

Geostatistics in the Presence of Extreme Values with Network Models of Regionalization

by

Benjamin Edward Harding

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

Doctor of Philosophy

in

Mining Engineering

Department of Civil and Environmental Engineering
University of Alberta

© Benjamin Edward Harding, 2024

Abstract

Geostatistical models are often generated with widely spaced data configurations. Data collection costs prohibit exhaustive sampling and necessitate statistical inference from limited samples. Spatial prediction with sparse data in the presence of extreme values is an enduring challenge in the mining industry. Extreme values may have significant local influence, leading to overstated resources and the risk of production shortfalls. Practitioners are presented with difficult decisions for restricting extreme value influence and characterizing their spatial continuity. Inputs to numerical geologic models consist of observed data, a representative histogram and spatial controls on mineralization. Each of these components presents challenges in the presence of extreme values. Extreme values are rare events, making inferences about their probability of occurrence difficult. The influence of extreme values is often controlled in practice through grade capping, which could significantly impact the final resource. Extreme values' spatial continuity often differs from the barren or lower grade background. Traditional estimation and simulation methodologies are limited in adapting to extreme values and asymmetric spatial continuity features. These challenges motivate the development of a framework for the simulation of continuous variables with explicit consideration of high-order extreme value features. The proposed network model of regionalization (NMR) framework constructs a continuous regionalized variable as a non-linear mixture of latent Gaussian factors and does not require capping or modification of extreme grade values. The network parameters are inferred via optimization, considering two- and multi-point connectivity features at grade thresholds. This permits the reproduction of high-order connectivity features and asymmetric spatial continuity of high and low grades that cannot be captured by a single Gaussian random function (RF) model. The latent Gaussian factors are imputed such that they exactly reproduce the observed data values when mixed. The applicability of the proposed methodology is demonstrated on a mineral deposit where the project operators note non-Gaussian, extreme value features in drillhole data. In this deposit the NMR approach shows a 7% improvement in expected metal relative to traditional approaches using a hold-out data set for validation.

Preface

This thesis is an original work by B.E. Harding. No part of this thesis has been previously published.

Dedication

To my wife J.E.R

Acknowledgments

I would like to thank my supervisor Clayton Deutsch for his continued support throughout this process. Your ability to intrinsically motivate your students is unparalleled. I am grateful to the Centre for Computational Geostatistics for financial support. I would also like to thank Art Frye for the initial opportunity to travel this path.

Table of Contents

1	Introduction	1
1.1	Problem Setting	1
1.1.1	Thesis Statement	2
1.2	Geostatistical Background	4
1.2.1	Overview	4
1.2.2	Factorial Kriging	5
1.2.3	Simulating Continuous Variables	7
1.2.4	Multiple Point Statistics	8
1.2.5	High-Order Simulation	8
1.2.6	Imputation	9
1.3	Extreme Value Background	11
1.3.1	Outlier Detection	11
1.3.2	Geospatial Outlier Detection	11
1.3.3	Geospatial Outlier Management	12
1.3.4	Extreme Value Theory (EVT)	14
1.3.5	Spatial Extreme Value Theory	15
1.4	Optimization Background	16
1.4.1	Inverse Problems	16
1.4.2	Genetic Algorithms	17
1.5	Thesis Outline	18
2	Outlier Management	20
2.1	Outliers and Extreme Values	20
2.1.1	Outlier Detection	21
2.2	Mining Industry Practices	22
2.2.1	Tools for Outlier Management	23
2.2.2	Current State of the Art	27
2.3	Spatial Outlier Detection	28
2.3.1	Methodology	29
2.3.2	Application	33
2.4	Analytical Extreme Value Models	36
2.4.1	Methodology	36
2.4.2	Probability of Drilling EHG	38
2.4.3	Application	39

2.5	Discussion	42
3	Network Model of Regionalization Framework	43
3.1	Problem Overview	43
3.2	Connectivity and Non-Gaussianity	47
3.2.1	Distribution of Runs	48
3.2.2	N -point Connectivity Function	49
3.2.3	Measures of Non-Gaussianity	50
3.3	Network Methodology	53
3.3.1	Notation	54
3.3.2	The NMR Inverse Problem	54
3.3.3	Latent Spatial Structure	55
3.3.4	Non-Linearity and Mapping	56
3.3.5	Two-Point Statistics	58
3.3.6	Multi-Point Statistics	60
3.3.7	Latent Imputation	61
3.3.8	Simulation and Mapping	62
3.4	Effects of High-Order Continuity	62
3.5	Discussion	65
4	Network Implementation	67
4.1	Network Design	67
4.1.1	Architecture	67
4.1.2	Activation Function	69
4.2	Latent Factor Design	71
4.3	Parameter Inference	74
4.3.1	Differential Evolution (DE)	74
4.3.2	Objective Function	76
4.3.3	Precedence	77
4.3.4	Complete Algorithm	78
4.3.5	Checks	78
4.4	Implementation Details	82
4.4.1	Non-Uniqueness	82
4.4.2	Number of Data	86
4.4.3	Computational Considerations	86
4.5	Discussion	88
5	Latent Factor Imputation	90

