
5	4	1	541
5	4	2	542
5	7	4	574
5	8	1	581
5	8	4	584
6	1	1	611
6	1	2	612
6	1	3	613
6	2	3	623
6	5	1	651
6	5	2	652
6	8	1	681
6	8	2	682

**Table 3.2:** Valid combinations of vectors for the GSLIB angle convention for 3-D datasets. Part 2.

Octant for major	Octant for minor	Octant for tertiary	Octant index
7	2	3	723
7	2	4	724
7	3	3	733
7	3	4	734
7	5	2	752
7	6	2	762
7	6	3	763
8	3	1	831
8	3	4	834
8	4	1	841
8	4	4	844
8	6	3	863
8	6	4	864
8	7	3	873
8	7	4	874

matrix that is compared to the correct set of eigenvectors. If the error of the second comparison is smaller than the error of the first comparison, these angles are set as output; otherwise, the initial angles are set.

$$OF_{eig} = \sum_{i=1}^3 \sum_{j=1}^3 \left| (E_{(i,j)})^2 - (Tr_{(i,j)})^2 \right| \quad (3.6)$$

Where  $OF_{eig}$  is a measure of comparison between rotation matrices,  $E_{(i,j)}$  is the correct rotation matrix and  $Tr_{(i,j)}$  is the matrix generated with the inferred GSLIB angles,  $i$  and  $j$  are the rows and columns of the matrix.

Equation 3.6 works because it allows the comparison between opposite vectors, which makes no difference for the search of pairs. Once the rotation matrix and angles are calculated, the program calculates the experimental variogram points.

### 3.3 Experimental variogram calculation

The third step is to get the experimental variograms for each principal direction. *Autovar* will use the original or stratigraphic coordinates, depending on which set was used to infer the directions. Before using the rotation matrix, the program calculates 90% of the extension in X, Y, and Z to have as a reference for the ranges of the variograms. The 90% reference extension has been determined after testing synthetic and real examples; it ensures the proper functionality of the program if the extension is not properly recognized using the rotated coordinates. The product of the selected coordinates and the rotation matrix gives the rotated coordinates. The variogram calculation uses the rotated coordinates. The maximum ranges are printed on the screen as a reference for the practitioner.

The subroutines *datspac\_horz* and *datspac\_vert* calculate the horizontal and vertical spacing of the rotated coordinates. These distances are the data spacing in each principal direction, and they are set as the lag spacing for each corresponding direction. A constraints is applied to ensure that the lag sizes is reasonable and respect order relations: the tertiary lag is set to 0.75 times the size of the average vertical spacing if the ratio between the tertiary lag and the average vertical spacing is less than 0.25. These ratio and factor (0.75 and 0.25) were selected after testing real datasets deposits because they gave consistent results.

For 2-D datasets, all parameters related to the tertiary direction are set as constants and not analyzed. Inconsistencies in lag distances of the variograms can appear when the rotation matrix is inadequate or when the spacing on the horizontal presents high variability. An inadequate matrix is produced when pairs with big mass and far away from the origin of rotation coordinates are present, even after all the measures taken to avoid this situation (see Section 2.2).

The next step is to calculate the extension of the variogram in each direction. It is recommended that the maximum range of the variogram in each direction does not exceed more than half of the domain size (A. G. Journel & Huijbregts, 1978). Different constraints exist based on the type of deposit. The maximum extent is half the maximum distance in that direction (i.e., a factor of 0.5) for the disseminated deposits. In contrast, there are different factors per direction for tabular deposits: 0.75 for the major, 0.5 for the minor and 0.35 for the tertiary direction. The ratios for the tabular deposits have been calculated experimentally after testing and ensuring the program's functionality. The factors for the tabular deposits are different per direction because the horizontal continuity on a tabular deposit is considerably greater than in the vertical direction.

When the geolocated option is selected, the program reduces the maximum allowed extension to emphasize data closer to each anchor. For 2-D data, the extension is set to 0.3 times the maximum extension ( $0.3 \times \text{max\_dist}$ ) of the dataset, while for 3-D data, the extension is set to 0.4 times the maximum extension ( $0.4 \times \text{max\_dist}$ ). The factors 0.3 and 0.4 were selected after testing synthetic and real datasets. These factors ensure that the analysis is local and representative for each anchor. After the calculation of the extension of the variogram, *autovar* checks the order relations to ensure that they are respected; otherwise, it enforces them, as shown in the next equations.

$$Ext_{maj} \geq Ext_{min} \geq Ext_{ter} \quad (3.7)$$

$$Ext_{min} = 0.85 \times Ext_{maj} \quad (3.8)$$

$$Ext_{ter} = 0.85 \times Ext_{min} \quad (3.9)$$

Where  $Ext_{maj}$ ,  $Ext_{min}$ ,  $Ext_{ter}$  are the calculated extensions for the experimental variogram in the major, minor and tertiary direction, respectively. The 0.85 factor has been selected because it works consistently and it ensures the program functionality.

Equations 3.7, 3.8, and 3.9 ensure order relations of the rotated coordinates. Nevertheless,

another set of consistency tests is implemented to compare the inferred extensions with respect to the dataset extension. For disseminated deposit types, the extension in the tertiary direction is checked to ensure that it has a reasonable value. The extension is set to 0.5 times the referential extension in  $Z$  if the rotated coordinates are inconsistent. This factor follows the general recommendation for the variogram extension (A. G. Journel & Huijbregts, 1978). A wrong extension in the vertical direction could happen if the rotation matrix is inadequate for the tertiary direction. For tabular deposits, the extension of the horizontal directions (major and minor) is checked to ensure they have at least half the domain's extension.

After the second test in the extension of the variograms, a double check on the order relations is performed. The results are printed on the screen when the debug option must be selected. At this point, the extension of the variogram and lag size are coherent for each direction. This allows to calculate a reasonable number of lags using Equation 3.10.

$$Nlag_{dir} = \frac{Ext_{dir}}{Lsz_{dir}} \quad (3.10)$$

Where  $Nlag_{dir}$  is the number of lags in the direction  $dir$ ,  $Ext_{dir}$  and  $Lsz_{dir}$  are the total extension and lag size for the direction being analyzed.

All necessary arrays for the experimental variogram calculation are set according to the required number of lags per direction. Then, the subroutine *varexp* is called. This subroutine was adapted from the GSLIB *varcalc* program (J. Deutsch, 2015c).

The subroutine *varexp* takes all the parameters previously inferred in the main code that is necessary for the experimental variogram points calculation. Angle tolerances are set to  $22.5^\circ$  for the azimuth and  $22.5^\circ$  for the dip. These angles are the most common to use in an initial variogram calculation in the industry, and they generally allow fair stability and precision for the experimental variogram points. The subroutine identifies if the stratigraphic coordinates are being used. The rotated coordinates are aligned with the principal directions: the major direction is  $X'$ , the second is  $Y'$ , and the third is  $Z'$ ; this allows to easily set the constraints for the pair search, such as bandwidth and angle tolerances.

The calculation of the general variogram is indirect; the program calculates the correlogram first, and then the variogram is obtained using Equation 2.37. The correlogram has shown to be a robust alternative to the direct variogram calculation, and it is preferred since the goal of the general variogram is to give a stable variogram from the dataset. In contrast, the geolocated variograms will provide more detail at a local scale.

Table 3.3 shows the tolerances for deposit type. Each lag size is used according to each direction for the horizontal bandwidth (e.g.  $lsz_{maj}$  is used for the restriction on the horizontal bandwidth of the major direction). At the same time, for the vertical, a factor of the vertical spacing is set. For the horizontal directions (major and minor), the bandwidth tolerances for the tabular deposits are half of the disseminated tolerances, and this better represented the continuity for each deposit type.

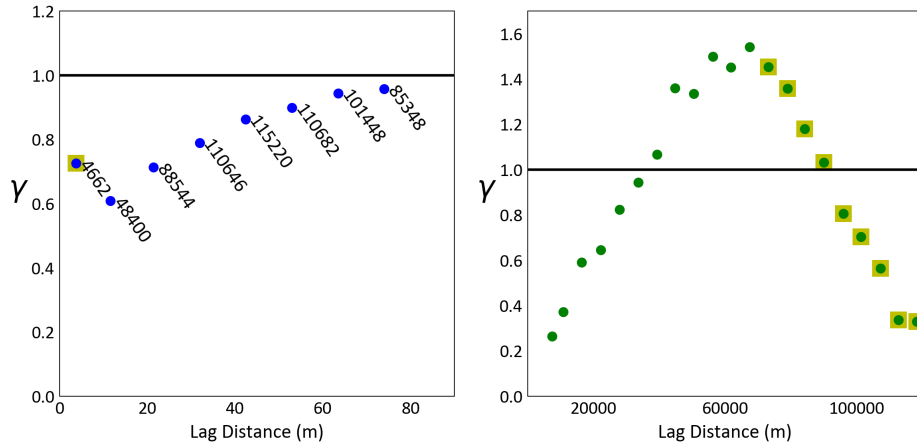
For the tertiary direction, 4 and 0.5 times the vertical spacing were the most stable parameters for disseminated and tabular deposits, respectively. The factors in Table 3.3 were selected after testing with the real and synthetic examples.

**Table 3.3:** Tolerance parameters for pair selection for covariance and correlogram. Where  $lsz_{maj}$ ,  $lsz_{min}$ , and  $lsz_{ter}$  are the lag sizes for the major, minor and tertiary directions, respectively, and  $vspc$  is the vertical spacing. The tolerances for the tabular deposit are more restrictive than the ones for the disseminated deposit type since the tabular deposit has strong controls on the horizontal directions.

Bandwidth tolerances:	Major direction		Minor direction		Tertiary direction	
	Horizontal	Vertical	Horizontal	Vertical	Horizontal	Vertical
Disseminated	$lsz_{maj}$	$vspc$	$lsz_{min}$	$vspc$	$lsz_{ter}$	$4 \times vspc$
Tabular	$0.5 \times lsz_{maj}$	$0.5 \times vspc$	$0.5 \times lsz_{min}$	$0.5 \times vspc$	$lsz_{ter}$	$0.5 \times vspc$

The experimental variogram points are evaluated to decide which points are going to be considered for the optimization process in the variogram modeling; this weight will be referred to as modeling weight ( $m_{wt}$ ). Simple conditions are set to determine the consistency of the first and last points. The  $m_{wt}$  of each experimental point is set to one as a default value, allowing to change the  $m_{wt}$  to zero after evaluation. Points with  $m_{wt}$  of zero will not be considered in the optimization.

The first point of the variogram is compared to the second point. The weight of the first point is set to zero if the first point has fewer pairs and the variogram value is larger than the second point. This analysis is done for the three directions. This is a common practice in variogram modeling since outliers, proportional effect, or preferential sampling could generate an inconsistent variogram value for short distances. The number of pairs is the support for each lag, and if few pairs are being analyzed, it will not represent the spatial variability in the deposit.



**Figure 3.4:** Experimental variogram points selection for modeling. Left: Experimental variogram showing the number of pairs per lag; the first point will have a modeling weight ( $m_{wt}$ ) of zero. Right: Experimental variogram with some experimental points decreasing as the lag distances increase; these points will have a  $m_{wt}$  of zero.

A second check compares the last points of the variogram in the secondary direction for 2-D or in the tertiary direction for 3-D datasets. The weight of these points will be set to zero if the

variogram value is below the sill and if they decrease as the lag distance increases. Typically, the variogram value increases when the lag distance increases; if this is not observed, it could be related to a hole effect or not enough pairs being analyzed for larger lags. It is common practice to not model these points and focus on the precision of the shorter range. Figure 3.4 shows examples where both checks are applied.

The mentioned constraints are similarly executed for both general and geolocated variograms. For the general variogram, all steps are as described. For the geolocated option, each pair will have a different weight. Weighted formulas for calculating the geolocated version of the variogram, covariance and correlogram consider a standardized weight, as explained in Equations 2.28, 2.30, and 2.33.

The output is the experimental standardized variogram, including the number of pairs, lag distances, standardized variogram value and a binary weight of each point that will be used for the modeling stage.

### 3.4 Automatic modeling

*Autovar* embeds a modified version of *varmodel* in the main subroutine. A detailed explanation of *varmodel* can be found in (J. Deutsch, 2015c). The optimal variogram model is calculated by minimizing the mean square error as a metric. The subroutine compares different randomly generated models against the experimental points; these different models come from small random variations of the modeling parameters.

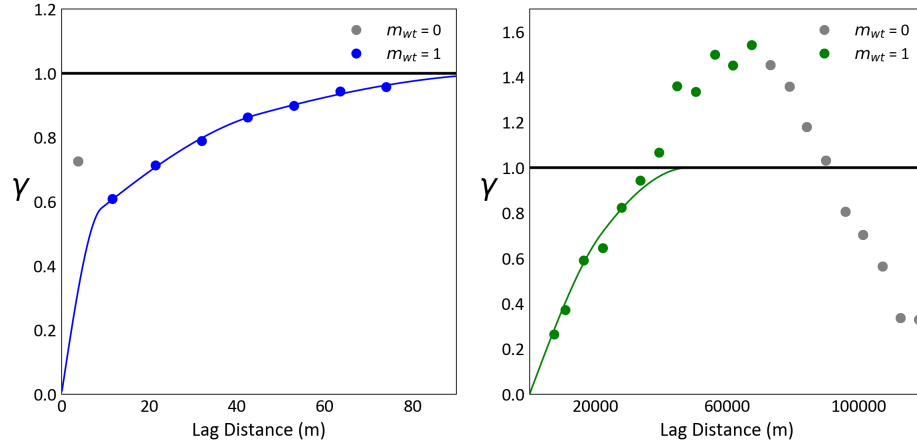
Initial parameters are set as the starting point for the *varmodelfit* subroutine. The number of structures is set to three, the spherical model is set as default, the nugget effect is set to zero, and each of the three contributions is set to 1/3 (the standardized sill is 1). The angles are set to follow the GSLIB convention. This ensures that the fit is performed in the recognized directions of continuity, following the same convention as in the experimental variogram calculation. Three structures give enough flexibility to generate a consistent model.

*Varmodelfit* subroutine takes the output of the experimental variogram: variogram values, lag distances, number of points and modeling weights ( $m_{wt}$ ). Points with values above the sill are set as one since the focus of the *autovar* program is to recognize the variability in the short to medium range. The experimental points are temporally modified inside the subroutine, but the actual experimental point values are reported in the corresponding output file. The subroutine does not model the points that have a modeling weight of zero. Figure 3.5 shows the points that are considered for modeling.

Some internal checks are performed (as in the original *varmodel* code) and hardcoded: the minimum number of points to model is two, and a minimum of ten pairs per point are necessary to consider modeling it. Initial parameters from the main subroutine are used to calculate an initial



objective function. Random shifts change the initial parameters by  $\pm 7.5\%$ . Different combinations of parameters are compared to the experimental points until the one with the smaller mean squared error is selected as the best model. Equation 3.11 shows the objective function for the algorithm.



**Figure 3.5:** Variogram models fitting selected points. Left: Variogram model excluding the first point for the optimization process because it has a modeling weight ( $m_{wt}$ ) of zero. Right: Variogram model that disregards experimental points after 60000 because they have a  $m_{wt}$  of zero.

$$\text{Objective Function} = \sum_{j=1}^{nlag} \sum_{i=1}^{ndir} [\gamma_{(j,i)}e - \gamma_{(j,i)}m]^2 \quad (3.11)$$

Where  $nlag$  is the number of lags,  $ndir$  is the number of directions,  $\gamma_{(j,i)}e$  is the experimental variogram value, and  $\gamma_{(j,i)}m$  is the variogram model value at the same lag distance as the experimental value.

The original *varmodel* program allows to optimize the contributions of structures, the ranges for each structure, the angles of the directions of continuity, and the type of structure (spherical, exponential or Gaussian); in contrast, *autovar* does not need to optimize the directions and the structure is fixed as spherical.

The subroutine performs the following iterative random changes: (1) a random value is used to select if the ranges or structure will be modified; both have the same probability of change (0.5). The contribution of a random structure will be modified if the random value is less or equal to 0.5; otherwise, the range of one random structure in one random direction will be modified.

There are two extra steps when the random iteration is modifying the structure contribution: (1) another random number is generated to pick which structure will be modified using the random shift, and (2) the sills are recalculated to add up to one. Similarly, there are two other steps when changing the range: (1) two different random numbers are used to pick the structure and direction, (2) check if the order relations of the structures are respected, and enforce a change on the modified structure if needed.

The new set of parameters is used to calculate the spherical variogram model value at the same

distances as the experimental variogram points; then, all the points are compared using the mean square error (MSE). If the new set gives a smaller MSE, the set of parameters is saved. If the new set of parameters gives a larger MSE, the set of parameters is discarded, and another random change is made. 2 million random changes are performed because the error is ensured to be small and the time of processing is short. The internal output is the optimized spherical model with the optimum ranges and contributions for each direction and structure.

A future consideration could be to allow the user to set a value for the nugget effect. The program prints the optimized model on the screen: the values of the structures and ranges are displayed. The model is saved in point and text file format following a GSLIB-like convention.

### 3.5 Geolocated function

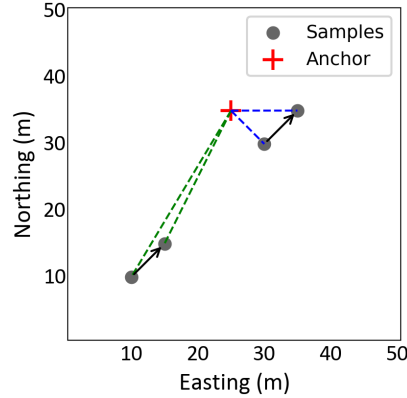
The practitioner picks the location of the anchors to identify local variability in different areas of the domain. These areas could exhibit different directions of continuity or different anisotropy characteristics. The program calculates the general variogram first using all the data points; then, it loops over each anchor to calculate the geolocated variograms.

The program selects points up to 0.3 times the maximum distance (see Equation 3.1) to identify the direction locally. A bigger factor captures regional characteristics, and a smaller factor produces non-representative results while 0.3 showed a consistent and correct identification of directions. After the principal directions are locally identified, the program looks at all pairs and groups them by lag. The extension of the variogram is restricted to one third of the domain because it looks to focus on the characterization of the first lags at each anchor. Each pair will have a different weight. This forces to centre the spatial analysis at the anchor, giving more importance to pairs located at a closer distance without disregarding information for the initial lags from the whole domain.

Figure 3.6 shows two different pairs that are located at different distances from the anchor to illustrate the different weighting concept. The first step is to get the weight of each point with respect to the anchor using a Gaussian weighting function (see Equation 2.25). The second step is to calculate the weight of each pair using the geometric average (see Equation 2.26). The third step is to standardize all the weights per lag distance (see Equation 2.27). The fourth step is to apply the standardized weight to each pair for the correlogram or covariance calculation (see Equations 2.33 and 2.30). Thus, the closer pair will receive a bigger weight.

The calculation of the geolocated variograms is similar to the general variogram. Nevertheless, the calculation and standardization of weights represent an extra step. All possible pairs are analyzed and grouped by lag distance to get the total weight of pairs per lag.

The subroutine calculates the weighted non-ergodic mean and variance of the heads and tails and the geolocated variance. All the variograms are standardized to the sill of one. The respective equations are presented in Section 2.3. The results are printed on the screen if the debug option is



**Figure 3.6:** Schematic representation of weighting for two pairs with the same lag distance but located at different distances from the anchor location (red cross). The pair on the south will receive less weight than the pair closer to the anchor.

selected.

The geolocated variograms are calculated indirectly from the geolocated covariance using Equation 2.36. Although the covariance is less stable than a correlogram, it is preferred since the correlogram could not recognize the anisotropy correctly (Kumara & Deutsch, 2017). Correct recognition of anisotropy is important to characterize local spatial variability, which is why the geolocated option is used.

### 3.6 Model evaluation

The model evaluation option allows to compare the anisotropy on the variogram model. Most deposits in mining or oil exhibit anisotropy, and when the variogram does not detect it, it indicates that the parameters could be improved. This option is recommended only for the disseminated deposit type. The practitioner declares this optional process in Line 11 of the parameter file. Line 12 informs the anisotropy ratios between the major against minor and tertiary; this line is optional and can be skipped even when the model evaluation is active.

In default mode, the program compares the ratio of the major to the minor of the model's third nested structure (*major/minor*). When the ratio is less than 1.1, the program will recalculate the variogram, allowing more points per lag to stabilize the variogram. The 1.1 ratio was selected after testing synthetic and real datasets. The program does not produce better but similar results when a ratio is above 1.1, while smaller ratios generally lead to improvement after the constraints in Table 3.4 are used.

If the practitioner has provided the desired anisotropy ratios in line 12, the program will pick the model that presents the closest ratios to the declared input. A set of more flexible constraints allows more pairs per lag interval when this option is used.

The *varexpflx* subroutine works slightly different than the *varexp* subroutine. *Varexpflx* finds

the relative distance in X, Y, and Z and then uses the rotation matrix to get the components by matrix multiplication of both. Nevertheless, the main difference is in the parameter configuration for the pair selection.

The constraints in Table 3.4 will locate each pair on the respective lag, comparing the components of the major, minor and tertiary directions. Table 3.4 shows the parameters that are different for each deposit type and direction. The variogram is calculated indirectly as well; the program gets a correlogram first, and the variogram is calculated as in Equation 2.37. The factors presented in Table 3.4 were calculated experimentally and are the ones that give experimental variograms that have a balance between precision and stability.

**Table 3.4:** Tolerance parameters for pair selection - correlogram to variogram. Where  $dmaj$  is the current distance of this pair in the major direction,  $lsz_{maj}$  is the major lag size,  $dmin$  is the current distance of this pair in the minor direction,  $lsz_{min}$  is the minor lag size,  $lsz_{ter}$  is the tertiary lag size, and  $vspc$  is the tertiary lag size.

Bandwidth tolerances:	Major direction		Minor direction		Tertiary direction	
	Horizontal	Vertical	Horizontal	Vertical	Horizontal	Vertical
Disseminated	$0.5 \times dmaj$	$4 \times vspc$	$0.5 \times dmin$	$4 \times vspc$	$2 \times lsz_{ter}$	$4 \times vspc$
Tabular	$lsz_{maj}$	$1.5 \times vspc$	$lsz_{min}$	$1.5 \times vspc$	$2 \times lsz_{ter}$	$3 \times vspc$
Tabular and Stratigraphic transformation	$0.25 \times dmaj$	$1.5 \times vspc$	$0.25 \times dmin$	$1.5 \times vspc$	$0.125 \times lsz_{ter}$	$0.5 \times vspc$

A loop evaluates all pairs according to its restrictions (see Table 3.4), and each pair is classified on the corresponding lag. In an attempt to stabilize the variogram, the correlogram is used because it is a robust tool to reduce the influence of outliers; it also ensures the standardization of the variogram; however, it is theoretically incorrect and could mask zonal anisotropy or trends (Kumara & Deutsch, 2017). The user is encouraged to provide the best data possible before using the option to attempt for a more stable variogram.

The program saves the first calculated variogram and proceeds to calculate the second variogram. The program shows the experimental and model results on the screen when the debug mode is selected; however, only the best variogram will be the output. There will be cases in which this option will not detect a better anisotropy. When the program could not calculate a variogram that shows better anisotropy ratios, a variogram with similar behavior for the major and minor will be the output, and it is recommended to tune the parameters manually in programs that allow more customization.

*Autovar* evaluates the general and all geolocated variograms when selecting this option. The execution time increases but not significantly. Table 3.3 shows the parameters for point selection for the initial variogram, and Table 3.4 shows the less restricted parameters for the second search. The optimization function considers the same conditions and constraints as explained in Section 3.4

because the main change must be identified in the experimental variogram.

## 3.7 Outputs

*Autovar* generates outputs in GSLIB format. The main outputs are the experimental variogram and its model, including the angles to define each principal direction. The model has two output files: (1) GSLIB text format and (2) GSLIB point format. The first one is suitable for input to other GSLIB programs such as *kt3dn* and *sgsim*; the point file is for plotting using the *pygeostat* library. The model text file presents the three angles to rotate the coordinates for the GSLIB convention. The model point file gives the angle values for each of the vectors.

The experimental variogram from *autovar* has the same format as the *varcalc*'s output. It presents the following columns: variogram index, lag distance, number of pairs, variogram value, inferred azimuth and inferred dip, which defines each direction of continuity (not the angles in GSLIB convention).

The rotation matrix is also presented as a debug file. It could be plotted as vectors to give an idea of the orientation of the principal directions and to visually assess if the directions are coherent with the dataset.

The weights file gives the coordinate of each point and the weight with respect to the anchor being analyzed; all the weights in this file will be one for the general variogram.

This chapter describes the *autovar* program. It explains the application of several equations and theoretical concepts covered in Chapter 2. The following chapter will show a set of examples of variogram calculation and modeling using *autovar*.

## Chapter 4

# Variograms

---

This chapter summarizes the results of *autovar* in datasets with different characteristics. There are four sections to group different results, including (1) synthetic datasets with regular and irregular distribution of samples, (2) real datasets for disseminated and tabular deposits, (3) a comparison between *autovar* and *varcalc – varmodel*, and (4) an example using the geolocated functionality.

The first group of examples shows the results for synthetic datasets generated with different sets of GSLIB angles. This section shows the consistency of *autovar* to identify the principal directions correctly and produce stable variograms for the synthetic datasets.

The second group of case studies shows the results for real 3-D datasets for disseminated and tabular deposits. The samples on these datasets are irregularly spaced. These datasets have different spatial characteristics that the program identifies. The sample locations and *autovar*'s results are presented.

The third group of case studies compares results between *autovar* and *varcalc – varmodel* since these programs have been used for years, giving consistent results. The geolocated option is not used for these first three groups.

The fourth group of case studies presents the use of the geolocated option for a 2-D synthetic dataset. Another example showing the functionality of the geolocated option can also be reviewed in Chapter 5.

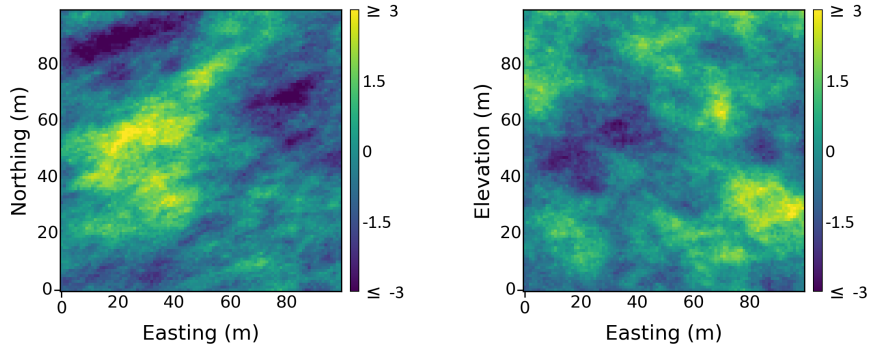
### 4.1 Synthetic datasets

Synthetic datasets are a control group for every test in this thesis because there is a clear idea of what to expect as results. The main goal of these tests is to evaluate the identification of the principal directions of continuity and the variogram structure for every generated dataset. The datasets are generated with simulation using the SGSIM program and then.

The grid definition for each synthetic dataset is 100 nodes with a spacing of 1 meter, where the first node starts on the origin of coordinates for every direction. Consequently, the total number of grid nodes is  $10^6$  per synthetic dataset. The variogram to generate them has no nugget effect and one structure with ranges of 64, 32 and 16 meters on the major, minor and tertiary directions, respectively. The orientation of the principal directions are defined by three GSLIB angles.

The simulations are sampled in a regular grid with a spacing of 5 meters in X, Y and Z or in a irregular pattern. Figure 4.1 shows a simulated dataset, Figure 4.2 and Figure 4.6 show a regular and irregular sampled data.

In this thesis we refer to octant to characterize the location of a vector. An octant is one of the eight divisions of a 3-D coordinate system (Brailov, 2016). For this thesis the first octant is where all coordinates are greater than 0 ( $x \geq 0$ ,  $y \geq 0$ , and  $z \geq 0$ ). In the second octant all coordinates are greater or equal than zero, except for Y ( $x \geq 0$ ,  $y \leq 0$ , and  $z \geq 0$ ). In the third octant all coordinates are less than zero, except for Z ( $x < 0$ ,  $y < 0$ , and  $z \geq 0$ ). In the fourth octant all coordinates are greater than zero, except for X ( $x < 0$ ,  $y < 0$ , and  $z \geq 0$ ). Octants five, six, seven and eight follow the same sequence as the first four described octants, just that in every case the Z coordinates are negative.



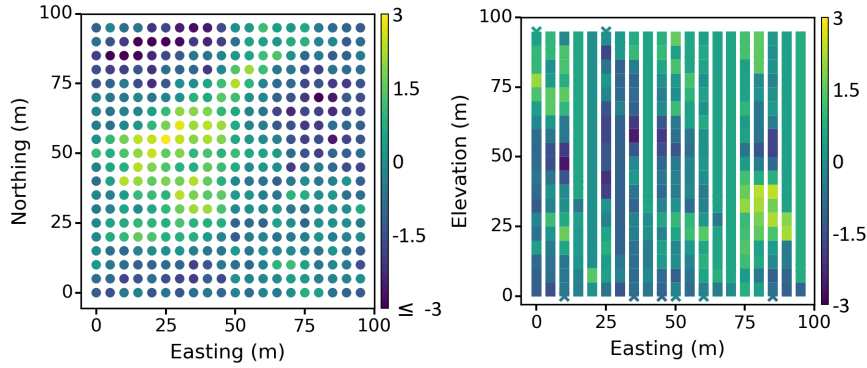
**Figure 4.1:** Synthetic data generation with 40 as the azimuth, 40 as the dip and 40 as the tilt. Left: Plan view of simulated values. Right: XZ view of simulated values.

Five datasets with different continuity directions were generated per octant. Five selected directions were generated per octant, which gives a total of 40 test datasets. Results from sampled values related to only five of them are shown in this section. The angles in this thesis follow the GSLIB convention; the notation used is ‘70,50,10’, where 70 is the azimuth, 50 is the dip, and 10 is the tilt. The selected directions have angles separated every  $20^\circ$  in azimuth, dip and tilt. For instance, the angles that define two directions in the second octant are ‘90,0,0’ and ‘110,20,20’. The simulation process generates values that follow a Gaussian distribution. After the new datasets’ sampling, the behavior is not exactly Gaussian, but it is very similar to it.

#### 4.1.1 Regular sampling

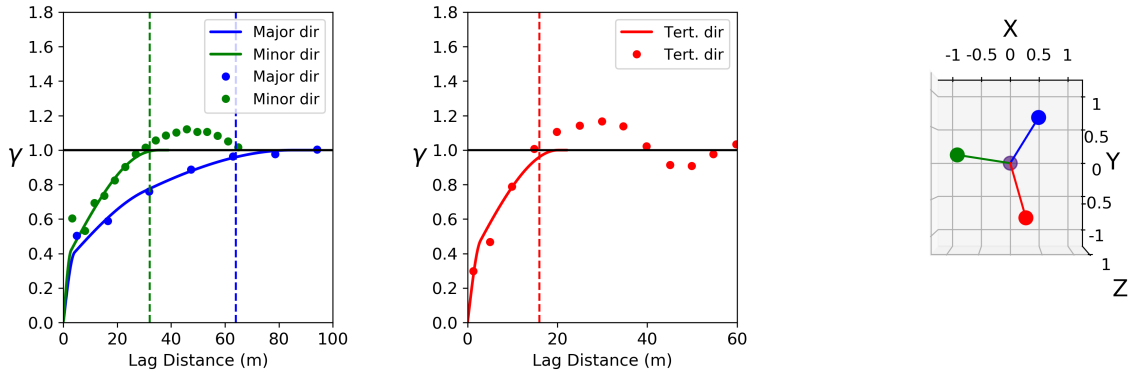
The synthetic simulated datasets are sampled in a regular grid with spacings of 5 meters in X, Y and Z. The number of samples is 8000 per dataset. These datasets were the first set of tests used to evaluate the accuracy of the program to infer the directions and to give the experimental variogram and its model. Figure 4.2 shows the selected samples for the dataset with azimuth, dip and tilt of  $40^\circ$ .

Figure 4.3 shows the variograms and directions for the dataset shown in Figure 4.2. Listing 4.1 shows the variogram model parameters in GSLIB format for Figure 4.3. The inferred angles (‘35,38,40’) are close to the initial set of angles used for generating the datasets (‘40,40,40’). The ranges of each principal direction are consistent with the initial setup; however, the ranges for the



**Figure 4.2:** Regularly spaced samples to test the program. Left: Plan view of sampled values. Right: section looking north showing the selected drillholes.

major are bigger than the original data. The option of attempting to stabilize the variogram (Line 11 of the parameter file) was used because it generated results closer to the original dataset.



**Figure 4.3:** Experimental variogram and its model for samples taken from a dataset with orientation ‘40,40,40’. The dashed lines represents the ranges of the original variogram used to generate the synthetic data. The major direction is in blue, the minor in green and the tertiary in red. The anisotropy relations and the directions are consistent with the truth data.

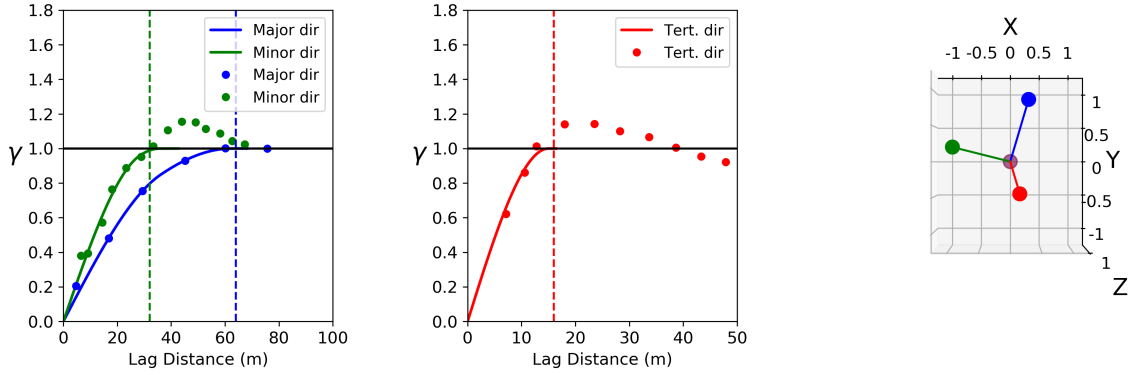
**Listing 4.1:** Variogram model for Figure 4.3

3	0.00				–nst , nugget effect
1	0.34	35.57	38.48	40.36	–it , cc , azm , dip , tilt
	4.58		4.04	4.04	–a_hmax , a_hmin , a_vert (ranges)
1	0.17	35.57	38.48	40.36	–it , cc , azm , dip , tilt
	30.95		30.95	20.13	–a_hmax , a_hmin , a_vert (ranges)
1	0.49	35.57	38.48	40.36	–it , cc , azm , dip , tilt
	84.51		35.41	20.14	–a_hmax , a_hmin , a_vert (ranges)

Figure 4.4 shows the result for the dataset that was generated with ordinary kriging using the principal directions defined by an azimuth, dip and tilt of 20° (‘20,20,20’). List 4.2 shows the variogram model parameters in GSLIB format for Figure 4.4. It is clear from the figure and the text model that the principal directions (‘18,22,16’) and anisotropy relations are accurately recognized



by *autovar*.



**Figure 4.4:** Experimental variogram and its model for samples taken from a dataset with orientation ‘20,20,20’. The dashed lines represents the ranges of the original variogram used to generate the synthetic data. The major direction is in blue, the minor in green and the tertiary in red. The anisotropy relations and the directions are consistent with the synthetic data.

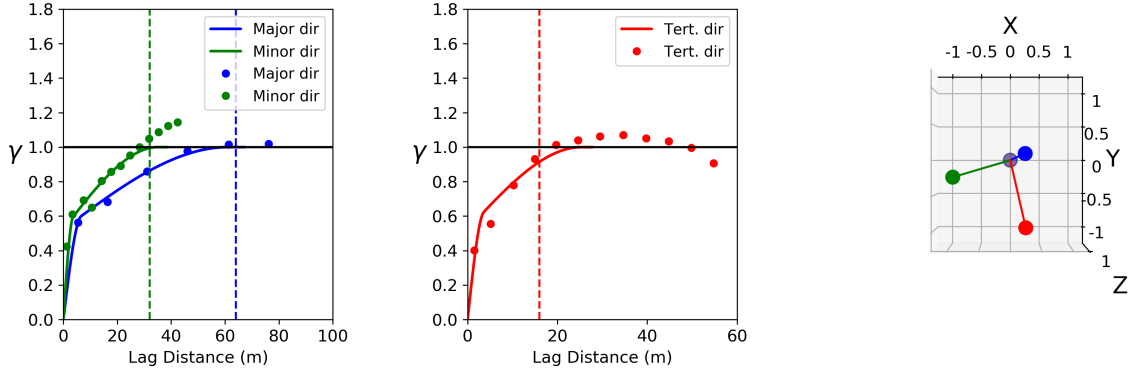
**Listing 4.2:** Variogram model for Figure 4.4

3	0.00				–nst , nugget effect
1	0.34	18.87	22.75	16.65	–it , cc , azm , dip , tilt
		36.02	28.85	15.03	–a_hmax , a_hmin , a_vert (ranges)
1	0.01	18.87	22.75	16.65	–it , cc , azm , dip , tilt
		61.64	39.08	15.02	–a_hmax , a_hmin , a_vert (ranges)
1	0.65	18.87	22.75	16.65	–it , cc , azm , dip , tilt
		62.35	35.79	15.03	–a_hmax , a_hmin , a_vert (ranges)

The results for many of the tests showed consistency with identifying angles similar to its respective synthetic dataset as in the examples of Figures 4.3 and 4.4. However, there are cases where the inferred angles are not as close to the input angles that define the principal directions of the synthetic datasets.