5.1	Imputation Concepts	90
5.2	Gibbs Sampler	92
5.3	Sequential Gaussian Rejection Imputation	93
5.3.1	Precedence and Constraints	94
5.3.2	Conditional Moments	95
5.3.3	Coarse Search	96
5.3.4	Solution Polishing	97
5.3.5	Imputation Checks	98
5.4	Latent Factor Simulation	102
5.4.1	Simulation Checks	104
5.5	Discussion	106
6	Case Study	108
6.1	Exploratory Data Analysis	108
6.2	Network Parameter Inference	112
6.3	Latent Factor Imputation	117
6.4	Latent Factor Simulation	120
6.5	Post Processing	122
6.6	Model Comparison	125
6.6.1	Global Results	125
6.6.2	Local Results	127
6.7	Discussion	132
7	Conclusions and Future Work	134
7.1	Summary of Contributions	134
7.1.1	Extreme Value Connectivity	134
7.1.2	Network Model of Regionalization	135
7.1.3	Sequential Gaussian Rejection Imputation	135
7.1.4	Simulation of Continuous High-Order Features	136
7.1.5	Spatial Outlier Detection	137
7.2	Limitations	137
7.2.1	Network Model of Regionalization	137
7.2.2	Latent Imputation	138
7.3	Future Work	139
	References	140
	A Software	153

Table of Contents

A.1 NMROPT	153
A.2 NMRIMP	157

List of Tables

2.1	Parrish decile analysis for a positively skewed distribution with 334 samples. The data is binned by deciles (0-9) and the upper decile is further split into percentiles (90-99). The Parrish methodology suggests a capping limit of 33.66.	24
2.2	Inferred and calculated model parameters.	41
2.3	Overall metal proportions by category.	41
2.4	Analytical versus simulated EHG probability.	41
3.1	Non-Gaussian metrics calculated for ten drillholes considering the 0.1 and 0.9 quantile indicator transforms. y -scores ≥ 2.5 are considered strongly non-Gaussian. DHID corresponds to Figure 3.7.	53
3.2	Selective mining unit (SMU) scale resources above the 0.1, 0.5, and 0.9 quantiles as a fraction of the true resources. Cutoff values are calculated from the true image. g/t=grams per tonne.	65
5.1	Six adjacent samples from a single drillhole, where z is the observed value, \tilde{z} is the mapped standard normal value, Factors 1-3 are the imputed latent Gaussian values, b is the ratio in Equation 5.10 and nresim is the number of resimulations at the given data location. The polishing tolerance is 0.01.	101
6.1	Original unit variogram model parameters.	110
6.2	Indicator variogram model parameters. All models have zero nugget.	110
6.3	Normal score variogram model parameters.	112
6.4	Covariance structures of the Gaussian pool (excluding the nugget).	114
6.5	ω bounds by factor.	114
6.6	Differential Evolution parameters.	114
6.7	Imputation parameters.	117
6.8	Simulation grid parameters.	120
6.9	SMU grid parameters.	124
6.10	% difference in expected value between ordinary kriging (OK) mean, sequential Gaussian simulation (SGS) expected value and network model of regionalization (NMR) expected value relative to test set mean by stope (≥ 2.0 unit cutoff grade (COG)). The table shows stopes with greater than 30 test data and test set deviation of less than 20%. Values are sorted in ascending order by training data mean value.	129

List of Figures

2.1	cumulative probability plot (CPP) showing a possible capping threshold based on upper tail decomposition (left), and the same CPP showing outlier and far outlier thresholds based on Tukey’s fences (Tukey, 1977).	24
2.2	Cutting curve showing the relationship between the top cut grade and the mean of the adjusted data (left) and cumulative coefficient of variation (CV) versus sample grade for a lognormal-like dataset (right). The cumulative CV plot shows a rapid increase at approximately CV=1.25. The capping limit corresponds to the value determined with the Parrish method.	25
2.3	Distribution of 1000 bootstrapped mean values with the top 2% of values removed (left), and the cumulative distribution functions (CDFs) of 50 of those realizations (right). . .	26
2.4	Summary of capping methodology (left) and, if capped, the data support to which the cap is applied (right) from 125 national instrument (NI) 43-101 reports. CPP - cumulative probability plot; P99 - 99 th percentile; CV - coefficient of variation; MIK - multiple indicator kriging; Parrish - decile analysis after Parrish (1997).	27
2.5	A sketch illustrating the spatial and density components of the spatial outlier algorithm. (1) shows an example search neighbourhood with $k = 5$ where $d(i, j)$ is the distance between location \mathbf{u}_i and \mathbf{u}_j , and r_{max} is the distance to the furthest neighbour. (2) shows the probability density function (PDF) of all samples fitted by a Gaussian mixture model (GMM) with $J = 2$ components. (3) shows an example domain with samples classified as inliers (black) and outliers (red); thresholding the outlier measure in Equation 2.5 results in the binary classification of each sample.	31
2.6	(a) one-dimensional (1-D) function of the coordinate position x ; (b) the difference between each sample the neighbourhood median, $h(x)$; (c) fitted GMM model with $J = 3$ components; (d) the combined outlier measure $g(x)$; (e) detrended $g(x)$ measure with flagged outliers; (f) the 1-D function from (a) with outliers identified. The red x’s denote the samples with added noise, and the green circles are the samples classified as outliers by the algorithm.	32
2.7	Location map of platinum group elements (PGE) sample locations.	33
2.8	CDF plot and CPP for the PGE distribution. The green shaded area in the CPP are Tukey’s fences (Tukey, 1977), identifying potential outliers.	33