The differences between the angles used to generate the input sampled dataset and the inferred angles could be generated for the selected sampled values as a matter of chance. *Autovar* identifies the directions with less accuracy but it is consistent with the available sampled data. Figure 4.5 and Listing 4.3 show an example of this scenario. Nevertheless, the ranges are quite close to the initial anisotropy relationships.

Figure 4.5 shows the result for the dataset that was generated with kriging with an azimuth, dip and tilt of 80° (‘80,80,80’). The anisotropy is consistent as in the previous examples. There are approximate differences of 12°, 5°, and 3° in the azimuth, dip and tilt, respectively. Nonetheless, these three angles (‘68,75,83’) are consistent with the rotation matrix generated from the sampled values (see Section 2.2). The variogram model presents ranges similar to the variogram used to generate the initial dataset.



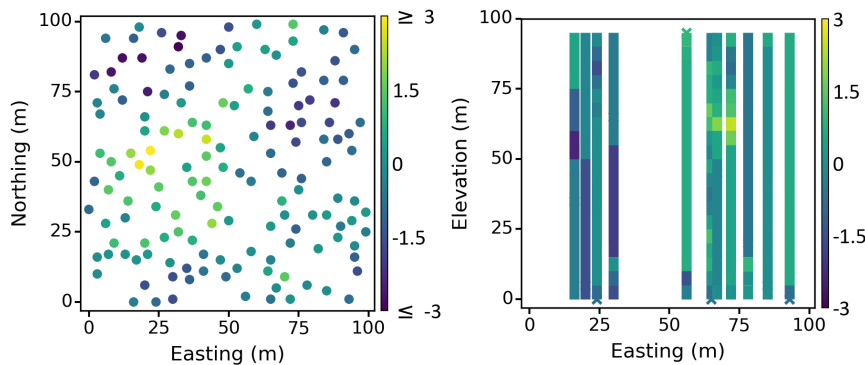
**Figure 4.5:** Experimental variogram and its model for samples taken from a dataset with orientation ‘80,80,80’. The dashed lines represents the ranges of the original variogram used to generate the synthetic data. The major direction is in blue, the minor in green and the tertiary in red. The anisotropy relations are consistent with the sampled data. The inferred angles are consistent with the rotation matrix.

**Listing 4.3:** Variogram model for Figure 4.5

3	0.00				–nst , nugget effect
1	0.52	68.91	75.33	83.50	–it , cc , azm , dip , tilt
	6.63		3.62	3.62	–a_hmax , a_hmin , a_vert (ranges)
1	0.31	68.91	75.33	83.50	–it , cc , azm , dip , tilt
	61.00		35.08	25.31	–a_hmax , a_hmin , a_vert (ranges)
1	0.17	68.91	75.33	83.50	–it , cc , azm , dip , tilt
	61.14		35.13	25.32	–a_hmax , a_hmin , a_vert (ranges)

### 4.1.2 Irregular sampling

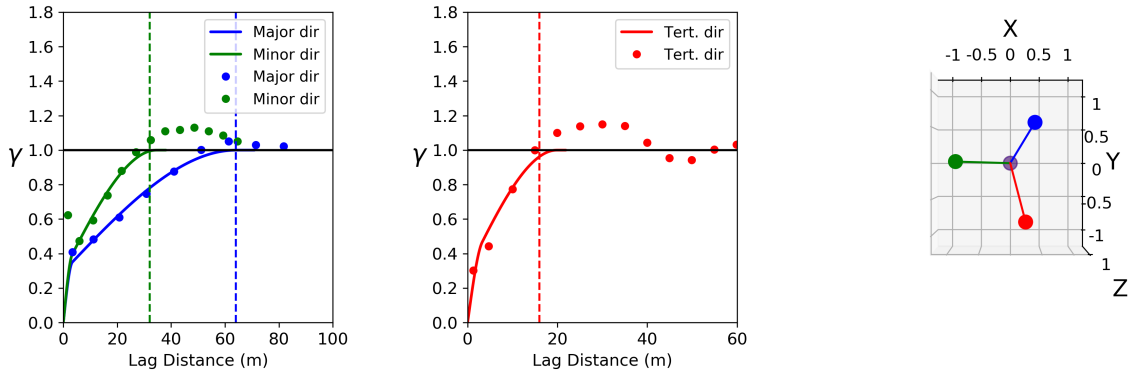
The same synthetic initial datasets (see Figure 4.1) were sampled irregularly for these tests. The purpose of testing irregular datasets is to assess the consistent identification of the principal directions of continuity and variogram structure with unequally spaced samples. These results can be compared to the results of the regularly sampled datasets.



**Figure 4.6:** Irregularly spaced samples to test the program. Left: Plan view of sampled drillhole values. Right: section looking north showing drillholes (section includes drillholes from 0 to 10 meters in the Y direction).

These datasets consisted of 150 vertical drillholes in an irregular pattern, considering a minimum spacing of 4 by 4 meters in X and Y directions, respectively. The spacing in the Z direction is a constant value of 5 meters. This sampling setup produces 3000 samples per irregular spacing dataset. Figure 4.6 shows the selected samples from Figure 4.1, which has principal directions set by an azimuth, dip and tilt of  $40^\circ$  ('40,40,40').

Figure 4.7 shows *autovar*'s experimental variogram points, the variogram model and the representation of the principal directions of continuity for the dataset shown in Figure 4.6. Listing 4.4 shows the variogram model parameters in GSLIB text format for Figure 4.7. Comparing with the principal directions of the synthetic dataset ('40,40,40'), the inferred azimuth has an approximate difference of  $6^\circ$ , but the dip and tilt are closer to  $40^\circ$ . The anisotropy ratios are consistent with the original synthetic dataset and keep a relationship of 3:2:1 (major:minor:tertiary).

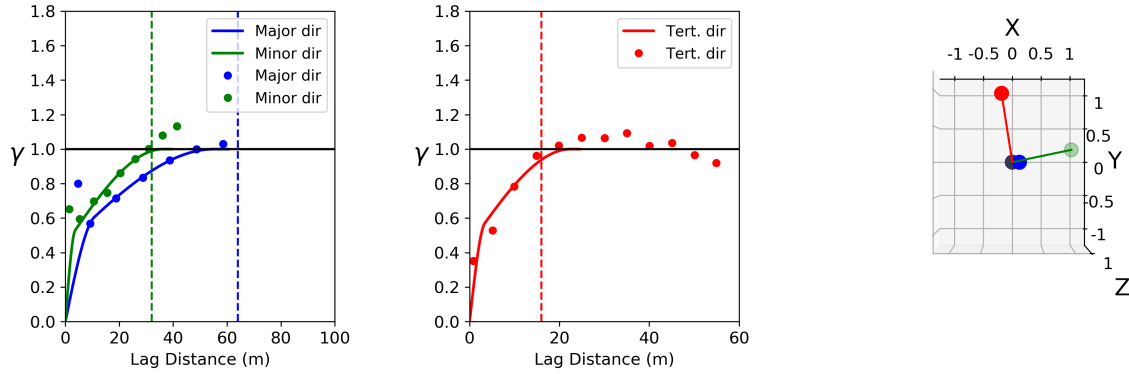


**Figure 4.7:** Experimental variogram and its model for the irregularly sampled from a dataset with orientation '40,40,40'. The major direction is in blue, the minor in green and the tertiary in red. The anisotropy relations and the directions are consistent with the truth data.

**Listing 4.4:** Variogram model for Figure 4.7

3	0.00				-nst , nugget effect
1	0.30	34.92	46.31	42.44	-it , cc , azm , dip , tilt
	3.34		3.31	3.25	-a_hmax , a_hmin , a_vert (ranges)
1	0.62	34.92	46.31	42.44	-it , cc , azm , dip , tilt
	64.45		34.58	19.89	-a_hmax , a_hmin , a_vert (ranges)
1	0.09	34.92	46.31	42.44	-it , cc , azm , dip , tilt
	60.75		34.48	19.80	-a_hmax , a_hmin , a_vert (ranges)

Figure 4.8 shows *autovar*'s result for the irregularly sampled dataset from a synthetic dataset with principal directions defined by an azimuth, dip and tilt of  $80^\circ$  ('80,80,80'). The plot of the sampled data is not presented, but it is similar to the plot in Figure 4.7. Listing 4.5 shows the variogram model parameters in GSLIB format for figure 4.8. The principal directions and anisotropy relations are closer to the directions of the synthetic dataset, and *autovar* recognizes the directions from the available samples.



**Figure 4.8:** Experimental variogram and its model for the irregularly sampled from a dataset with orientation ‘80,80,80’. The major direction is in blue, the minor in green and the tertiary in red. The anisotropy relations and the directions are consistent with the synthetic data.

**Listing 4.5:** Variogram model for Figure 4.8

3	0.00				–nst , nugget effect
1	0.44	90.73	83.49	–79.06	–it , cc , azm , dip , tilt
	4.69		2.99	2.99	–a_hmax , a_hmin , a_vert (ranges)
1	0.34	90.73	83.49	–79.06	–it , cc , azm , dip , tilt
	55.21		36.03	22.86	–a_hmax , a_hmin , a_vert (ranges)
1	0.22	90.73	83.49	–79.06	–it , cc , azm , dip , tilt
	55.25		36.05	21.80	–a_hmax , a_hmin , a_vert (ranges)

The presented examples of synthetic datasets for regular and irregular sampled spacing show the consistency of the program in calculating and modelling the variogram in the principal directions of continuity. The models that come from the same synthetic dataset are fairly similar independently of the sampling configuration (see Figures 4.3 and 4.7). The variogram model could be fit with less than three structures in some of the presented cases. However, this could generate weak fitting for a more complicated set of experimental variogram points. The following section shows other examples of the variogram calculation and modeling of real datasets using *autovar*.

## 4.2 Real datasets

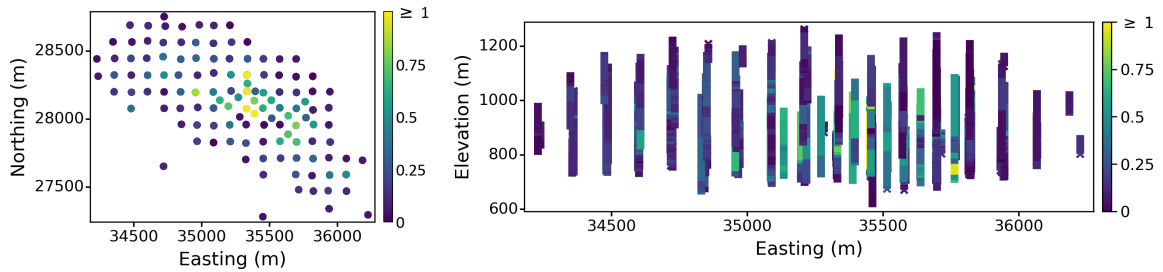
Real datasets give a better idea of the program’s functionality, while synthetic datasets are valuable at the initial stages of testing a program. Real datasets usually present irregularly spaced samples, which could be sparse, scarce, or both. This section presents examples of real datasets of disseminated and tabular deposits.

### 4.2.1 Disseminated deposits

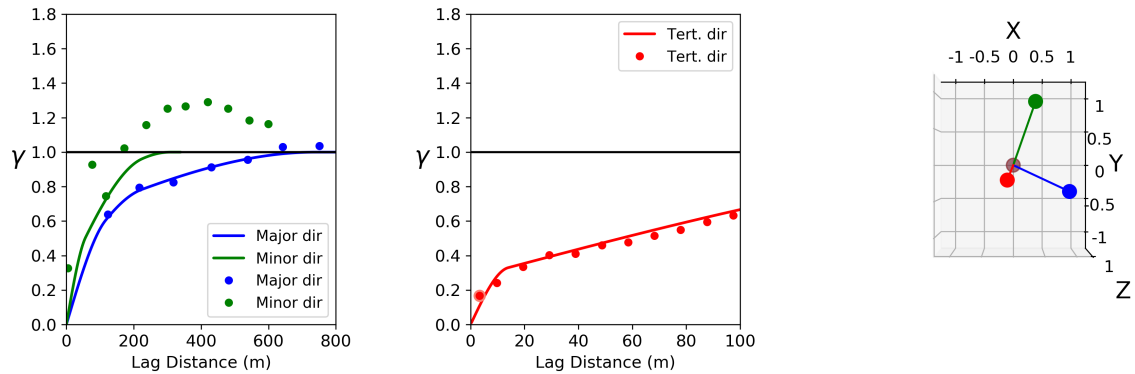
Two datasets were tested for disseminated deposit types: Large and Training datasets. These datasets have porphyry characteristics such as big volume and predominance of low grades. The type

and depth of drilling and sampling space of the drillholes are also characteristics of these deposits.

The Large dataset represents a copper-molybdenum porphyry deposit. There are sparse samples in some areas; however, only the area with the highest density of samples is evaluated. The dataset consists of 124 vertical drillholes with an average length of 240 meters. The sampling spacing downhole is 10 meters, and the average sampling spacing in the horizontal plane is 107 meters. There are 3107 samples for copper with an average of 0.262 and 0.053 variance. Figure 4.9 shows the data location of the sampled values. Figure 4.10 shows the variogram outputs using *autovar*. Listing 4.6 shows the variogram model parameters for Figure 4.10. In the second structure there all the ranges are the same because of the order relation rules implemented on the code.



**Figure 4.9:** Left: Plan view of the Large dataset for the copper variable. Right: XZ view of the large dataset. The range scales exhibits relatively low anisotropy, which is characteristic of a disseminated deposit.

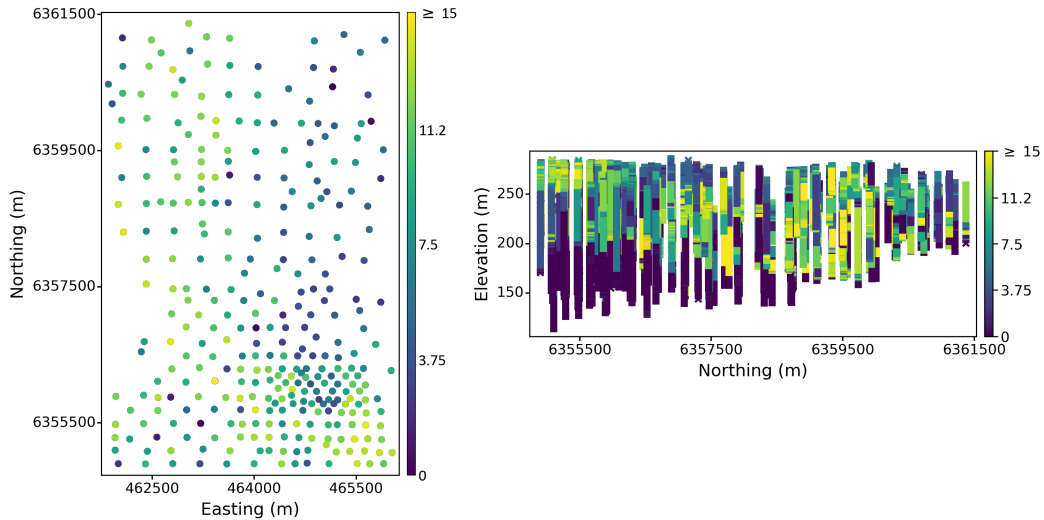


**Figure 4.10:** Variogram, model and directions for the Large dataset using *autovar* program. The anisotropy relationships are typical for a disseminated deposit.

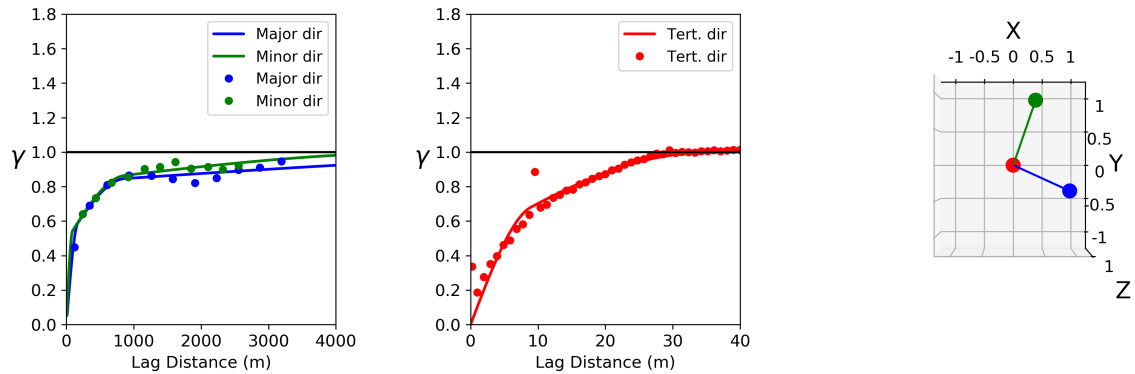
**Listing 4.6:** Variogram model for Figure 4.10

3	0.00				—nst, nugget effect
1	0.26	112.10	0.63	13.01	—it, cc, azm, dip, tilt
		105.65	58.28	7.87	—a_hmax, a_hmin, a_vert (ranges)
1	0.36	112.10	0.63	13.01	—it, cc, azm, dip, tilt
		221.81	221.81	221.81	—a_hmax, a_hmin, a_vert (ranges)
1	0.38	112.10	0.63	13.01	—it, cc, azm, dip, tilt
		742.57	305.96	305.96	—a_hmax, a_hmin, a_vert (ranges)

The Training dataset consists of 322 vertical drillholes with an average length of 90 meters. The sampling spacing downhole is 1 meter, and the average sampling spacing in the horizontal is 276 meters. The variable being analyzed is gold, with 30774 samples with an average of 6.361 and 29.344 variance. Figure 4.11 shows the data location of the sampled values. Figure 4.12 shows the variogram outputs using *autovar*. The Listing 4.7 shows the variogram model for Figure 4.12.



**Figure 4.11:** Left: Plan view of Training dataset for the gold variable. Right: YZ view of the training dataset with vertical exaggeration for visualization purposes.



**Figure 4.12:** Variogram, model and directions for the training dataset using *autovar* program

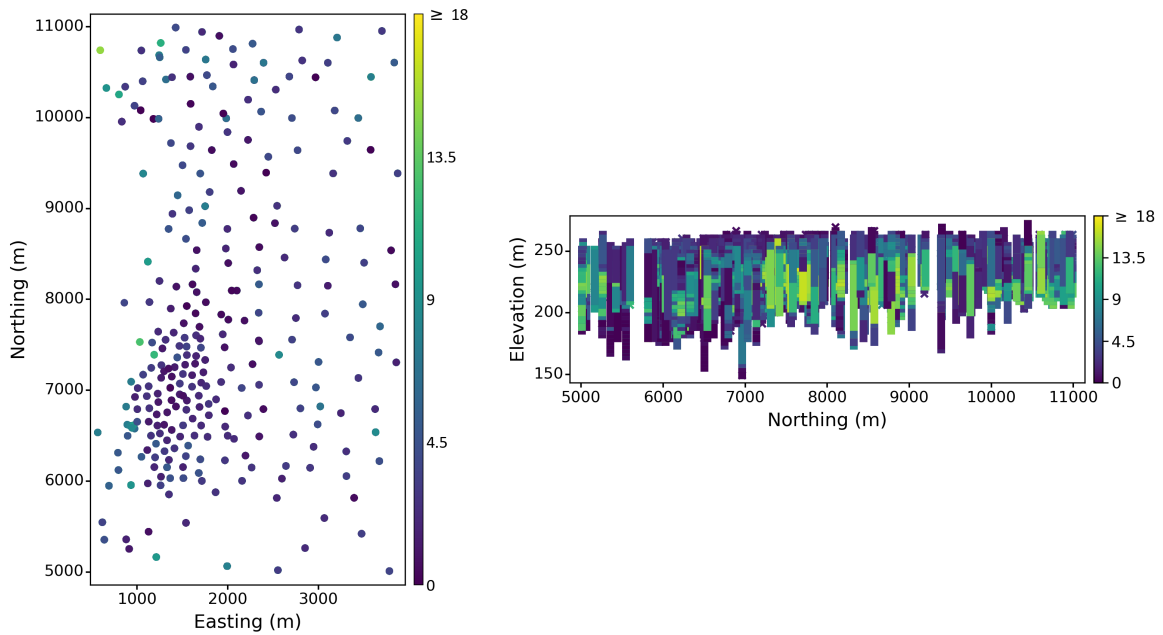
**Listing 4.7:** Variogram model for Figure 4.12

3	0.00				—nst , nugget effect
1	0.49	111.60	0.12	0.05	—it , cc , azm , dip , tilt
	170.74		91.90	9.00	—a_hmax , a_hmin , a_vert (ranges)
1	0.34	111.60	0.12	0.05	—it , cc , azm , dip , tilt
	865.53		865.53	37.00	—a_hmax , a_hmin , a_vert (ranges)
1	0.18	111.60	0.12	0.05	—it , cc , azm , dip , tilt
	10001.97		5519.27	30.77	—a_hmax , a_hmin , a_vert (ranges)

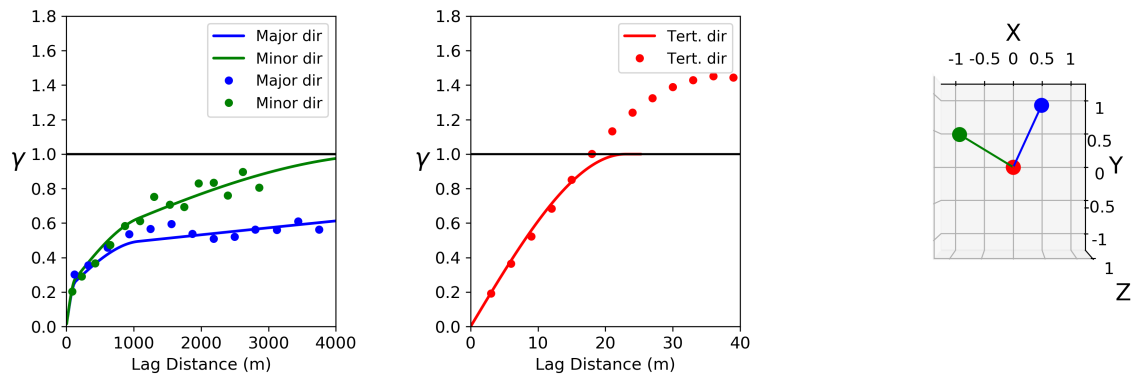
The results of *autovar* for both disseminated datasets are representative of their spatial continuity. It provides the correct identification of principal directions and variograms that show structure in each direction. The attempt to stabilize the variogram option has been used in the first case.

### 4.2.2 Tabular deposits

Tabular deposits present significantly more continuity in the horizontal directions (major and minor) than in the vertical (tertiary) direction. This difference makes the variogram very sensitive to the vertical bandwidth tolerance parameter. The program implements a more restrictive vertical bandwidth for tabular deposits. This section presents two examples: Oilsands and Conklin datasets.



**Figure 4.13:** Left: Plan view of Oilsands dataset for the bitumen variable. Right: YZ view of the Oilsands dataset, which presents a vertical exaggeration in the vertical for visualization purposes.

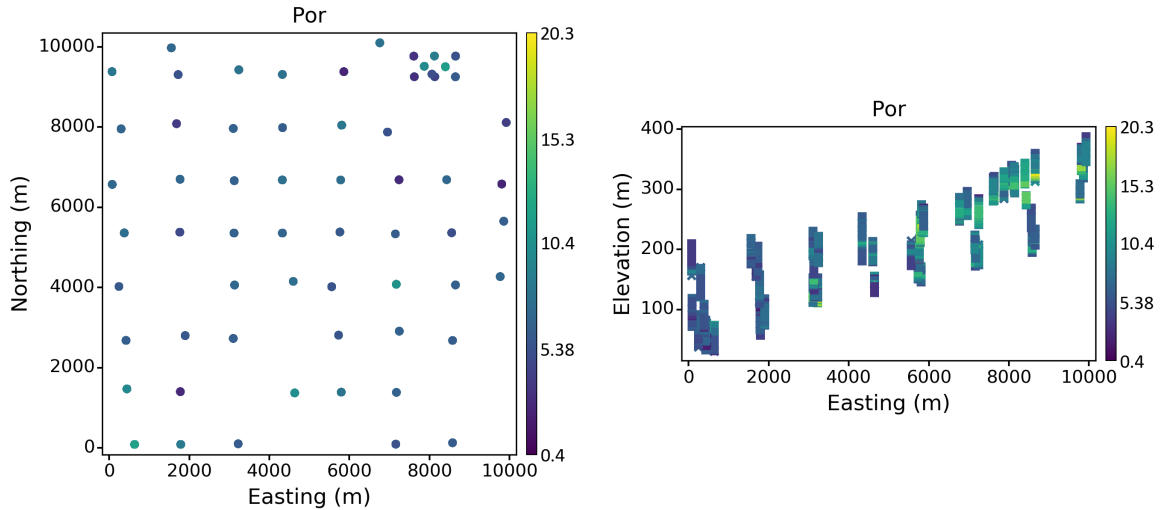


**Figure 4.14:** Variogram, model and directions for the Oilsands dataset using *autovar* program. The major directions exhibits local anisotropy.

The Oilsands dataset is a clear example of the differences in continuity in different directions. The anisotropy relations of the major to the tertiary is around 50-100:1 while major to minor is around 1:2. The dataset consists of 280 vertical drillholes with an average length of 59 meters. The sampling spacing downhole is 3 meters, and the average sampling spacing in the horizontal is 235 meters. The variable being analyzed is bitumen, with 5808 samples with an average of 7.7 and 26.3 variance. Figure 4.13 shows the location map, and Figure 4.14 shows the variograms and directions using *autovar*. Listing 4.8 shows the variogram model in text format for the Figure 4.14.

**Listing 4.8:** Variogram model for Figure 4.14

3	0.00				-nst , nugget effect
1	0.23	27.83	0.14	0.16	-it , cc , azm , dip , tilt
	111.30		90.28	22.99	-a_hmax , a_hmin , a_vert (ranges)
1	0.22	27.83	0.14	0.16	-it , cc , azm , dip , tilt
	1114.62		1114.62	22.99	-a_hmax , a_hmin , a_vert (ranges)
1	0.55	27.83	0.14	0.16	-it , cc , azm , dip , tilt
	20014.62		4874.58	23.00	-a_hmax , a_hmin , a_vert (ranges)

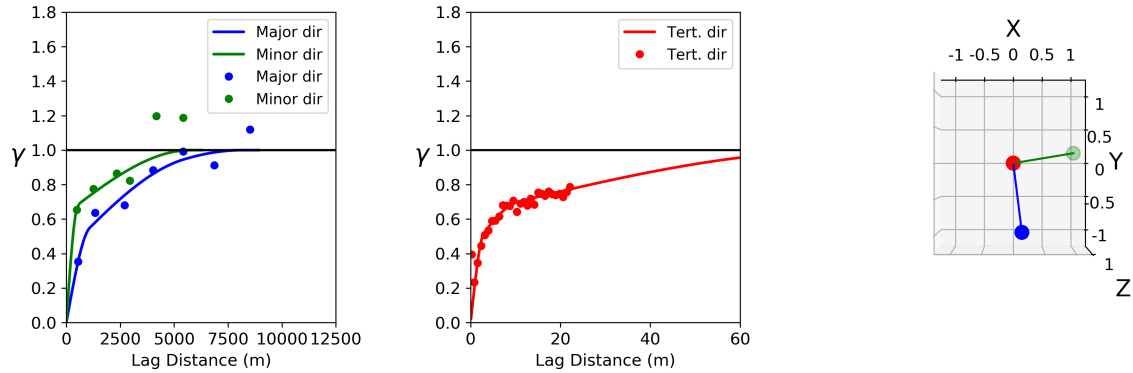


**Figure 4.15:** Left: Plan view of Conklin dataset collars for the porosity variable, a cluster of collars is visible on the northeast area. Right: XZ view of the Conklin dataset showing the difference on the vertical position of the collars of each drillhole.

The Conklin dataset represents a tabular dataset with 62 vertical drillholes and an average length of 51 meters. The sampling spacing downhole is 1 meter, and the average sampling spacing in the horizontal is 1500 meters. The variable being analyzed is porosity, with 3257 samples with an average of 8.3 and 11.4 variance. This dataset is more challenging for automatic analysis because it presents a clustering on the Northeast area and the collars of the drillholes are in an inclined surface. Figure 4.15 shows the location map for the variable porosity. Figure 4.16 shows the variograms and principal directions of continuity using *autovar*. The stratigraphic transformation option has been



used automatically by the program. Listing 4.9 shows the variogram model in text format for the Figure 4.16.



**Figure 4.16:** Variogram, model and directions for the Conklin dataset using *autovar* program.

**Listing 4.9:** Variogram model for Figure 4.16

3	0.00				–nst , nugget effect
1	0.41	171.83	0.07	–0.10	–it , cc , azm , dip , tilt
	889.44	408.23		2.98	–a_hmax , a_hmin , a_vert (ranges)
1	0.21	171.83	0.07	–0.10	–it , cc , azm , dip , tilt
	5442.51	2326.69		9.97	–a_hmax , a_hmin , a_vert (ranges)
1	0.39	171.83	0.07	–0.10	–it , cc , azm , dip , tilt
	7923.91	5734.48		75.25	–a_hmax , a_hmin , a_vert (ranges)

The results of *autovar* for both tabular datasets are reasonable. It provides the right identification of principal directions and variograms that show structure. The ranges of the variograms reflects the characteristics of tabular deposits. The next section presents a comparison between *autovar* and *varcalc – varmodel*.

The execution time for the program using a disseminated 3-D dataset with 3107 points, including the option of attempting to stabilize the variogram, is 9.4 seconds. For the same dataset without the stabilization option, the program has an execution time of 4.6 seconds. The program ran another 3-D dataset with 30774 without attempting to stabilize the variogram in 13.2 seconds. Considering the last two cases where no attempt to stabilize the variogram is made, the number of samples grows approximately ten times while the execution time of the program grows three times. The program has similar or slightly less execution times for tabular deposits.

### 4.3 Comparison with *varcalc* and *varmodel*

This section compares the results between *autovar* and *varcalc – varmodel* (GSLIB programs) to show the similarities and differences between both. It could be challenging to select good parameters to get a variogram model that reasonably represents the spatial continuity of the variable. *Autovar*

prompts the user with fewer parameters than *varcalc – varmodel*, which is convenient for new practitioners and reduces the variogram’s calculation and modeling.

A variogram calculation and modeling require several steps. Rossi and Deutsch (2013) recommend some requirements before calculating a variogram: (1) data visualization and understanding considering the geology, (2) definition of the coordinate system, and (3) outlier management and data transformation. These requirements are important to met for any dataset at the beginning of the workflow.

The next step is to identify the directions of continuity. Several tools are available, such as the variogram maps, the moment of inertia, and neutral models (Kim & Deutsch, 2022). The geology interpretation will provide valuable insight into correctly identifying the directions and the positioning of the data (Rossi & Deutsch, 2013).

Then, all the parameters for pair selection must be established as detailed in Section 2.1. After calculating the experimental variogram points, they are fitted with a positive definite model. The *varcalc-varmodel* methodology considered in this thesis has the following steps:

1. Identification of directions of continuity: visually or other options (*varmap* program).
2. Experimental variogram calculation: set all paramaters in *varcalc*.
3. Semiautomatic variogram modeling using *varmodel*.
4. Evaluate if the model represents the spatial variability of the domain.

The best characteristics of *varcalc* and *varmodel* are the high customization of options, parameters and constraints according to the deposit characteristics. A complete explanation of both CCG programs can be found in previous CCG papers (J. Deutsch, 2015c).

The methodology of *autovar* is described in Chapter 3. Essentially, *autovar* infers the principal directions of continuity, calculate the experimental variogram points and fit a positive definite model to them. All these steps are automatic and speeds the spatial characterization process for disseminated and tabular deposits. The practitioner needs to evaluate the final result to see if it is consistent with the characteristics of the deposit.

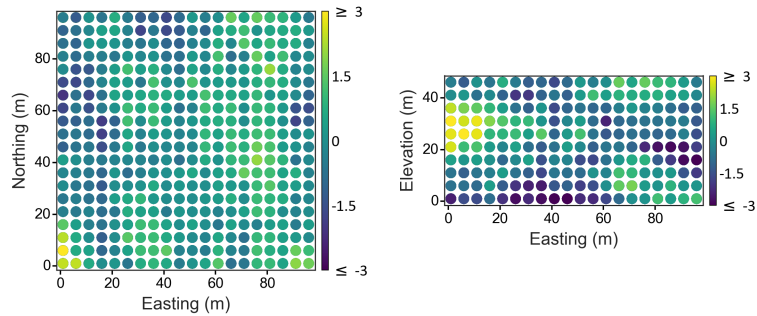
A major difference between both workflows is *autovar*’s geolocated option, which allows centring the variogram calculation with respect to anchors the user selects. Nevertheless, just the general variogram will be compared with *varcalc – varmodel*.

### 4.3.1 Comparison using a synthetic example

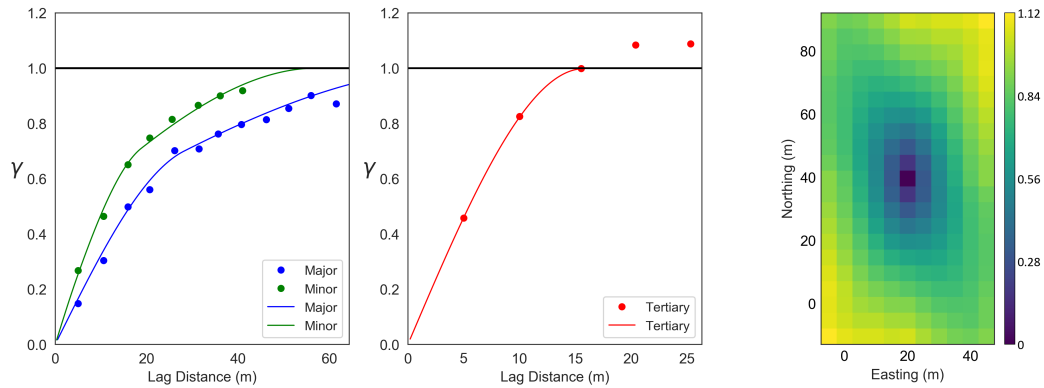
The first example is a synthetic dataset generated by sequential Gaussian simulation. The dataset is regularly sampled every 5 meters in each direction to get the sampled data. Both workflows

are tested with the same sampled data. Figure 4.17 shows the samples, and Figure 4.18 shows the results using *varmap*, *varcalc* and *varmodel*.

Nevertheless, the angles were known, the principal directions were calculated using *varmap*, and they were set to the corresponding identified angles ('170,0,0'). The lags are set according to the sampling spacing by direction (5 meters for all directions), and lag tolerances are half that value; bandwidths had the same sizes as lag sizes. Figure 4.19 shows the result of *autovar*. Only basic



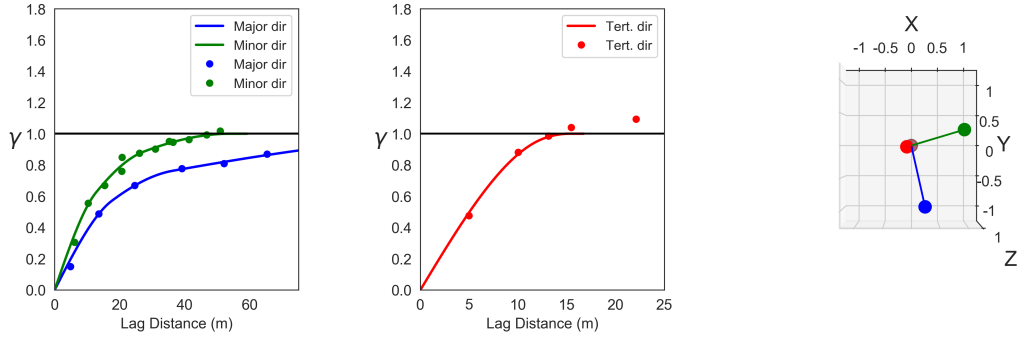
**Figure 4.17:** Location plot of the synthetic dataset generated for *autovar* and *varcalc* – *varmodel* comparison. Left: Plan view of the synthetic dataset. Right: XZ view of the synthetic dataset.