2.9	Absolute difference between the sample value and corresponding neighbourhood (1 st row); area of influence for each sample (2 nd row); density measure (3 rd row); final outlier measure (4 th row) with flagged outliers above the $g(\mathbf{u}_i) = 0.6$ threshold. The left column shows values in the east-west direction, and the right column shows the north-south direction.	35
2.10	Location map of PGE sample locations with flagged outliers.	35
2.11	Trimodal distribution model of mineralized (M), high-grade (HG) and extreme high-grade (EHG).	37
2.12	Au CDF with log scaling (left) and log probability plot with Tukey's fences (Tukey, 1977) (right).	39
2.13	Location plot of all samples with extreme high-grade (EHG) locations highlighted in red.	40
3.1	Reference image showing the volume of shale for interbedded sands and shale of the McMurray Formation (left), a Gaussian realization generated with SGS (middle), and the corresponding n -point connectivity functions (right). The reference image is part of a core photograph.	44
3.2	High level flow chart illustrating the key components of the NMR framework. There are multiple junctions where previous steps can be revisited and refined. Reverse-flow arrows leading to multiple actions emphasize the interconnected nature of each component.	46
3.3	Indicator transform of a Gaussian random variable (RV) using the 0.1, 0.5, and 0.9 quantiles as thresholds.	48
3.4	Example of lower-order runs for a single run of 3. One run of 3 is also two runs of 2 and three runs of 1.	49
3.5	n -Point connectivity function for a string of non-Gaussian data (red) and the expected value of Gaussian realizations (gray) with the same covariance structure.	50
3.6	Schematic illustrating the calculation of the non-Gaussian Score y where z_{dh} is the metric from the drillhole, and μ_r and σ_r are the mean and standard deviation of the Gaussian distribution.	52
3.7	Image of a meandering river system sampled in ten locations to generate sequences of pixel data. DHID = drillhole ID, px = pixel.	53
3.8	Flowchart showing an overview of the optimization workflow for determining the parameter vector θ	55
3.9	Unconditional realizations of Gaussian RFs with orthogonal directions of continuity.	56
3.10	Power law relationship between the linear input y and the non-linear output $\phi(y, \omega)$, where $\phi(\dots)$ is the power law function.	57
3.11	Factors 1 (left) and 2 (center) after application of the power law function, and the final mapped, normal score transformed realization (right).	58

3.12	Histograms of factors 1 (left) and 2 (center) after application of the power law function, and the final mapped, normal score transformed realization (right).	58
3.13	Normal score experimental variogram (blue), the 0.1 quantile indicator experimental variogram (black), and the 0.9 indicator experimental variogram (red) of the mapped model in the north-south (left) and east-west (right) directions, respectively.	59
3.14	The final NMR model (left) and a SGS model generated with the same normal score variogram model.	59
3.15	Synthetic drillhole samples and corresponding 0.9 quantile indicator transform for the NMR model (left) and the SGS model (right).	60
3.16	N -point connectivity function for the 0.9 quantile indicator transform of the drillholes in Figure 3.15.	61
3.17	Bivariate scatter plots comparing the observed value, z , and factors 1 and 2. The histograms are the marginal distributions of each variable.	62
3.18	The non-stationary reference truth image of fork lightning (left) and the drillhole samples extracted from the image (right).	63
3.19	The reference truth block averaged to a 5 x 5 pixel SMU (top left) and corresponding 0.9 quantile indicator transform (bottom left), with the SMU scale NMR model and indicators (top middle, bottom middle), and the SMU scale SGS model and indicators (top right, bottom right)	65
4.1	Schematic representation of the NMR with $M = 3$ latent factors. $\phi(\dots)$ is the modified power-law (MPL) activation function. For clarity the ω parameter is excluded.	68
4.2	The MPL (left) and scaled MPL (right) activation function for various values of ω and input $\mathbf{y} \in [-5, 5]$. The scaled activation uses $\xi = G^{-1}(0.999) \approx 3.09$.	70
4.3	The top row shows factor 1 activated with $\omega_1 = 4$ (left), factor 2 activated with $\omega_2 = 1/\omega_1$ (center), and the normal score transform the mixture (right). The bottom row shows the gridded indicator variograms for the 0.1 quantile (black) and 0.9 quantile (red) in the north-south (left) and east-west (right) directions.	73
4.4	Sigmoid weighting function where $x = -1.5$. The x-axis is the range of \mathbf{y}_p , and the y-axis is the weight to the remaining $M-1$ factors.	78
4.5	Plan-view (left) and long-section oriented 340° (right) through the synthetic data set. Both sections have a tolerance of ± 25 meters.	80
4.6	Fitted experimental variograms for the normal score variable (a), the 0.1 indicator (b) and 0.9 indicator (c). Note the asymmetric destructuring of the indicator variograms.	80
4.7	N -point connectivity function above the 0.9 quantile indicator for the synthetic data set.	80