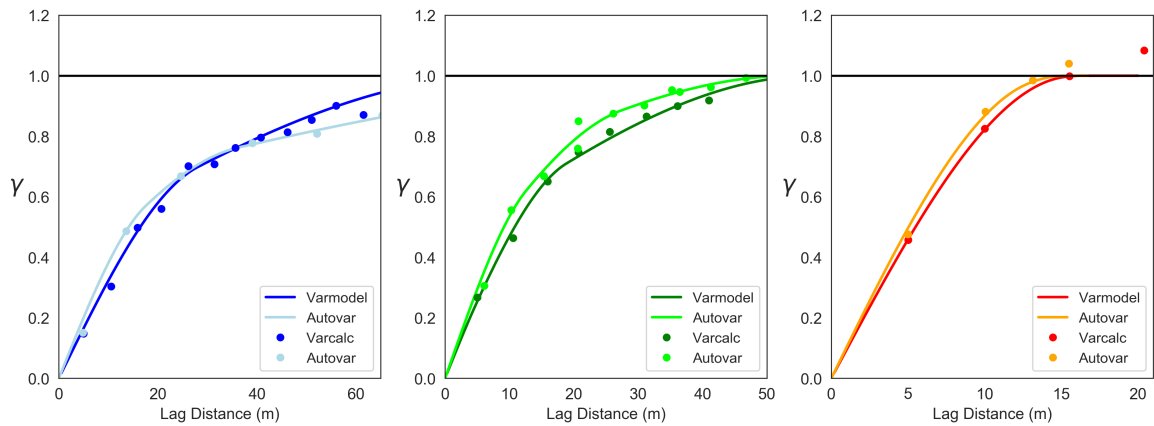
**Figure 4.18:** *Varcalc* – *varmodel* results for dataset in Figure 4.17. Left: Experimental variogram points and their models for the major and minor direction. Centre: Experimental variogram points and fits models for the tertiary direction. Right: Plan view of the variogram map.

parameters are declared by the user. Comparing Figures 4.18 and 4.19 is clear that the variogram map and the directions inferred by *autovar* are very similar. The variogram map shows  $170^\circ$  as the principal direction. This direction coincides with the major direction inferred by *autovar*, which is slightly less than  $180^\circ$ . The same similarity is observed for the minor and tertiary directions.

Figure 4.20 shows results from *varcalc* – *varmodel* and *autovar* in one plot. The results for the major are similar, while the model on the minor and tertiary shows differences. Nevertheless, the points behavior are similar in all directions. The lag sizes inferred by *autovar* are slightly larger than those used for *varcalc* and *varmodel*.



**Figure 4.19:** Autovar results for dataset in Figure 4.17. Left: Experimental variogram points and their models for the major and minor directions. Centre: Experimental variogram points and their models for the tertiary direction. Right: Principal directions recognized by *autovar* (major in blue, minor in green, and tertiary in red).

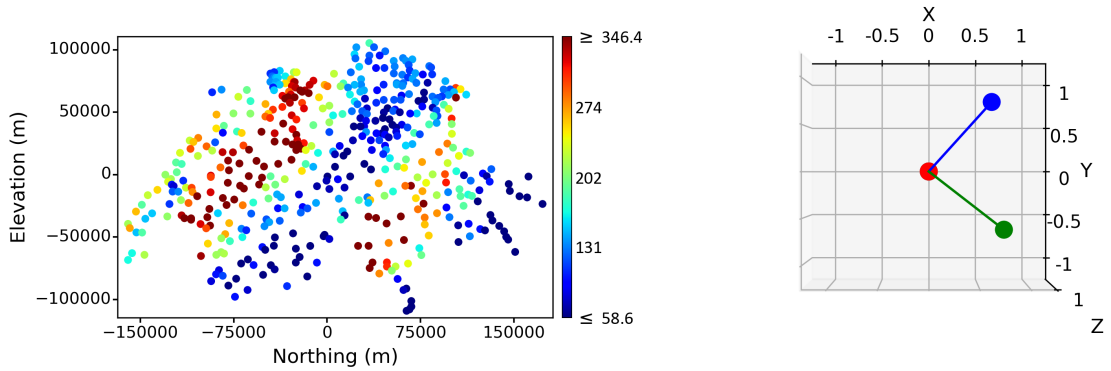


**Figure 4.20:** Comparison of *autovar* and *varcalc - varmodel* results. Left: Comparison of experimental variogram points and its models for the major direction. Centre: Comparison of experimental variogram points and its models for the minor direction. Right: Comparison of experimental variogram points and its models for the tertiary direction.

### 4.3.2 Comparison using a 2-D real dataset

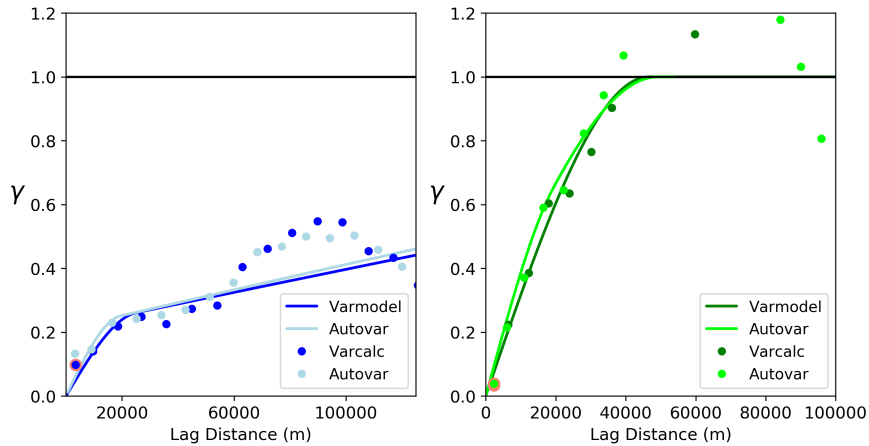
The following example is a real 2-D dataset that shows clear directions of continuity and cyclicity. Dubois (1998) describes the Spatial Interpolation Comparison (SIC) dataset. This dataset consist of measurements of rainfall in Swiss. There are different websites to access it; however, the ‘SIC’ dataset used in this thesis was retrieved online from (Clark & Harper, 2020). The anisotropy is obvious and has an approximate relation of 5:1 (major: minor); the zonal anisotropy is shown in the variograms calculated with both workflows. The principal directions of continuity are clear to identify visually. Figure 4.21 shows the distribution of the values and the principal directions of continuity calculated by *autovar*; they are consistent with the visually identified directions of continuity.

Figure 4.22 compares results for the *varcalc - varmodel* against *autovar*. For the *varcalc - varmodel* workflow, the lag size is 9000 in the major direction and 6000 in the minor direction, while the lag tolerances are half each value, respectively. The azimuth of the principal direction is set to



**Figure 4.21:** Location plot of 2-D real dataset and plot of principal directions. Left: Plan view of 2-D real dataset. Right: Plot of principal directions: major in blue and minor in green.

40° after a visual analysis. The horizontal bandwidth tolerances have the same values as the lag size. The only parameters given to the *varmodel* are the angles for the directions of continuity; then, the program automatically calculates the contributions and ranges for each of the three structures.



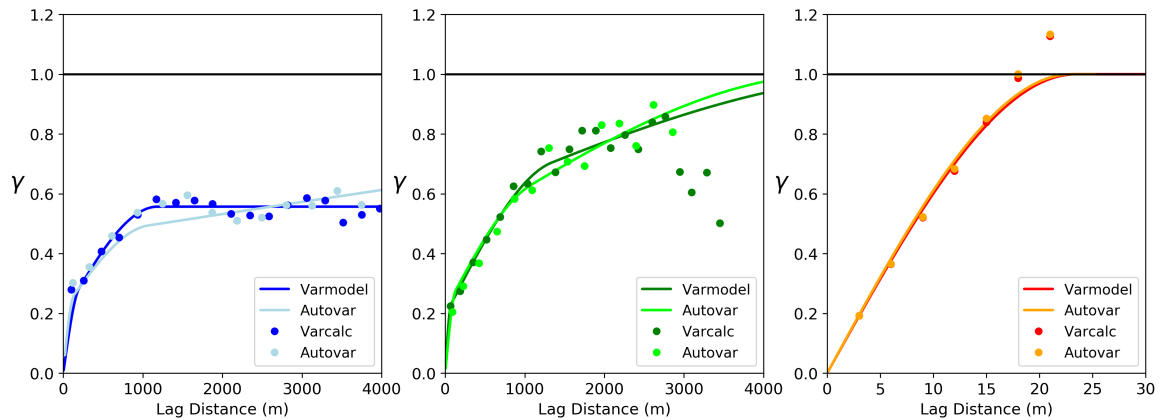
**Figure 4.22:** Comparison of *autovar* and *varcalc* – *varmodel* for real 2-D dataset in Figure 4.21. Left: Comparison of experimental variogram points and its model for the major direction. Right: Comparison of experimental variogram points and its models for the minor direction.

The experimental points for the major direction show similar behavior but are not the same lag distance since there are differences in the tolerances for pair selection in each workflow. Consequently, the automatically fitted models show very similar behavior.

There are slightly more differences between the experimental points of both workflows in the minor direction. However, the models converge to a similar behavior in the initial part of the variogram, and when it reaches the sill. This difference is related to the horizontal bandwidth settings for both workflows.

### 4.3.3 Comparison using a 3-D real dataset

The last example compares the results of both methodologies for the Oilsands dataset (see Figure 4.13). Figure 4.23 compares the results of both workflows. The experimental points of the major and minor show similar behavior, while the points in the third direction exhibit almost equal behavior and virtually the same model. The model of the major and minor directions exhibits more differences due to the more severe constraints of *autovar* for the maximum modeling ranges with respect to *varmodel*. This difference is noticeable when the experimental variogram does not reach the sill of one. Otherwise, the behavior will be similar to *varmodel*, which model has a better fit in the major direction.



**Figure 4.23:** Comparison of *autovar* and *varcalc* – *varmodel* for Oilsands dataset. Left: Comparison of experimental variogram points and its models for the major direction. Centre: Comparison of experimental variogram points and its models for the minor direction. Right: Comparison of experimental variogram points and its models for the tertiary direction.

Both workflows have advantages, and the different examples show similar results. The flexibility to pick parameters with *varcalc* and *varmodel* shows that different applicability and customization can be set to have the best possible variogram. Nevertheless, it requires some expertise to tune it correctly.

Another advantage of *varcalc* is that it could calculate different measures of spatial continuity, such as variogram, covariance, correlogram, indicator variogram (continuous or categorical variables), cross-variogram, and others.

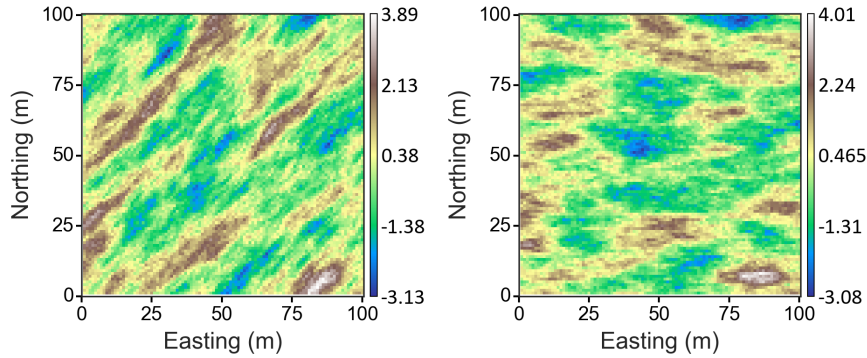
Additionally, *varcalc* allows to generate more than one of the spatial measures of continuity at once, while *autovar* results are only set for the classic variogram, and consequently using one variable at the time.

The main advantage of *autovar* is the automation and speed to calculate a variogram for the variable. This gives an initial idea of the spatial variability characteristics of the deposit; additionally, the few parameters that need to be declared make it suitable for new practitioners. The next section describes another useful functionality of *autovar*.

#### 4.4 Geolocated option

The geolocated option is useful when there are different directions of continuity in some areas of the dataset. The practitioner needs to inform the number and location of the anchors. The program will calculate geolocated variograms for each anchor. The following is an example of a 2-D dataset with two anchors.

This example was generated by mixing two synthetic datasets. Both datasets have the same anisotropy relations of 3:1 (major:minor) and ranges, but the directions of continuity are different. These datasets are generated using the GSLIB *sgsim* program with different random seeds. Figure 4.24 shows the initial datasets with major directions given by GSLIB angles of  $45^\circ$  ('45,0,0') and  $90^\circ$  ('90,0,0').

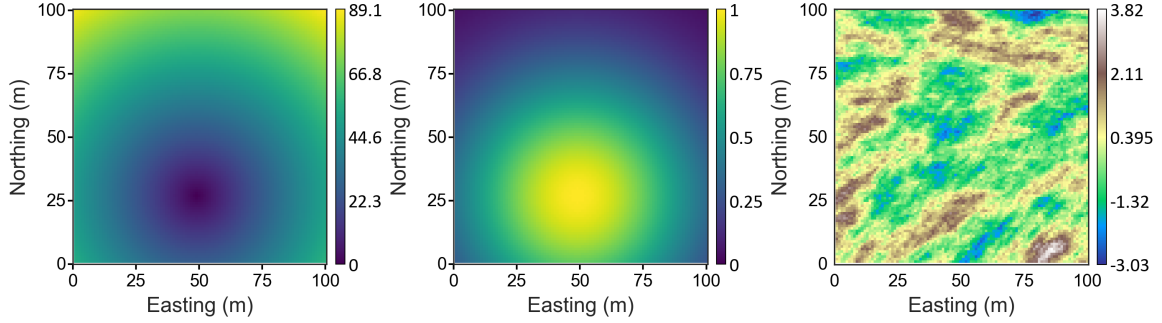


**Figure 4.24:** Synthetic data generation of two datasets with different directions of continuity. Left: Plan view of the first simulated dataset with a principal direction of  $45^\circ$ . Right: Plan view of the second simulated dataset with a principal direction of  $90^\circ$ .

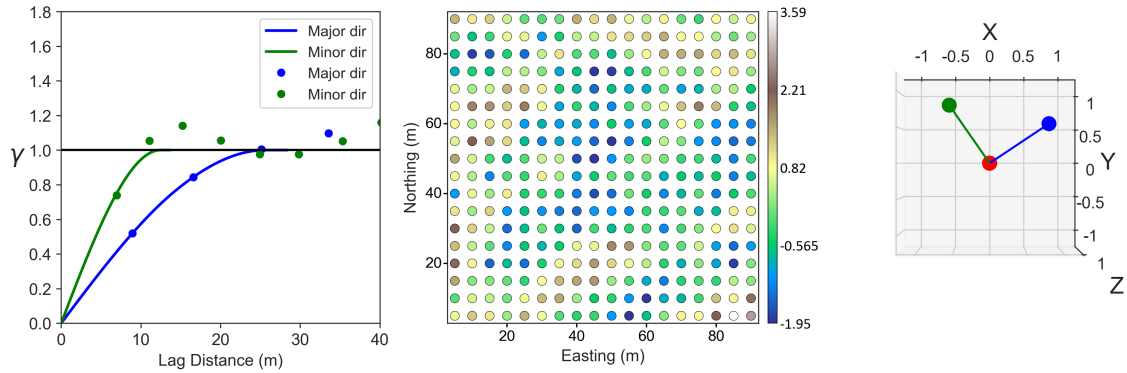
A Gaussian weighting function using the distance to the anchor is generated to assign weights when mixing both datasets. Figure 4.25 shows the distances and weight with respect to a point in the south ( $x_1 = 49, y_1 = 27$ ). A similar set of weights are calculated for a point located in the north ( $x_2 = 49, y_2 = 71$ ). Then, the weights are standardized to add up to one because it ensures a smooth mixing of both datasets. These weights are multiplied by the values of the corresponding datasets ( $45^\circ$  for the south and  $90^\circ$  for the north). Both weighted datasets are added together. Figure 4.25 shows the final mixed dataset.

The mixed dataset shows different orientations in the south and north areas, which is suitable to be analyzed using the geolocated option. This dataset was regularly sampled (5 by 5 meters) to analyze with *autovar*. The general variogram (see Figure 4.26) and the geolocated variograms for the south (see Figure 4.27) and north (see Figure 4.28) are shown alongside the dataset and the anchor's position (red cross). The identified directions for each anchor are plotted to the right of each image.

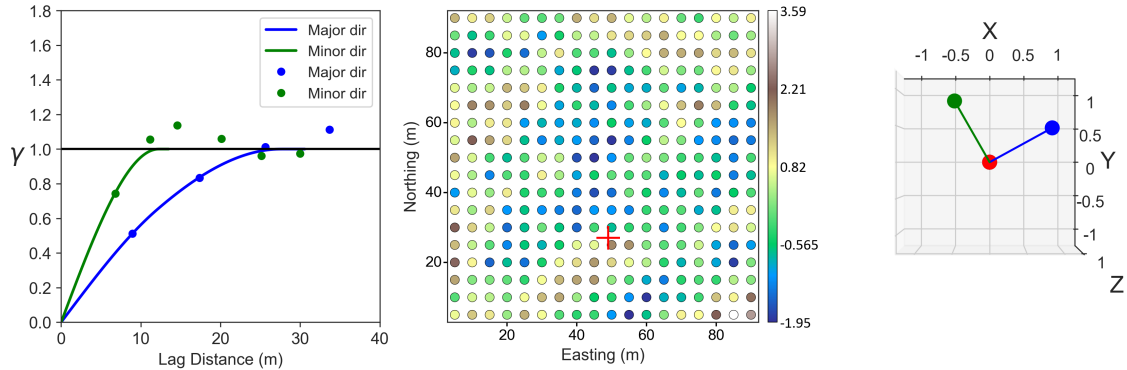
The general variogram (see Figure 4.26) considers all pairs with the same weight, the principal direction is  $55^\circ$ ('55,0,0'). The variogram shows structure in both directions at the first meters. Both



**Figure 4.25:** Mixed synthetic datasets generation. Left: Plan view of Gaussian weights for the simulated dataset with principal direction of  $45^\circ$ . Centre: Plan view of Gaussian weights for the simulated dataset with principal direction of  $90^\circ$ . Right: Mixed dataset using the datasets in Figure 4.24.



**Figure 4.26:** General variogram for mixed dataset. Left: General variogram and its model for the major and minor direction. Centre: Plan view of the dataset's samples. Right: Principal directions of the general variogram (blue for major and green for minor).



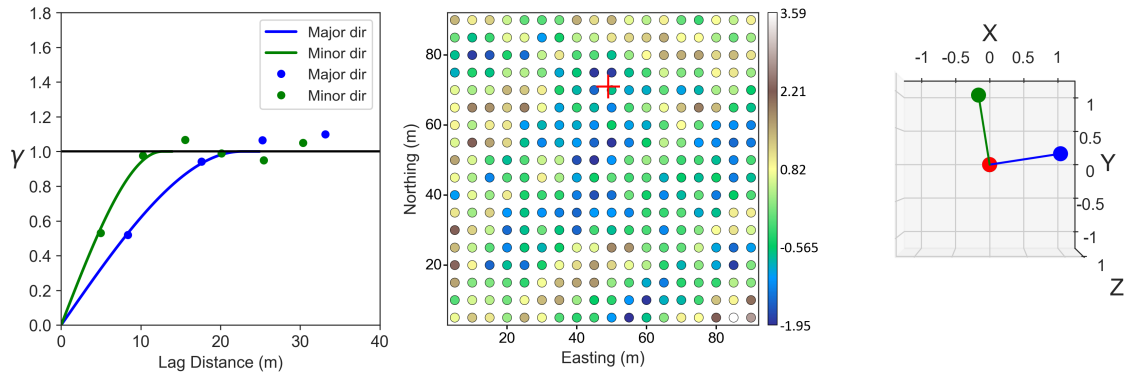
**Figure 4.27:** Geolocated variogram for a mixed dataset for the South anchor. Left: General variogram and its model for the major and minor direction. Centre: Plan view of the dataset's samples with the location of the anchor as a red cross. Right: Principal directions of the geolocated variogram.

directions are fairly fitted. The experimental points show some cyclicity.