4.8	Continuous variogram reproduction for 25 realizations. The black line is the variogram model, the red line is the average variogram, and the shaded red area encloses the minimum and maximum variogram values. Left to right are the major, minor and vertical directions, respectively.	81
4.9	Indicator variogram reproduction for 25 realizations for the 0.1 (a) and 0.9 (b) quantiles. Left to right are the major, minor and vertical directions, respectively. The black line is the variogram model, the red line is the average variogram, and the shaded red area encloses the minimum and maximum variogram values.	82
4.10	N -point connectivity function reproduction for 25 realizations for the 0.1 quantile indicator (left) and 0.9 quantile indicator (right). The black line is the target model, the red line is the average experimental value, and the shaded red area encloses the minimum and maximum experimental values.	82
4.11	Normal score variogram reproduction (a) in the major, minor and vertical directions, respectively, and the components of θ (b) for the first scenario. PFI=permutation feature importance.	84
4.12	Normal score variogram reproduction (a) in the major, minor and vertical directions, respectively, and the components of θ (b) for the second scenario. PFI=permutation feature importance.	85
4.13	Normal score variogram reproduction (a) in the major, minor and vertical directions, respectively, and the components of θ (b) for the third scenario. PFI=permutation feature importance.	85
4.14	Square data spacings and corresponding numbers of data (a) and corresponding minimum sum of squared error values from NMR parameter inference(b). The total number of data increases to the right of the plot in (b).	87
4.15	Objective function value versus iterations for runs of the differential evolution (DE) and parallel differential evolution (PDE) algorithms.	87
4.16	Sensitivity of experimental variogram points to the total number of pairs per lag.	88
5.1	A schematic drillhole showing the multivariate relationships between observed and latent spaces. The observed data values are a function $\mathcal{F}_\theta(\mathbf{y})$ of the imputed latent factors at location \mathbf{u}_i . The imputed values must reproduce the true observed values, and each regionalized latent factor must reproduce its covariance structure γ_m	94
5.2	Schematic representation of latent factor sensitivity as a tornado chart. The x-axis shows $\Delta\tilde{z}$, or how much \tilde{z} changes for a given perturbation of Y_m . In this scenario, the target delta is a positive value. The y-axis shows $m = 1, \dots, M$ latent factors sorted by descending sensitivity. $a = \Delta z_t - c$; $b = \Delta z_t$; $c =$ the largest possible change to \tilde{z}	97

5.3	Scatter plot matrix for a single imputed realization highlighting the perfect correlation between z and \tilde{z} , the uncorrelated nature of all latent factors, and the standard normal nature of all marginal distributions. These features are the first three minimum acceptance criteria for sequential Gaussian rejection imputation (SGRI) imputed realizations.	101
5.4	Latent factor variogram reproduction for the first two factors in the major (left), minor (center) and vertical (right) directions, respectively. The expected variograms of the imputed realizations reproduce the models well, given the data density.	101
5.5	Plan view sections through the first conditional realization of the three gridded factors with imputed conditioning data.	102
5.6	Plan (a), east-west (b) and north-south (c) sections through the first NMR realization mapped to Gaussian space with observed data values in Gaussian units.	103
5.7	Plan (a), east-west (b) and north-south (c) sections through the first NMR realization back-transformed to original units.	103
5.8	CDF and data reproduction for the NMR realizations in Gaussian units (left and center) and CDF reproduction in original units (right). The original units reference distribution shown in red considers declustering weights.	105
5.9	NMR continuous variogram reproduction. The black line is the variogram model; the red dots are the experimental variogram of the imputed data; the red line is the average variogram of the gridded NMR realizations, the shaded red area encloses the minimum and maximum gridded NMR variogram values. Left to right are the major, minor and vertical directions, respectively.	105
5.10	NMR indicator variogram reproduction for the 0.1 (a), 0.5 (b) and 0.9 (c) quantiles. The black line is the variogram model; the red dots are the experimental variogram of the imputed data; the red line is the average variogram of the gridded NMR realizations, the shaded red area encloses the minimum and maximum gridded NMR variogram values. Left to right are the major, minor and vertical directions, respectively.	106
5.11	NMR 0.9 quantile cumulative run-length frequency reproduction. The black line is the target calculated from the drillholes, the red line is the average gridded value, and the shaded red area encloses the minimum and maximum gridded values.	106
6.1	CDFs for training (grey) and testing (red) datasets.	109
6.2	Plan (a), east-west (b) and north-south (c) sections through the training dataset.	109
6.3	Fitted original unit variograms in the major, minor and tertiary directions (left to right).	110
6.4	Fitted experimental indicator variograms for the 0.1 (a), 0.5 (b) and 0.9 (c) quantiles in the major, minor and tertiary directions (left to right).	111
6.5	Clustered (gray) and declustered CDFs (red) considering inverse distance squared (ID2) weights.	111

6.6	Fitted normal score variograms in the major, minor and tertiary directions (left to right).	111
6.7	Normal score variogram leave-one-out cross-validation (LOOCV) (left) and accuracy plot (right).	112
6.8	Cumulative run length frequencies above the 0.9 quantile from the drillhole data and the base-case SGS model.	113
6.9	DE objective function value (a) and final optimized network parameters (b).	115
6.10	Plan (a), east-west (b) and north-south (c) sections through the first unconditional realization of Gaussian values generated through parameter inference. Objective function components are evaluated using these realizations.	115
6.11	Continuous variogram reproduction for 25 realizations. The black line is the variogram model, the red line is the average variogram, and the shaded red area encloses the minimum and maximum variogram values. Left to right are the major, minor and vertical directions, respectively.	116
6.12	Indicator variogram reproduction for 25 realizations for the 0.1 (a), 0.5 (b) and 0.9 (c) quantiles. Left to right are the major, minor and vertical directions, respectively. The black line is the variogram model, the red line is the average variogram, and the shaded red area encloses the minimum and maximum variogram values.	116
6.13	Cumulative run-length frequency reproduction for 25 realizations. Thresholds 1, 2, and 3 correspond to the 0.1, 0.5 and 0.9 quantiles. The black line is the target run-length frequency model, the red line is the average experimental value, and the shaded red area encloses the minimum and maximum experimental values.	116
6.14	Scatter plot matrix coloured by kernel density estimate (KDE) illustrating the relationships between the observed data values, mapped imputed values and latent factors in Gaussian units. The histograms show the marginal distributions of each variable by row.	118
6.15	Variogram reproduction for 25 imputed realizations of each factor. The black line is the variogram model, the red line is the average variogram, and the shaded red area encloses the minimum and maximum variogram values. Table 6.4 specifies the Gaussian pool.	119
6.16	East-west section through the first realization of the gridded factors, plus the 0.9 quantile indicator transform of the latent factors mapped to observed space.	120
6.17	Plan (a), east-west (b) and north-south (c) sections through the first NMR realization mapped to Gaussian units.	121
6.18	CDF reproduction for 25 NMR realizations mapped to Gaussian units (left) and original units (right).	122
6.19	Plan (a), east-west (b) and north-south (c) sections through the first NMR realization back-transformed to original units.	122

6.20	NMR normal score variogram reproduction for 25 realizations. The black line is the variogram model, the red dots are the experimental variogram of the imputed data, the red line is the average variogram of the gridded NMR realizations, the shaded red area encloses the minimum and maximum gridded NMR variogram values, and the dashed grey line is the average variogram of the gridded SGS realizations. Left to right are the major, minor and vertical directions, respectively.	123
6.21	NMR indicator variogram reproduction for 25 realizations for the 0.1 (a), 0.5 (b) and 0.9 (c) quantiles. The black line is the variogram model, the red dots are the experimental variogram of the imputed data, the red line is the average variogram of the gridded NMR realizations, the shaded red area encloses the minimum and maximum gridded NMR variogram values, and the dashed grey line is the average variogram of the gridded SGS realizations. Left to right are the major, minor and vertical directions, respectively.	123
6.22	NMR 0.9 quantile cumulative run-length frequency reproduction for 25 realizations. 0.9 quantile from SGS is shown for reference.	124
6.23	Plan view section through the reference OK model showing stope shapes and the test dataset.	124
6.24	Distributions of prediction error for NMR (left) and SGS models (right) at all test set locations.	126
6.25	Cross validation scatter plots of NMR (left) and SGS (right) models and true data values at all test set locations. The black line is 1:1, and the red line is the regression of the truth on the estimate. RMSE=root mean squared error; SoR=slope of regression.	126
6.26	Accuracy plots of NMR (left) and SGS (right) models at all test set locations. n_p is the number of probability intervals.	126
6.27	Distributions of prediction error for NMR (left) and SGS models (right) at test set locations flagged within stope volumes.	127
6.28	Cross validation scatter plots of NMR (left) and SGS (right) models and true data values at test set locations flagged within stope volumes. The black line is 1:1, and the red line is the regression of the truth on the estimate. RMSE=root mean squared error; SoR=slope of regression.	127
6.29	Accuracy plots of NMR (left) and SGS (right) models at all test set locations. n_p is the number of probability intervals.	128
6.30	Grade-tonnage (GT) curves for all stope volumes for the NMR (left) and SGS (right) models.	128
6.31	Percent deviation from the test data set by stope ID. Stopes where train/test deviation is less than 20% are retained for comparison.	129
6.32	Mean grades by stope ID where test set deviation is less than 20%, and at least 30 data are present in the stope volume.	130

6.33	GT curves for stope 39 for NMR (left) and SGS (right) models. Stope 39 is a high-grade stope where the NMR performs well relative to SGS.	130
6.34	Sections through stope 39 showing the test data, OK model, and the first block averaged realization of the SGS and NMR models. Stope 39 is a high-grade stope where the NMR performs well relative to SGS. Training data is outlined in gray circles, while test data is in black.	131
6.35	GT curves for stope 47 for NMR (left) and SGS (right) models. Stope 47 is a high-grade stope where SGS performs well relative to the NMR.	131
6.36	Sections through stope 47 showing the test data, OK model, and the first block averaged realization of the SGS and NMR models. Stope 47 is a high-grade stope where SGS performs well relative to the NMR. Training data is outlined in gray circles, while test data is in black.	132

List of Symbols

Symbol	Description
α	Data matching tolerance
$A(\mathbf{u})$	Search area
argmin	Value at which the function output is minimized
F_Z	Cumulative distribution function of random variable Z
ρ	Correlation coefficient
$C(\mathbf{h})$	Covariance value at lag \mathbf{h}
CR	Differential evolution crossover probability
DHID	Drillhole ID
$d(i, j)$	Distance between locations i and j
\mathcal{D}	Domain of interest
$E\{ \ }$	Expected value
\mathcal{F}_θ	Parameterized NMR mapping function
g_i	Outlier measure
\mathbf{h}	Lag vector
i	Index for summation
$I(\mathbf{u}_i; z_k)$	Indicator value at threshold z_k
L	Number of realizations
M	Number of latent factors
μ	Mean
m	Meter
O_c	Objective function value of component c
ω	Power law activation function exponent
$\psi_{z_k}(n; \mathbf{h})$	n -point connectivity function at threshold z_k
$\phi(\mathbf{y}, \omega)$	Power law activation function of vector \mathbf{y} and exponent ω
$P(y z)$	Probability of y given z
σ	Standard deviation
F	Differential evolution scaling factor
θ	NMR parameter vector
\mathbf{u}	Location in space
$\gamma(\mathbf{h})$	Variogram value at lag \mathbf{h}

Symbol	Description
ix, iy, iz	Block model indices
ξ	Power law activation scaling factor
Z	Random variable
$z(\mathbf{u})$	Random variable outcome at location \mathbf{u}