The weighted variogram for the south anchor (see Figure 4.27) shows a direction of continuity of  $60^\circ$  ( $'60,0,0'$ ). This direction is fairly close to the original  $45^\circ$  dataset of the south, considering the added points of the  $90^\circ$  dataset.

The weighted variogram for the north anchor (see Figure 4.28) shows a principal direction of  $81^\circ$  ( $'81,0,0'$ ), which is reasonable considering that the dataset is generated by mixing two datasets.





**Figure 4.28:** Geolocated variogram for a mixed dataset for the North anchor. Left: General variogram and its model for the major and minor direction. Centre: Plan view of the dataset’s samples with the location of the anchor as a red cross. Right: Principal directions of the geolocated variogram.

Both geolocated variograms show clean experimental points and well fitted models, similar to the ones of the general variogram. In both of the weighted variograms, the experimental points in the major direction suggest the presence of a trend. The execution time of the program for this dataset and considering two anchors was 10.5 seconds.

This chapter summarizes the functionality of *autovar* for different datasets. It presents several examples of synthetic and real datasets, a comparison with *varcalc – varmodel*, and an example of using the geolocated option. The next chapter presents further applicability of the geolocated option to generate a mixed estimate with a novel methodology.

## Chapter 5

# Estimation with geolocated variograms

---

Estimation with ordinary kriging is a widely used technique in mining. This chapter presents a novel technique of mixed estimation using geolocated variograms. A variogram is essential for calculating weights in kriging, and *autovar* gives the option to get it automatically. Section 2.5 presents the theory for kriging. The mixed estimation with different variograms is recommended for interim estimates and visualization or trend model estimates.

Previous authors had shown different methodologies to consider the local variability in the estimation. Boisvert and Deutsch (2011) present kriging or simulation with local varying anisotropy (LVA). This technique first generates an LVA field (strike, dip, plunge, anisotropy ratios); second, it calculates the experimental variogram (GSLIB program prepared for LVA: *gamv\_LVA*); third, it selects the multidimensional scaling (MDS) parameters and finally perform the kriging or simulation (Boisvert & Deutsch, 2011). Martin et al. present a case study comparing the standard workflow and LVA estimation techniques (Martin, Machuca-Mory, Leuangthong, & Boisvert, 2019); they conclude that LVA techniques outperform stationarity methodologies.

Machuca (2010) presents a different approach to recognizing the local variability by considering anchor points that the user defines in the domain. A complete set of programs to develop the workflow are prepared (Machuca, 2010). In summary, first, it is necessary to calculate a matrix of distance weights. Second, build normal score transformation tables and model local transformation functions using Hermite polynomials. Third, calculate ‘location-dependent experimental variograms’ (Machuca, 2010). Fourth, perform a joint fit of the local variograms, and finally, perform kriging or simulation.

*Autovar* offers the option to declare  $n$  anchor locations, and there will  $n$  geolocated variograms, which can be used to generate different models. These models will be mixed at each node location based on the distance to each anchor. The kriging models are generated using the *kt3dn* program. The complete workflow to mix the models and get the final estimate can be set using a *jupyter* notebook.

### 5.1 Proposed methodology

The *autovar* program offers two options for variogram calculation: general and geolocated variograms. The general variogram is the standard way to get the experimental pairs, where all pairs have the same weight. The geolocated variogram is calculated after the local recognition of the principal directions of continuity. The geolocated variogram considers different weights for each

pair based on the distance to an anchor point. This local analysis allows to recognize the variability relative to the anchor. The user could define any number of anchors. For instance, a domain with four anchors will have four geolocated variograms and one general variogram.

If the user declares  $n$  anchors, the outputs will be  $n$  geolocated and 1 general variogram ( $n + 1$  variograms). Having  $n + 1$  variograms allows the estimation of the domain  $n + 1$  times. The parameters for each kriging estimation change related to the variogram characteristics (whether it is a general or geolocated variogram). Having  $n + 1$  block estimates allows to combine them to produce a unique mixed estimate that recognizes the local variability better.

The geolocated option is useful when the domain exhibits local variability. The practitioner decides how to group the data in domains at the beginning of the statistical analysis, and it looks to group data with similar spatial continuity. The decision of stationarity could lead to avoiding the use of the geolocated option. Nevertheless, suppose there is not enough data or the practitioner prefers to consider all available data for variogram calculations, allowing certain flexibility to recognize local variability. In that case, *autovar* is an option to consider.

This thesis proposes that the final estimated value at each node of the model is a weighted combination of the  $n + 1$  models at the respective node. The weights are standardized to add up to 1. This standardization ensures that kriging remains an exact estimator. Weighting many models allows different influences for the same data points, reflecting a better recognition of local variability. A detailed explanation of the geolocated option can be found in Section 3.5. The following workflow can be considered for the estimation using the geolocation option of *autovar* after the EDA and all prerequisites before the variogram calculation are reasonably met.

1. Run *autovar* with  $n$  number of anchors.
2. Estimate using all the  $n + 1$  generated models.
3. Calculate the weights at each node of the block model for the estimated value of the corresponding geolocated variogram. Gaussian weighting is recommended.
4. Standardize weights for all anchors at each node (see Equation 5.1).
5. Mix the different estimation models using the standardized weights (see Equation 5.2).
6. Perform a cross validation of the block model to ensure that there is no bias and evaluate the final estimate model comparing against a model that would be generated with a regular workflow.

$$w_{\gamma g} + \sum_{i=1}^n w_{\gamma i} = 1 \quad (5.1)$$

$$Z^* = w_{\gamma g} (Z_{\gamma g}^*) + \sum_{i=1}^n w_{\gamma i} (Z_{\gamma i}^*) \quad (5.2)$$

Where  $Z^*$  is the final mixed estimator at the current node (unsampled location),  $w_{\gamma g}$  is the selected constant weight for the general variogram,  $Z_{\gamma g}^*$  is the estimated value with the general variogram ( $\gamma g$ ),  $w_{\gamma i}$  is the weight for the estimated value with the geolocated variogram ( $\gamma i$ ),  $Z_{\gamma i}^*$  is the estimated using the corresponding geolocated variogram, and  $n$  is the number of anchors.

A weight for each node is calculated as a function of the distance from the node to the anchor of the geolocated variogram. A Gaussian weighting function is used because it ensures smooth transitions and no artifacts for the final mixed estimate. The weight for the general variogram ( $w_{\gamma g}$ ) is set as fixed for the whole domain to get an estimation with no artifacts. The general variogram weight ( $w_{\gamma g}$ ) ranges from 0 to 1. The total sum of weights for the geolocated estimations will be  $1 - wt_{gen}$  (see Equation 5.1). This ensures that kriging is exact at the data locations.

Sensitivity analysis with different datasets to evaluate the optimum weight for the general estimation (uses the general variogram) was performed. Weights from 0 to 1 (0.05 step) were tested several times to evaluate which weights would produce the lowest root mean squared error (RMSE). The results show that the value for the optimum weight is case dependent because each variogram can be different based on the available samples.

For instance, a more stable general variogram will have a more significant weight to produce a better estimate. In contrast, a more stable set of geolocated variograms could have more accumulated weight. Consequently, the general variogram will have a smaller weight. Generally, a bigger weight to the general variogram will produce a better estimate (less RMSE) if the general variogram is more stable than at least one of the geolocated variograms.

The execution time of a mixed estimator is increased according to the number of anchors plus the post processing (mixing  $n + 1$  estimated block models). Then, it is essential to pick a reasonable number of anchors that represent the local changes in the domain considering the subsequent estimations steps.

A regular workflow is expected to provide similar or better results if the domain is stationary. The estimation with *autovar*'s geolocated option is not recommended for stationary domains since the time to complete the task will be longer. The following section details a methodology considered regular for this thesis.

## 5.2 Regular methodology

There are several standard steps for estimation with kriging. This section explains the steps for a methodology considered regular in this thesis. This workflow assumes that the data was revised and prepared. This thesis uses GSLIB software to get the final block model: *varmap*, *varcalc*, *varmodel* and *kt3dn*. The workflow considers the following steps:

1. Decision of stationarity.

2. Identify principal directions (*varmap*).
3. Calculate the experimental variogram and models on the principal directions (*varcalc* and *varmodel*).
4. Estimate using ordinary kriging (*kt3dn*).
5. Model checks.

The GSLIB *kt3dn* program performs the estimation calculations, and the practitioner can pick different kriging types. The reader is referred to a previous CCG paper for a detailed description of the program (J. Deutsch, Barnet, & Deutsch, 2015). One of the kriging inputs is the variogram model (step 3). The practitioner needs to define the GSLIB angles to inform the principal directions for the kriging. These GSLIB angles are the same as the ones used for the variogram calculation. The search plan considers the variogram’s ranges as reference values.

The following section compares both methodologies with a synthetic 2-D dataset. It presents the calculated variograms, the final estimated block models, and a validation plot for both workflows.

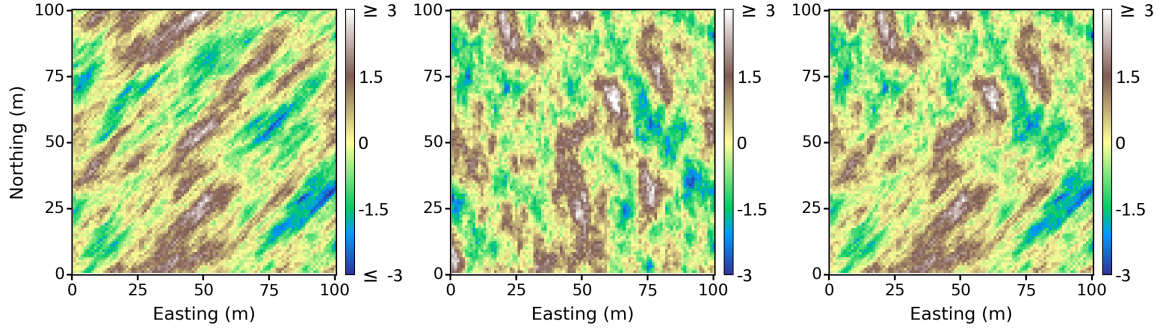
### 5.3 Comparison of estimation with autovar and regular workflow

This section shows a case study using workflows of Sections 5.1 and 5.2, where the main difference is the use of the geolocated variograms. The general variogram will produce results similar to those of the regular workflow. A comparison between the variograms of the *autovar* and *varcalc – varmodel* programs can be reviewed in Section 4.3.

A synthetic dataset was prepared to test both workflows. Two datasets were simulated with different principal directions and mixed with a weighting Gaussian distance function. Figure 5.1 presents the two initial and mixed datasets from the weighted combination of the first two. The initial datasets have the same anisotropy ratio between the major and minor directions (3:1). The final dataset has two different orientations: around 45° to the south and 0° to the north. An explanation of how to mix the datasets is in Section 4.4 (see Figure 4.25). Two anchors will be enough to characterize the local variability in this example. The mixed dataset is sampled regularly every 5 meters (in X and Y directions). There are 324 samples with a mean of 0.016 and 0.787 variance. Figure 5.2 shows the location plot of the samples.

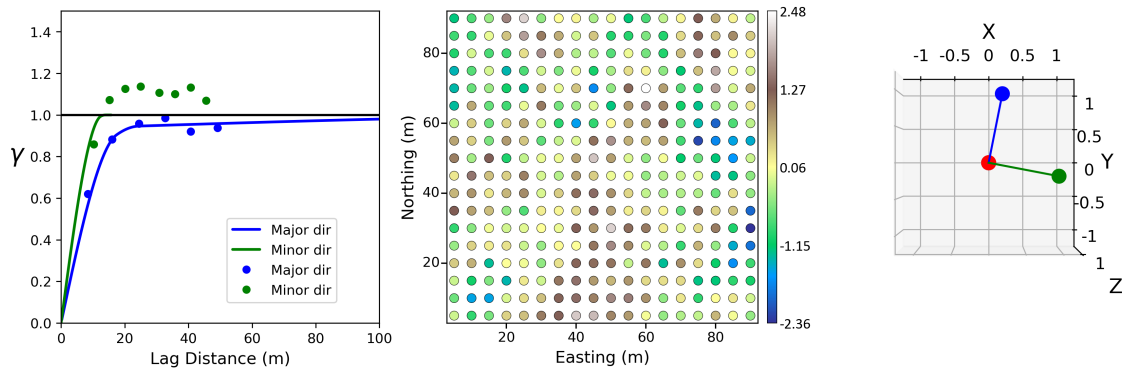
For the geolocated workflow, one anchor is located to the south, and the other one is located to the north (coordinates:  $x_1 = 27$ ,  $y_1 = 27$ , and  $x_2 = 27$ ,  $y_2 = 71$ ). Running *autovar* with the geolocated option selected will produce three variogram models as part of the outputs: one for each anchor plus the general variogram.

Figure 5.2 shows the general variogram where the principal direction is around 11°(‘11,0,0’). Figure 5.3 shows the weighted variogram for the anchor located in the south, where the principal

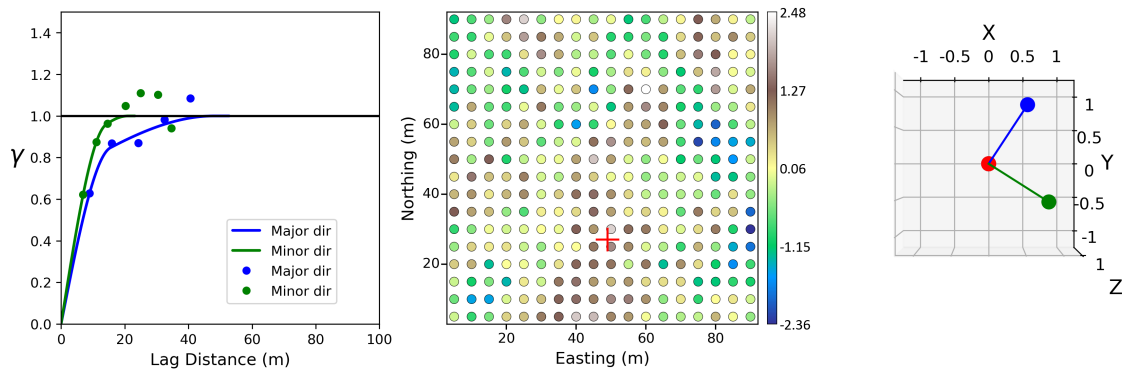


**Figure 5.1:** Left: First synthetic dataset with principal direction of  $45^\circ$  ('45,0,0'). Centre: Second synthetic dataset with principal direction of  $0^\circ$  ('0,0,0'). Right: Mixed dataset.

direction is around  $32^\circ$  ('32,0,0'). Figure 5.4 shows the weighted variogram for the anchor located in the north, where the principal direction is around  $6^\circ$  ('6,0,0').

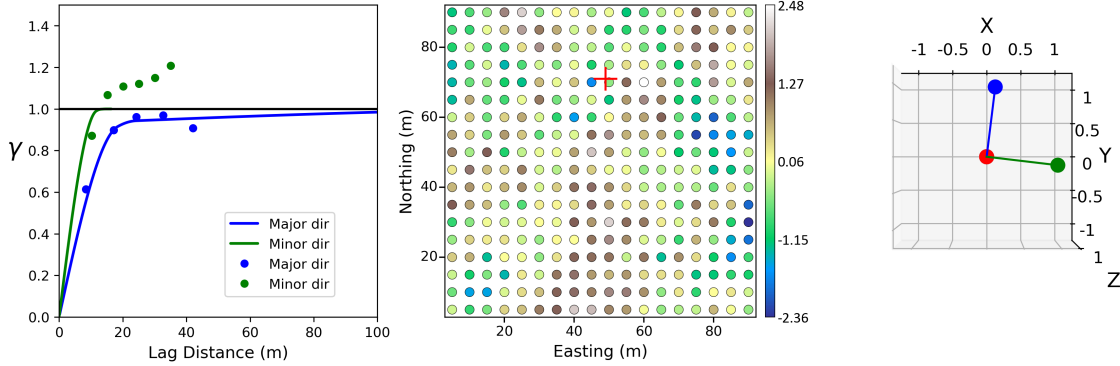


**Figure 5.2:** General variogram output for the mixed dataset. Left: variograms in the major and minor directions. Centre: location plot of samples. Right: Plot of the principal directions of continuity (blue for major and green for minor).



**Figure 5.3:** Variogram for anchor 1 (south). Left: variograms in the major and minor directions. Centre: Plan view of samples and anchor location (red cross). Right: Plot of the directions of continuity (blue for major and green for minor).

Kriging is calculated three times using each variogram output from *autovar* (see Figures 5.2, 5.3, and 5.4). The parameter file is set to use the respective inferred directions for the general and geolocated variograms. The variogram model text file has GSLIB format, which allows to integrate

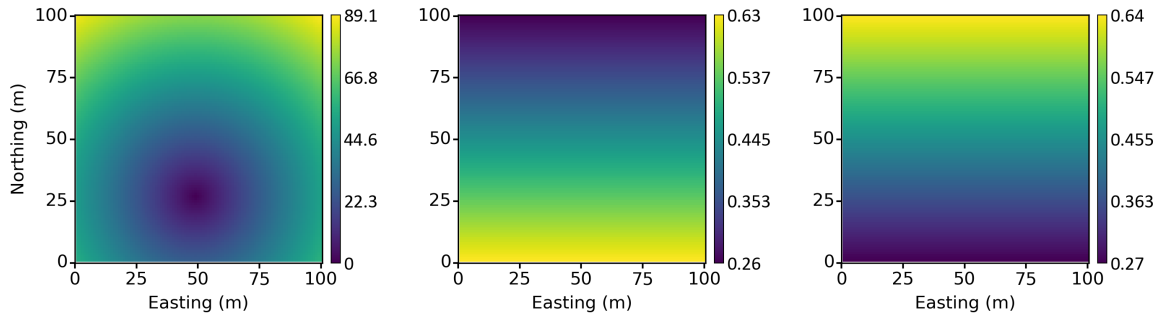


**Figure 5.4:** Variogram for anchor 2 (north). Left: variograms in the major and minor directions. Centre: samples and anchor location (red cross symbol). Right: Plot of the principal directions of continuity (blue for major and green for minor).

it easily in *kt3dn*.