List of Abbreviations

Abbreviation	Description
1-D	one-dimensional
2-D	two-dimensional
3-D	three-dimensional
CDF	cumulative distribution function
COG	cutoff grade
CPP	cumulative probability plot
CV	coefficient of variation
DE	differential evolution
EDA	exploratory data analysis
EHG	extreme high-grade
EM	expectation-maximization
EVT	extreme value theory
GA	genetic algorithm
GEV	generalized extreme value
GMM	Gaussian mixture model
GPD	generalized Pareto distribution
GSLIB	geostatistical software library
GT	grade-tonnage
HG	high-grade
ID2	inverse distance squared
iid	independent and identically distributed
KDE	kernel density estimate
LMC	linear model of co-regionalization
LMR	linear model of regionalization
LOOCV	leave-one-out cross-validation
M	mineralized
MCD	minimum covariance determinant
MCS	Monte Carlo simulation
MIK	multiple indicator kriging
ML	machine learning

Abbreviation	Description
MPDF	multiple-point density function
MPL	modified power-law
MPS	multiple point statistics
MSE	mean squared error
MWA	moving window average
NI	national instrument
NMR	network model of regionalization
OK	ordinary kriging
PCA	principal component analysis
PDE	parallel differential evolution
PDF	probability density function
PFI	permutation feature importance
PGE	platinum group elements
PPMT	projection pursuit multivariate transform
RGB	red-green-blue
RF	random function
RMSE	root mean squared error
RV	random variable
SCT	stepwise conditional transform
SGRI	sequential Gaussian rejection imputation
SGS	sequential Gaussian simulation
SK	simple kriging
SMU	selective mining unit
SOR	slope of regression
TI	training image

Chapter 1

Introduction

The field of geostatistics encompasses the problem of spatial prediction and the characterization of uncertainty within geological systems (Deutsch & Journel, 1992). We are concerned with describing the spatial aspects of mineralization and its variability or dispersion. Geostatistics was developed based on the need to forecast the recoverable resources at unsampled locations in mineral deposits (Matheron, 1963). Though initially developed in a mining context, geostatistics has found practical uses in many fields concerned with spatially correlated data, such as petroleum, hydrogeology, environmental science, remote sensing and others (Goovaerts, 1997).

Geostatistics utilizes observed categorical and continuous properties to generate exhaustive numerical models of the subsurface. These models are either deterministic or probabilistic, where equiprobable realizations are generated through stochastic simulation (Chilès & Delfiner, 2012). These realizations honour the spatial and multivariate characteristics of the input data with statistical fluctuations and provide a measure of joint uncertainty within the region of interest (Rossi & Deutsch, 2013). Characterizing and quantifying geologic uncertainty gives engineers and decision-makers practical tools for optimizing orebody extraction.

Mineral deposits, particularly precious metals, often exhibit strongly positively skewed grade distributions. These distributions pose challenges for spatial prediction as there is usually limited data characterizing the upper tail. Some components of the high values in the upper tail characterize “outliers” based on a subjective threshold. There is a risk of local overestimation with smooth kriging estimators if sparse, high-value data are left unmanaged (Leuangthong & Nowak, 2015). Standard practice in mining is to cap high values to a maximum to avoid local conditional bias, but those high values may have tremendous economic value. Appreciating the potential and the upside of such values in a quantitative and repeatable manner is of great practical interest. The term “extreme value” is not regularly used in the mining industry; rather, “outlier” is used. Some mineral distributions with high coefficients of variation likely do contain extreme values in the classic statistical sense. A mineral deposit is an extreme value in the context of regional geology.

1.1 Problem Setting

Mitigating the impact of extreme values on resource estimation is a long-standing issue. Extreme values present unique challenges because there are few samples, and understanding their spatial distribution is difficult. This lack of data is coupled with the potential of significant economic risk if mismanaged. High grades are typically capped in practice to mitigate the risk of overestimation.

Numerous problems remain outstanding concerning (1) objectively defining what an extreme value is, (2) explicit approaches for limiting extreme value influence, (3) characterizing the statistical or distribution component of extreme values, (4) characterizing the spatial component of extreme values and (5) develop practical advice to combine the statistical and spatial components.

The mining industry has not established a consensus regarding outlier management, and many approaches are developed case by case. Should high grades be capped? Should sub-regions be delineated to isolate higher grades? Should we use decile analysis, cumulative probability plots, cutting curves, coefficient of variation, production reconciliation, arbitrary percentiles, metal at risk, indicator correlations, multiple indicator kriging or no capping? If we cap, should it be before or after compositing? Answers to these questions influence the final metal content of resource estimates. The impact of restricting outliers may be significant depending on the distribution of the available data.

High-grade values are generally sparse in mining data sets. Though sparse, these high-grade values may contribute significantly to the project's economics. For example, at the Brucejack deposit in Northwest British Columbia, the top 1% of drill core samples contain 83% of the total deposit metal (Au) (Pretium Resources Inc., 2020). Anecdotal evidence from producing gold mines suggests this scenario poses an economic risk if samples in the upper tail are not explicitly managed. The use of smooth deterministic estimators such as kriging potentially exacerbates this issue. Many strongly positively skewed distributions also exhibit non-Gaussian spatial characteristics (Journel & Alabert, 1989). There is a need for a simulation framework that can correctly characterize the non-Gaussian spatial features related to extreme values without arbitrarily restricting their influence.

The spatial continuity of extreme values in the upper tail may differ from low values. The multivariate Gaussian assumption underlying many geostatistical algorithms does not allow for spatial connectivity of extreme values, nor does it allow for asymmetry in the loss of correlation away from the median (Journel & Alabert, 1989). Multiple indicator kriging (MIK) (Journel, 1983) was conceived for this purpose; however, it has proved difficult to implement effectively in practice, and simpler Gaussian techniques can outperform MIK (Vincent & Deutsch, 2021). The challenges associated with MIK and traditional extreme value management motivate the need for a new spatial model to characterize the non-Gaussian spatial continuity of extreme values.

1.1.1 Thesis Statement

A framework for the simulation of continuous variables in the presence of extreme values is proposed to address these challenges. The framework constructs numerical geologic models with explicit consideration for extreme values' presence and spatial structure. These models can better characterize high-grade geologic features by representing a regionalized variable by a non-linear combination of underlying latent Gaussian factors.

The framework consists of a non-linear network model of regionalization (NMR), which is an expansion of the linear model of regionalization (LMR) concept (Journel, 1974). Rather than a positive linear combination of latent factors, the network model of regionalization (NMR) introduces non-linear activations and a hidden layer to form a network structure. Non-linearity allows the spatial model to capture complex, high-order features and better control the known indicator asymmetry between low and high grades with non-Gaussian distributions (Journel & Alabert, 1989). High-order measures of connectivity are shown to characterize non-Gaussianity. These connectivity measures are extracted from drill strings and incorporated into the spatial models. Capturing richer spatial structures beyond what is possible with two-point statistics improves the prediction of high-grade in-situ resources. The NMR framework is particularly advantageous for strongly positively skewed distributions such as precious metals, uranium or diamonds.

Thesis Statement: *The breakdown of regionalized variables into fundamental latent components coupled with a non-linear network model of regionalization permits improved probabilistic modeling of strongly positively skewed grade distributions.*

The key contributions of this thesis are the development of:

1. The NMR framework for the simulation of high-order spatial features improves the modeling of continuous variables in the presence of extreme values. The framework includes:
 - a) Methodology for the parameterization of the network, permitting mapping between latent and observed spaces. This inverse problem is approached through stochastic optimization.
 - b) Methodology for stable imputation of latent factors that (1) reproduce the correct spatial statistics and (2) reproduce the observed data values.
 - c) A novel activation function to impose spatial features in the tails of the continuous distribution.
2. Tools for calculating high-order connectivity measures from drillhole sequences; these connectivity features are a proxy for non-Gaussianity.
3. An algorithm for identification of outliers in a spatial context. The algorithm considers the spatial arrangement and shape of the global empirical distribution to assign an outlier score.

Little research has been done on the continuous simulation of high-order spatial features without training images. In the NMR framework, all high-order features are extracted directly from the observed data with no assumptions made regarding the geological system or underlying physical processes. These one-dimensional (1-D) patterns are restricted to the drill strings; however, the network parameters enforce connectivity away from the data. Another key difference is the introduction of the latent Gaussian space. The local conditional cumulative distribution functions (CDFs) are

not approximated by a combination of high-order statistical moments but rather calculated directly under a multivariate Gaussian assumption. The non-Gaussian spatial model is constructed as a mixture of Gaussians; the transform from latent to observed space captures the high-order features.

1.2 Geostatistical Background

The proposed research covers a wide range of subjects involving modeling continuous, positively skewed variables with geostatistical simulation. The following section reviews the relevant geostatistical concepts. This section does not intend to be exhaustive but rather provide a review of concepts related to the NMR framework.

1.2.1 Overview

Geostatistics is a field of applied statistics concerned with characterizing and modeling spatially correlated variables. A variable dispersed in space and exhibiting spatial structure is said to be regionalized (Matheron, 2019). A foundation of geostatistics is the concept of random variables where unknown values, z , at an unsampled location are modeled as outcomes of a random variable Z (Deutsch & Journel, 1992). A random function represents a collection of spatially correlated, location-dependent random variables $Z(\mathbf{u})$ for every location \mathbf{u} within the study region (Goovaerts, 1997).

The decision of stationarity is one to group or pool relevant data together. As no data replicates are available at location \mathbf{u} to infer the random function $Z(\mathbf{u})$, geologically similar data must be pooled, permitting reliable inference of population statistics. This decision of stationarity allows the trade of unavailable replicates for data at other locations for statistical inference (Deutsch & Journel, 1992). Pooling too little data may lead to unreliable statistics, and too much data may lead to the masking of important geological features. *Stationarity* is a property related to the underlying random function model and cannot be checked or validated with data (Goovaerts, 1997).

Geostatistical estimation and simulation algorithms require inference of the random function's first and second-order moments (mean and covariance). When divided into a sub-region \mathcal{D} , the variable of interest is considered first-order stationary if the expected value is constant within \mathcal{D} . The variable is second-order stationary if the covariance depends only on the separation vector \mathbf{h} within \mathcal{D} . A random function $Z(\mathbf{u})$ is second-order stationary when:

$$\begin{aligned} E\{Z(\mathbf{u})\} &= \mu \\ E\{Z(\mathbf{u}) - \mu\}^2 &= C(0) = \sigma^2 \\ E\{Z(\mathbf{u}) \cdot Z(\mathbf{u} + \mathbf{h})\} - \mu^2 &= C(\mathbf{u}, \mathbf{u} + \mathbf{h}) = C(\mathbf{h}) \\ \forall \mathbf{u}, \mathbf{u} + \mathbf{h} &\in \mathcal{D} \end{aligned}$$

where μ , σ^2 and $C(\mathbf{h})$ are the mean, variance and covariance, respectively, and do not depend on location. The invariance of the random function parameters to a location within \mathcal{D} allows the relation $C(\mathbf{h}) = \sigma^2 - \gamma(\mathbf{h})$ which is the foundation of variogram interpretation (Pyrcz & Deutsch, 2014).