For this specific example, the weight used for the general estimation is 0.1 ( $w_{\gamma g}$ ). Figure 5.5 shows the distances for anchor 1 and the weights for both anchors. These weights are standardized to add up to 0.9 ( $1 - w_{\gamma g} = 0.9$ ) at each node of the estimated model.

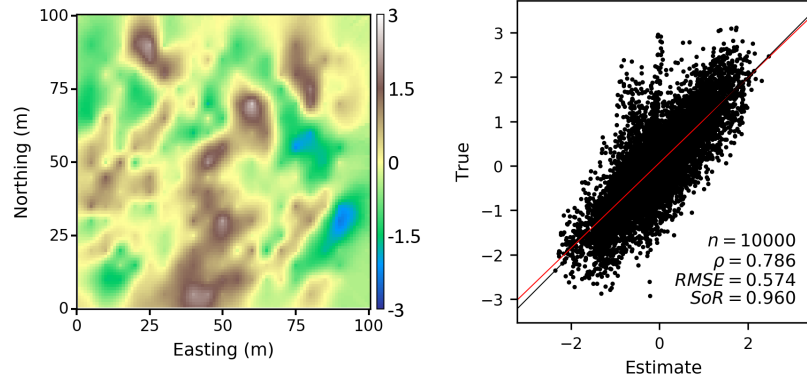
The three estimated models are mixed according to the standardized weights at each estimation node (see Equation 5.2). Figure 5.6 shows the final estimate considering the three models. The initial data was saved as true data to compare with the estimated values using a validation plot. The true and estimated data have the same grid with 10000 nodes.



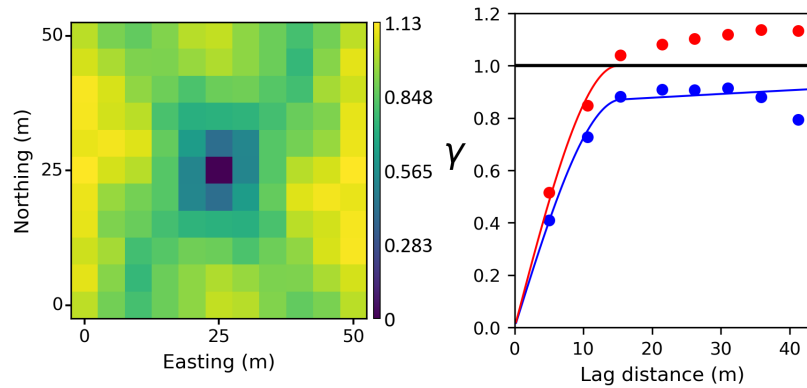
**Figure 5.5:** Plot of weights for the block estimation models using the geolocated variograms. Left: Plot of distances from anchor 1. Centre: Weights for estimates using geolocated variogram of anchor 1. Right: Weights for estimates using geolocated variogram of anchor 2.

For the regular estimation workflow (see Section 5.2), the same 324 samples are used. Figure 5.7 shows the variogram map and the experimental variogram and its model. The variogram map suggests a direction similar to the inferred directions of the general variogram ('15,0,0'). These angles are set for the variogram calculation and modeling to keep a fair comparison with *autovar*. The variograms are stable and consistent for both directions. Figure 5.8 shows the kriging results using the variograms of the regular workflow (Figure 5.7).

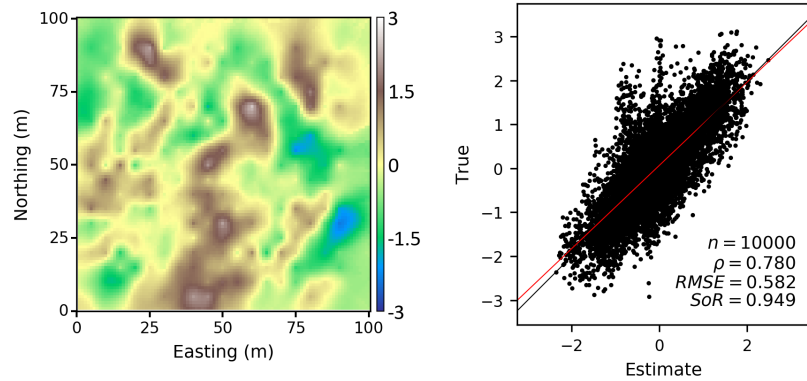
Both workflows have similar kriging search parameters; the main difference is the variogram models. Figures 5.6 and 5.8 show the kriging results and a validation plot, which compares the



**Figure 5.6:** Mixed estimation results using two anchors. Left: Final estimation model using *autovar*. Right: A validation plot comparing the exhaustive and estimated data using *autovar*.



**Figure 5.7:** Variogram map and experimental variograms points and their models for the regular workflow.



**Figure 5.8:** Estimation results using regular workflow. Left: estimation using the variogram model of Figure 5.7. Right: A validation plot comparing the exhaustive and estimated data using *autovar*.

estimated values to the true data. The results using *autovar* show a decrease in the RMSE of 1.37%.

This section compares the results of the regular workflow and *autovar* using two anchors. There is a slight improvement using the anchor points; however, another point to evaluate is the optimal number of anchors. The next section presents a sensitivity analysis concerning the number of anchors.

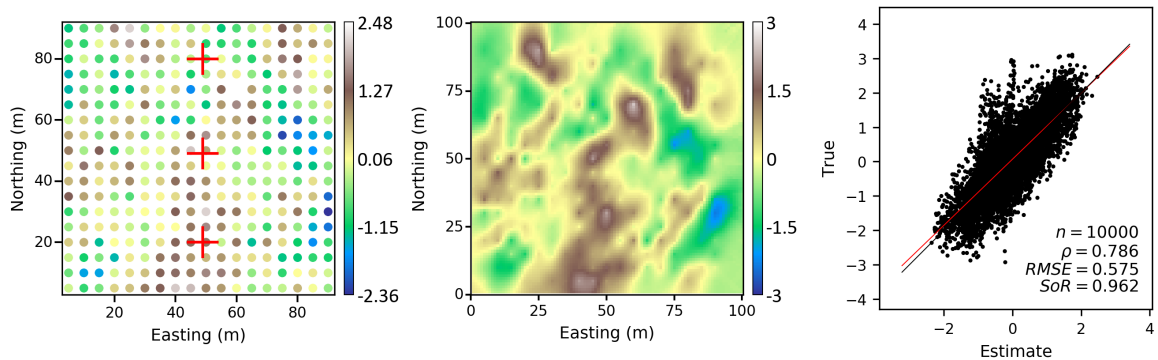


## 5.4 Sensitivity with respect to number of anchors

Another important factor is to select the number of anchors that will produce the best estimate possible. The practitioner could consider identifying zones with different directions of continuity and assign anchors there. Then, the suggestion is to increase the number of anchors until the variograms become too similar or unstable.

For instance, in the previous example (see Figure 5.6), two anchors were set; then, increase the number of anchors to three and compare the results. If the variograms are stable, the number of anchors could be increased or located in different positions until they become unstable or similar. Lets say that for the third anchor, the variograms are no longer representative, and the best option will be to go the previous number of anchors with stable variograms, 2 in this case.

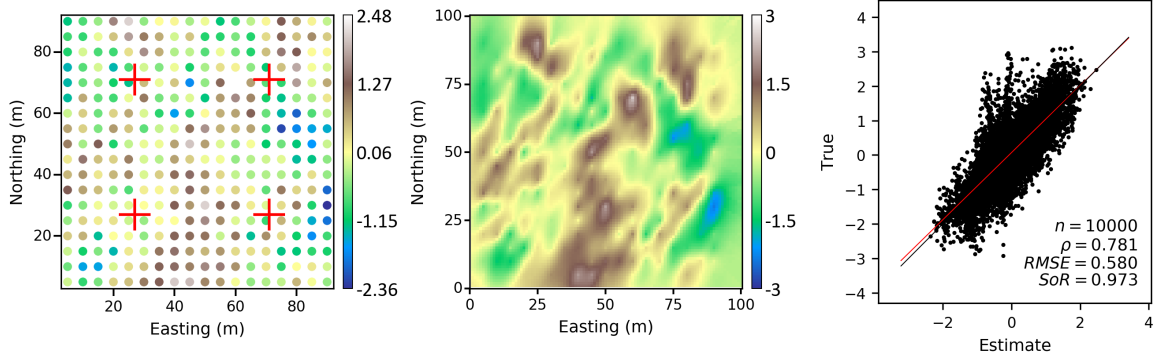
Close anchors will have similar directions of continuity and similar variograms. Even if these geolocated variograms are stable, it is better to reduce the number of anchors because the time for estimation will be increased based on the number of anchors. A 2-D example presents a mixed estimation using the same synthetic dataset as in the previous example. All the steps to mix the estimates are explained in Section 5.1.



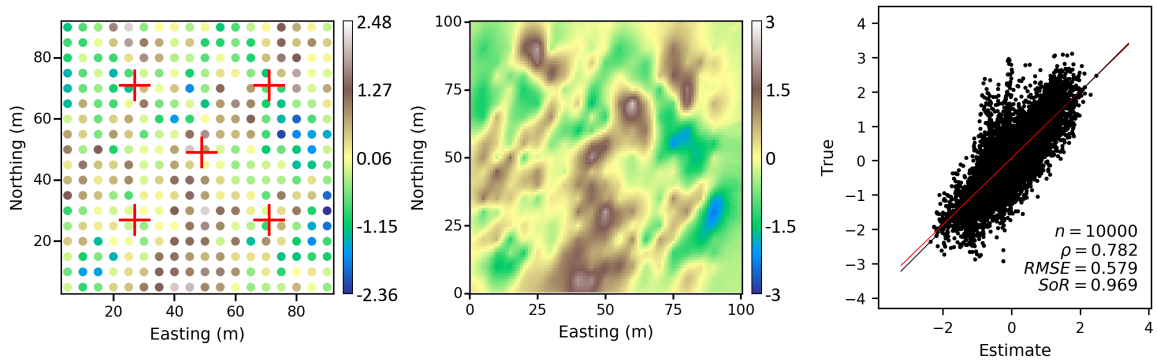
**Figure 5.9:** Mixed estimation results considering three anchors. Left: samples and anchors location. Centre: Final mixed estimation using the general and three geolocated variograms. Right: A validation plot comparing the exhaustive and estimated data using *autovar* considering three anchors.

Figure 5.9 shows the samples and position of the three anchors and the final estimate. The RMSE (0.575) of the mixed estimates using three anchors is slightly larger than the mixed estimation with two anchors, although the variograms are stable. Figure 5.10 shows the samples and positions of 4 anchors. The RMSE of the mixed estimates using four anchors has increased to 0.580. Figure 5.11 shows the results when five anchors are used; the RMSE has also increased to 0.579. The best option is to work with two anchors.

For this synthetic example, it is possible to do cross validation with the exhaustive true. For real examples, the practitioner can check with cross validation and visual inspection of the geolocated variograms, evaluating the stability and similitude between variograms. This section presented a case study as an example. Nonetheless, more tests were done to compare the results using both



**Figure 5.10:** Mixed estimation results considering four anchors. Left: samples and anchors location. Centre: Final mixed estimation using the general and four geolocated variograms. Right: A validation plot comparing the exhaustive data and the estimated data using *autovar* considering four anchors.



**Figure 5.11:** Mixed estimation results considering five anchors. Left: samples and anchors location. Centre: Final mixed estimation using the general and five geolocated variograms. Right: A validation plot comparing the exhaustive and estimated data using *autovar* considering five anchors.

methodologies. The next section summarizes their results.

## 5.5 Statistics of study cases

This section presents the statistics of the results for the mixed estimation for synthetic and exhaustive images datasets (Mokdad, Binakaj, & Boisvert, 2022). Several synthetic datasets were generated to test the geolocated option of the program for two anchor locations. Selected images from the data validation project (Mokdad, Binakaj, & Boisvert, 2022) were used to analyze the performance of this methodology. All examples are 2-D, and the position of the anchors is the same as the example in Section 5.3 (coordinates:  $x_1 = 27$ ,  $y_1 = 27$ , and  $x_2 = 27$ ,  $y_2 = 71$ ).

The methodology to create the synthetic mixed datasets is the same as the one explained in Section 5.3. The dataset in the south will have a fixed direction at  $0^\circ$  degrees, and the north dataset has variable directions every 10 degrees, from 0 to 180. This configuration generates 17 principal directions (e.g. ‘0,0,0’, ‘10,0,0’, ... ‘170,0,0’).

Different random seeds were used to generate three different sets of mixed data, generating 51 2-D datasets. This dataset is saved as the true dataset. Each synthetic dataset is sampled at a

regular space (2, 4, 6, 8, and 10 meters spacing in X and Y directions). Consequently, there will be five sets of samples per image and 255 datasets.

All the images from the data validation project can be customized to have a desired number of nodes. For this case study, all the real images were set to have the same grid definition as in the previous synthetic datasets. There are  $10^4$  nodes per image that are considered the true data. The sampling is regular and has the same spacing as in the synthetic datasets. There are 48 images and five subsets of samples per image, giving a total of 240 datasets.

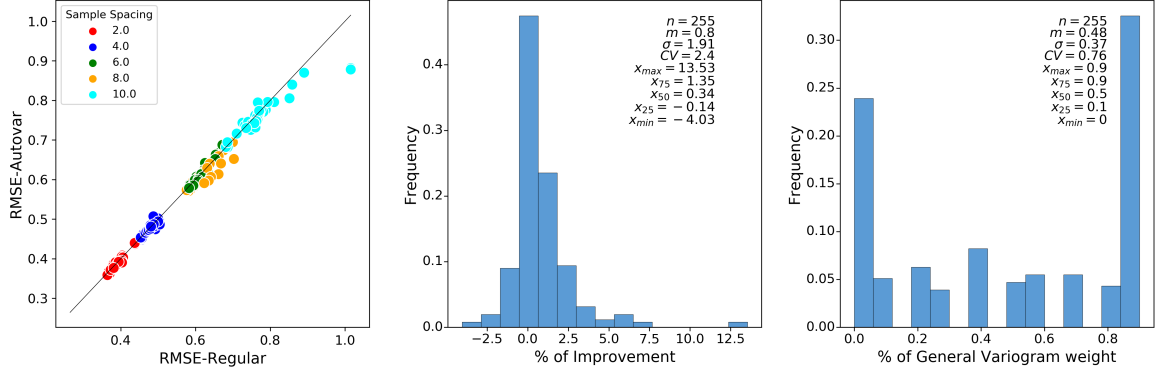
For each set of samples (synthetic or not), the *autovar* program calculates the general and geolocated variograms. The program was configured for disseminated deposits because the selected datasets mainly exhibit this deposit type characteristics. Each general and geolocated variogram model is used in the *kt3dn* program. Consequently, there will be three estimations. These estimations will be mixed, selecting weights that reduce the RMSE. This selection can be done since the exhaustive data (true data) is known. Then, each estimation is mixed with its respective weight at each node location, generating the final mixed estimate.

The regular workflow generates a variogram map that has been compared to the results of *autovar*, and the identification of directions is very similar. The same directions calculated from the *autovar*'s general variogram are set as the directions for *varcalc* to keep a fair comparison, and the semi-automatic fitting option is also picked for *varmodel* in the same directions. The lag spacing is set to be the same as the sample distance, and the lag tolerance is half of this value; the horizontal bandwidth is set to the same size as the lag distance, and the azimuth tolerance is set to 22.5 in all cases. These parameters are the most suitable and possible to automate since the sample spacing is regular. The selected azimuth tolerance show to give the most stable results for the tests. The kriging parameters are set from the results of the *varmodel*, and the search angles are also the same as those used for the variogram calculation and modeling.

The RMSE is calculated by comparing the estimates against the exhaustive truth. When less data is used, there will be more RMSE for the synthetic datasets (see Figure 5.12). The same is not observed for the exhaustive images datasets. This test allows the evaluation of different datasets, sample spacings, and the statistics of the results from both workflows.

**Table 5.1:** Summary results for synthetic datasets with *autovar*.

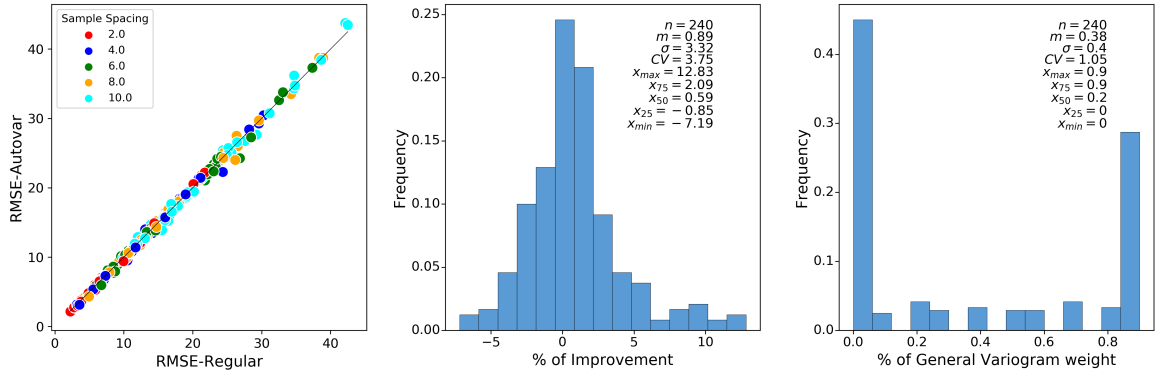
	RMSE regular	RMSE mixed	$w_{\gamma g}$	Improvement (%)	Succes (%)
count	255	255	255	255	255
mean	0.58	0.57	0.48	0.8	0.7
std	0.14	0.14	0.37	1.91	0.46
min	0.36	0.36	0	-4.03	0
25%	0.46	0.45	0.1	-0.14	0
50%	0.59	0.59	0.5	0.34	1
75%	0.66	0.65	0.9	1.35	1
max	1.02	0.88	0.9	13.53	1



**Figure 5.12:** Summary results of mixed estimation for synthetic datasets using two anchors. Left: Scatter plot comparing RMSE from the mixed estimation and the regular workflow. Centre: Histogram of improvement for each case. Right: Histogram of optimal weight for the general variogram.

**Table 5.2:** Summary results for exhaustive images datasets with *autovar*.

	RMSE regular	RMSE mixed	$w_{\gamma g}$	Improvement (%)	Succes (%)
count	240	240	240	240	240
mean	15.45	15.36	0.38	0.89	0.62
std	8.26	8.3	0.4	3.33	0.49
min	2.26	2.16	0	-7.19	0
25%	9.38	9.43	0	-0.85	0
50%	14.03	13.95	0.2	0.58	1
75%	19.94	19.55	0.9	2.09	1
max	42.51	43.75	0.9	12.83	1



**Figure 5.13:** Summary results of mixed estimation for exhaustive images datasets using two anchors. Left: Scatter plot comparing RMSE from the mixed estimation and the regular workflow. Centre: Histogram of improvement for each case. Right: Histogram of optimal weight for the general variogram.

Table 5.1 shows the summary of these results for the synthetic datasets. 70% of the time, the autovar workflow is performing better, and the mean of improvement of the RMSE is 0.8%. Table 5.2 shows the summary of results for the exhaustive images datasets. 62% of the time, the autovar workflow is performing better, and the mean of improvement of the RMSE is 0.9%.

Figure 5.12 shows the summary results for the synthetic datasets. The RMSE increases when the number of samples decreases for the synthetic datasets. Figure 5.13 shows the summary results for the exhaustive images datasets. In general, the errors are slightly larger for the regular workflow.

There are cases where *autovar* does not improve the final results. This is produced when the variograms are not stable or the geolocated variograms fail to recognize local characteristics on the domain. This is related to the dataset characteristic and the selected samples as a matter of chance.

The right part of each figure (see Figures 5.12 and 5.13) presents a histogram of the optimal weights. There is no clear distribution; however, the mean of the optimal weight for the general variogram is around 0.4.

This chapter shows the methodology for mixed estimates using the general and geolocated variograms. It also compares the mixed and regular methodology and presents a detailed case study. This section summarizes several tests for synthetic and exhaustive images datasets to test the effectiveness of the mixed estimates. The following section presents the conclusions and future work of this thesis.

## Chapter 6

# Conclusions and Future work

---

The characterization of the spatial variability of a regionalized variable is an essential step for resource estimation. The variogram is a widely used tool in geostatistics to characterize the continuity of the regionalized variable. A significant amount of time is dedicated to having a variogram model that reasonably represents the geological characteristics of the variable in the domain. Kriging stands as one of the most used techniques in mining. This thesis presented several contributions to optimize these processes.

*Autovar* allows to calculate experimental variogram points and model them automatically considering parameters inferred directly from the data and that considers common characteristics based on the deposit type. It provides a initial variogram that can be tuned. The mixed estimation technique generally characterizes better domains with local scale variability because it considers the recognition of this variability through the geolocated variograms.

This section summarizes the contributions of this research. This thesis presents the implementation of the *autovar* program (see Chapter 3), which integrates all steps to produce the spatial variability characterization using variograms. This program also gives the geolocated option (see Chapter 3 and 4), which helps to identify local variability representative of anchor locations. This thesis also compares the *varcalc – varmodel* and the *autovar* workflow (see Chapter 4).

A novel methodology for estimation is detailed as an alternative to a regular estimation (see Chapter 5). The proposed mixed estimation technique is compared to the regular methodology for estimation with ordinary kriging.

The practitioner is encouraged to ensure the quality of the data and correct selection of the variable, including but not limited to outlier management, normal score transformation, declustering, and stratigraphic transformation.

## 6.1 Contributions

The *autovar* program provides a reasonable experimental variogram and its model. It provides an initial model that could save time and be executed many times in different datasets, considering one variable at a time. It considers the most common conventions in variogram calculation (see Section 3.3) and modeling (see Section 3.4). *Autovar* enhances automation through parameter inference inside the program. The program looks to cover a variety of cases; nevertheless, a final evaluation by the practitioner is always the final test. Considering the geologic characteristics of the deposits will allow a better evaluation of the program's calculations.

This thesis presents the workflow and the logic of the algorithms inside the main program (see Chapter 3). The program infers all the parameters directly from the data for disseminated and tabular deposits. It starts with consistency checks and proceeds to infer the principal directions by eigendecomposition of the inertia tensor and subsequent ordered of vectors to get a rotation matrix. The inferred rotation matrix replaces the need to prompt the user for the principal directions. All the data points are rotated according to the principal directions, and the next stage is to calculate the experimental variogram. The program implements different search parameters for getting pairs for each lag based on the deposit type. The experimental points are evaluated to select the ones that would be modeled using an optimization by minimization algorithm. The outputs are the experimental variogram points, the variogram model, and the eigenvectors showing the principal directions of continuity.

This thesis also presents a geolocated option (see Sections 3.5 and 4.4) to identify local variability in a dataset without subdividing a domain. The program offers the geolocated option, where the user can define anchors in areas of local variability inside the domain. The directions are recognized centred at each anchor, and the geolocated variograms are calculated and modeled in these directions. Each pair of points will have a weighted contribution to calculating the experimental variogram points. This weight is based on the distance from the points to the anchor. The outputs for the geolocated option include the same outputs as the general variogram plus a file that contains the weights for the points with respect to each anchor.

This thesis compares the results of the *autovar*'s general variogram and the variograms produced by *varcalc* and *varmodel* (see Section 4.3). The selected examples show the similitude of results using both workflows, which supports the applicability of the *autovar* program. Both workflows have advantages that can be used according to the practitioner's goals. The flexibility to pick parameters with *varcalc* and *varmodel* shows the high customization to have the best possible variogram; however, it requires some expertise to tune it correctly. Another advantage of *varcalc* is its numerous options to measure spatial continuity, such as variogram, covariance, correlogram, and indicator variogram (continuous or categorical variables). In contrast, *autovar* results are only for the variogram. The main advantage of *autovar* is the automation and speed of calculating a variogram. Another advantage is the geolocated option and the few parameters required, which makes it suitable for new practitioners.

A novel methodology of mixed estimation with ordinary kriging is presented in this thesis (see Chapter 5). This methodology presents a weighted combination of different estimation block models. The block models are produced using the geolocated variogram models. This technique better recognizes local characteristics because it assigns different weights for each model at each node based on the distance from the node to the anchor point.

A case study shows a decrease in the RMSE comparing the estimated model using the mixed estimation technique and the regular methodology (see Section 5.3). A summary of a series of tests

is shown for synthetic and exhaustive datasets that exhibit local variability (see Section 5.5). The average improvement of the mixed estimation is 0.9%. Nevertheless, there are cases where there is no improvement. To get a good mixed estimate, the user must pick the most appropriate weight for the general variogram. The most stable set of variograms needs to be picked based on comparing the goodness of the geolocated variograms against the general variogram. This goodness depends on the appropriate selection of the directions, the structure of the variograms, and if there is recognizable anisotropy. A valuable insight is the knowledge of the deposit's geology.

## 6.2 Limitations

This thesis tested the presented methodologies many times to assess their potential and limitations. Automatic variogram calculation does not provide flexibility to manually tune parameters. This lack of flexibility is an issue when the dataset has few samples, or the spatial variability is too high. According to the data characteristics, the program could identify the directions incorrectly, leading to inconsistent or inappropriate results. A final evaluation of the accuracy and precision of the variogram model is recommended.

This thesis implements a workflow for continuous variables and one variable at a time, which could be cumbersome in a multivariate case. The algorithm is implemented to give only the variogram as output. There are no options for other kinds of measures of spatial variability such as covariance, correlogram or indicator variograms. The spherical model is the only model that fits the experimental variogram points. The program does not allow to specify a nugget effect and sets it with a constant value of zero.

The optimal weight for the general variogram has not been generalized for the mixed estimation workflow because the optimal weight is case dependent. This limitation translates the responsibility to the user to get the best mixed estimate and could be challenging to determine. There are cases where the mixed estimation does not improve the final results, and it is related to the dataset's characteristics.

## 6.3 Future work

The following steps will be to enhance the program for a larger number of deposits. The future *autovar* could consider a set of optional, more flexible parameters that experienced practitioners could customize. This set of parameters could include a ratio for defining searches and directions. Another essential feature to add is the flexibility to inform or infer the nugget effect for each case.

The uncertainty of the variogram could be evaluated using bootstrap, where the program could give some metrics to characterize the final model and provide numerical feedback to the practitioner. The number and location of anchors could be automatically inferred inside the subroutines, aiming to get an optimal number of anchors in areas of local variability.



Finally, the mixed estimation workflow requires using the *autovar* and the *kt3dn* program, plus the post processing, getting the final mixed block model. This workflow could be implemented in a unique program that could go from the data to the final estimates. This idea could also be implemented for simulation, including the calculation of a normal score variogram. The program could give several outputs as debug files for the practitioner. These debug files will help identify whether the final results are reasonable.

# References

---

- Bazania, J., & Boisvert, J. (2023). Analysis of classification workflows and industry practice. *CCG Paper 2023-134, Centre for Computational Geostatistics, University of Alberta, Canada, 25.*
- Blundell, M., & Damian, H. (2015). *Kinematics and dynamics of rigid bodies*. Elsevier.
- Boisvert, J. B., & Deutsch, C. V. (2011). Programs for kriging and sequential gaussian simulation with locally varying anisotropy using non-euclidean distances. *Computers & Geosciences, 37*(4), 495–510.
- Brailov, A. Y. (2016). Formation of the Complex Drawing. Octants. The Method of Gaspard Monge. *Engineering Graphics: Theoretical Foundations of Engineering Geometry for Design*, 15–17.
- Chilès, J.-P., & Delfiner, P. (2012). *Geostatistics: modeling spatial uncertainty* (Vol. 713). John Wiley & Sons.
- Chilès, J.-P., & Desassis, N. (2018). Fifty years of kriging. *Handbook of mathematical geosciences: Fifty years of IAMG*, 589–612.
- Clark, I., & Harper, W. (2020). *Teaching datasets*. Retrieved from [https://www.dropbox.com/sh/2lhtrqo4sy0mefo/AACvIaW1Xb\\_qqMc4i26xxJfBa/Teaching%20data%20sets?dl=0&subfolder\\_nav\\_tracking=1](https://www.dropbox.com/sh/2lhtrqo4sy0mefo/AACvIaW1Xb_qqMc4i26xxJfBa/Teaching%20data%20sets?dl=0&subfolder_nav_tracking=1)
- Darling, P. (2011). *4.5.8.4.3 deposit geometry*. Society for Mining, Metallurgy, and Exploration (SME). Retrieved from <https://app.knovel.com/hotlink/khtml/id:kt008JZ4A1/sme-mining-engineering/deposit-geometry>
- de Carvalho, P. R. M., & da Costa, J. F. C. L. (2021). Automatic variogram model fitting of a variogram map based on the Fourier integral method. *Computers & Geosciences, 156*, 104891.
- Deutsch, C. (2021). *Citation in Applied Geostatistics*.
- Deutsch, C. V., & Journel, A. G. (1997). *GSLIB Geostatistical software library and user's guide* (second ed., Vol. 369). New York: Oxford University Press.
- Deutsch, J. (2015a). Experimental variogram tolerance parameters. *Geostatistics Lessons*. Retrieved from <https://geostatisticslessons.com/lessons/variogramparameters>.
- Deutsch, J. (2015b). Variogram calculation for tabular deposits. *Geostatistics Lessons*. Retrieved from <https://geostatisticslessons.com/lessons/tabularvariograms>.
- Deutsch, J. (2015c). Variogram program refresh. *CCG Paper 2015-410, Centre for Computational Geostatistics, University of Alberta, Canada, 17.*
- Deutsch, J., Barnet, R., & Deutsch, C. (2015). Latest kriging program. *CCG Paper 2015-402, Centre for Computational Geostatistics, University of Alberta, Canada, 17.*
- Deutsch, J., & Deutsch, C. (2015). Introduction to choosing a kriging plan. *Geostatistics Lessons*. Retrieved from <https://geostatisticslessons.com/lessons/introkrigingplan>.
- Deutsch, M. (2015). The Angle Specification for GSLIB Software. *Geostatistics Lessons*. Retrieved

- from <https://geostatisticslessons.com/lessons/anglespecification>.
- Dubois, G. (1998). Spatial interpolation comparison 97: foreword and introduction. *Journal of Geographic Information and Decision Analysis*, 2(2), 1–10.
- Ellis, R., King, P., Richardson, S., & Allen, A. (2023). NI 43-101 Technical Report on the Neves-Corvo Mine, Portugal.
- Emery, X. (2010). Iterative algorithms for fitting a linear model of coregionalization. *Computers & Geosciences*, 36(9), 1150–1160.
- Glencore. (2022). GLENCORE Resources & Reserves, as at 31 December 2022.
- Graden, R. (2022). NI 43-101 Technical Report, Teck Highland Valley Copper.
- Gray, D., Sulway, R., Cameron, A., & Stone, R. (2022). Ravensthorpe Nickel Operations, Western Australia, NI 43-101 Technical Report.
- Greiner, W. (2010). *Classical mechanics: systems of particles and hamiltonian dynamics*. Springer.
- Gringarten, E., & Deutsch, C. V. (2001). Teacher’s aide variogram interpretation and modeling. *Mathematical Geology*.
- Hadavand, Z., & Deutsch, C. (2022). Derivation of Locally Varying Angles from Mass Moment of Inertia Tensor. *CCG Paper 2022-148, Centre for Computational Geostatistics, University of Alberta, Canada, 24*.
- Isaaks, E., & Srivastava, R. (1989). *An introduction to applied geostatistics*. Oxford University Press.
- Jo, H., & Pyrcz, M. J. (2022). Automatic semivariogram modeling by convolutional neural network. *Mathematical Geosciences*, 54(1), 177–205.
- Journel, A. (2018). Roadblocks to the evaluation of ore reserves—the simulation overpass and putting more geology into numerical models of deposits. *Advances in applied strategic mine planning*, 47–55.
- Journel, A. G., & Huijbregts, C. J. (1978). *Mining geostatistics*. Academic press.
- Kim, J., & Deutsch, C. (2022). Calculation and modeling of variogram anisotropy. *Geostatistics Lessons*. Retrieved from <https://geostatisticslessons.com/lessons/variogramanisotropy>.
- Korvin, G. (1982). Axiomatic characterization of the general mixture rule. *Geoexploration*, 19(4), 267-276. Retrieved from <https://www.sciencedirect.com/science/article/pii/001671428290031X> doi: [https://doi.org/10.1016/0016-7142\(82\)90031-X](https://doi.org/10.1016/0016-7142(82)90031-X)
- Krolak, T., Palmer, K., Lacouture, B., & Paley, N. (2017). NI 43-101 Technical Report, Red Dog Mine, Alaska, USA.
- Kumara, P., & Deutsch, C. (2017). Transforming a variogram of normal scores to original units. *Geostatistics Lessons*. Retrieved from <https://geostatisticslessons.com/lessons/convertnsvariograms>.
- Larrondo, P. F., Neufeld, C. T., & Deutsch, C. V. (2003). Varfit: A program for semiautomatic variogram modeling. *CCG Paper 2003-122, Centre for Computational Geostatistics, University*

- of Alberta, Canada, 05.
- Latifi, A. M., & Boisvert, J. B. (2022). Stratigraphic coordinate transformation. *Geostatistics Lessons*. Retrieved from: <http://www.geostatisticslessons.com/lessons/stratcoords>.
- Lehouiller, P., Lampron, S., Gagnon, G., Houle, N., & Bouchar, F. (2020). NI 43-101 Technical Report, Canadian Malartic Mine, Quebec, Canada.
- Li, Z., Zhang, X., Clarke, K. C., Liu, G., & Zhu, R. (2018). An automatic variogram modeling method with high reliability fitness and estimates. *Computers & Geosciences*, 120, 48–59.
- Lozada, L., & Espinoza, J. (2011). Technical Report Mineral Reserves and Resources Antamina Deposit, Perú 2010.
- Machuca, D. (2010). *Geostatistics with Location-Dependent Statistics*. Department of Civil and Environmental Engineering, University of Alberta, Canada.
- Mardia, K. V., & Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 71(1), 135–146.
- Martin, R., Machuca-Mory, D., Leuangthong, O., & Boisvert, J. B. (2019). Non-stationary geostatistical modeling: a case study comparing Iva estimation frameworks. *Natural Resources Research*, 28(2), 291–307.
- Matheron, G. (2019). *Matheron's theory of regionalised variables*. Oxford University Press.
- Mokdad, K., Binakaj, D., & Boisvert, J. (2022). Data Validation Project: Validation of 114 Spatial 2D Datasets (Non-synthetic Data). *CCG Paper 2022-110*, Centre for Computational Geostatistics, University of Alberta, Canada, 24.
- Mokdad, K., Koushavand, B., & Boisvert, J. (2022). Update on Automatic Variogram Inference Using CNN. *CCG Paper 2022-108*, Centre for Computational Geostatistics, University of Alberta, Canada, 24.
- Montero, J.-M., Fernández-Avilés, G., & Mateu, J. (2015). *Spatial and spatio-temporal geostatistical modeling and kriging*. John Wiley & Sons.
- Owusu, K., Solomon Dagdelen. (2019). Critical review of mineral resource classification techniques in the gold mining industry. *Mining Goes Digital – Mueller et al. (Eds)*.
- Pardo-Igúzquiza, E. (1999). *Varfit: a fortran-77 program for fitting variogram models by weighted least squares* (Vol. 25) (No. 3). Elsevier.
- Pardo-Igúzquiza, E., Mardia, K. V., & Chica-Olmo, M. (2009). Mlmatern: A computer program for maximum likelihood inference with the spatial matern covariance model. *Computers & Geosciences*, 35(6), 1139–1150.
- Pyrzcz, M. J., & Deutsch, C. V. (2014). *Geostatistical reservoir modeling*. Oxford University Press.
- Rossi, M. E., & Deutsch, C. V. (2013). *Mineral resource estimation*. Springer Science & Business Media.
- Samson, M., & Deutsch, C. (2021). Calculation and modeling of variogram anisotropy. *Geostatistics Lessons*. Retrieved from <https://geostatisticslessons.com/lessons/sillofvariogram>.

- Schalekamp, E., & McAllister, A. (2022). NI 43-101 Technical Report on Elkview Coal Operation, British Columbia, Canada.
- Seward, A., Seeger, C., Nahirniak, T., Royer, P., & Beranek, B. (2022). NI 43-101 Technical Report on Greenhills Coal Operation, British Columbia, Canada.
- Silva, D., & Boisvert, J. (2013). Mineral Resource Classification (NI 43-101): An Overview and a New Evaluation Technique. *CCG Paper 2013-307, Centre for Computational Geostatistics, University of Alberta, Canada, 15.*
- Slabaugh, G. G. (1999). Computing euler angles from a rotation matrix. *Retrieved from <http://eecs.qmul.ac.uk/gslabaugh/publications/euler.pdf>.*
- Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley Pub. Co.
- Wilde, B., & Deutsch, C. (2012). Automatic variogram modeling from censored variogram volumes. *CCG Paper 2012-127, Centre for Computational Geostatistics, University of Alberta, Canada, 14.*
- Winston, W., & Goldberg, J. (2004). *Operations research: Applications and algorithms*. Thomson Brooks/Cole.