The linear model of regionalization (LMR) is used to fit experimental variograms with models that ensure positive definite covariance matrices. As not all combinations of variogram models may lead to a permissible model, the LMR constructs a random function $Z(\mathbf{u})$ to be the linear combination of $L + 1$ independent, standard random functions $\{Y_\ell(\mathbf{u}), \ell = 0, \dots, L\}$, each with its permissible variogram function (Goovaerts, 1997), plus the stationary mean:

$$Z(\mathbf{u}) = \sum_{\ell=0}^L a_\ell Y_\ell(\mathbf{u}) + \mu(\mathbf{u})$$

where L is the number of nested structures in the model. By convention, the isotropic nugget effect is the 0^{th} structure. The variogram model of can $Z(\mathbf{u})$ then be expressed as the sum of the variograms for each of the factors:

$$\gamma_z(\mathbf{h}) = \sum_{\ell=0}^L b_\ell \Gamma_\ell(\mathbf{h}), \quad b_\ell = (a_\ell)^2$$

where $\Gamma_\ell(\mathbf{h})$ is the variogram of Y_ℓ and b_ℓ represents the variance contribution of each $\ell = 0, \dots, L$ factors.

The LMR extends to the multivariate case with $k = 1, \dots, K$ coregionalized variables. With the liner model of coregionalization (LMC), each coregionalized random function, $Z_k(\mathbf{u})$, is also the sum of the standard, uncorrelated factors:

$$Z_k(\mathbf{u}) = \sum_{\ell=0}^L a_{\ell,k} Y_\ell(\mathbf{u}) + \mu_k(\mathbf{u})$$

where $a_{\ell,k}$ is the contribution of the ℓ^{th} factor the k^{th} variable. The direct and cross variograms can be expressed as:

$$\gamma_{k,k'}(\mathbf{h}) = \sum_{\ell=0}^L a_{\ell,k} a_{\ell,k'} \Gamma_\ell(\mathbf{h}), \quad k, k' = 1, \dots, K$$

The LMC is commonly modeled in the cokriging paradigm for multivariate covariance inference.

1.2.2 Factorial Kriging

As demonstrated with the LMR notation, the regionalized variable is characterized by $m(\mathbf{u})$, the $L + 1$ a_ℓ values and the $L + 1$ variograms $\Gamma_\ell(\mathbf{h})$. Factorial kriging aims to model each nested spatial structure present in the LMR for filtering or feature extraction. The idea is that each regionalized factor has a correlation structure responsible for a different scale of continuity, and they can be estimated independently (Matheron, 1982). The factors are estimated as linear combinations of the

data values Deutsch (2007):

$$z_\ell^*(\mathbf{u}) = a_\ell Y_\ell(\mathbf{u}) = \sum_{i=1}^n \lambda_{i,\ell} z(\mathbf{u}_i)$$

By convention, ordinary kriging is used, though there is no reason simple kriging cannot be used (Hong & Deutsch, 2007). The estimation weights for the ℓ^{th} factor are obtained by minimizing the estimation variance, leading to the factorial kriging equations:

$$\begin{cases} \sum_{j=1}^n \lambda_{j,\ell} C(\mathbf{u}_i, \mathbf{u}_j) + \mu_\ell = C_\ell(\mathbf{u}_i, \mathbf{u}), & i = 1, \dots, n \\ \sum_{j=1}^n \lambda_{j,\ell} = 0, & \ell = 0, \dots, L \end{cases}$$

The right hand side covariances are the covariances corresponding to the particular structure ℓ being estimated. The sum of the estimated factors returns the original ordinary kriging estimate:

$$z^*(\mathbf{u}) = \sum_{\ell=0}^L z_\ell^*(\mathbf{u}) + \mu^*(\mathbf{u})$$

Filtering properties of factorial kriging may be helpful if one is interested in removing a particular factor from estimated maps. For example, high-frequency variation from the nugget effect could be filtered by only considering the $\ell = 1, \dots, L$ factors. Factorial kriging extends to the multivariate context by considering the LMC fitted to the direct and cross variograms where each coregionalized variable $\{Z_k(\mathbf{u}), k = 1, \dots, K\}$ is a linear combination of the standard, independent factors $\{Y_\ell^v(\mathbf{u}), v = 1, \dots, K; \ell = 0, \dots, L\}$ (Wackernagel, 1988):

$$z_{\ell,k}^*(\mathbf{u}) = \sum_{v=1}^K a_{\ell,k}^v Y_\ell^v(\mathbf{u}) = \sum_{v=1}^K \sum_{i=1}^n \lambda_i^v z^v(\mathbf{u}_i)$$

The cokriging equations for a particular spatial component are then:

$$\begin{cases} \sum_{k'=1}^K \sum_{j=1}^n \lambda_{k',j} C_{k,k'}(\mathbf{u}_i, \mathbf{u}_j) + \mu_k = C_{k,k}^\ell(\mathbf{u}_i, \mathbf{u}), & k = 1, \dots, K; i = 1, \dots, n \\ \sum_{j=1}^n \lambda_{k,j} = 0 & k = 1, \dots, K \end{cases}$$

Again, the sum of the estimated factors returns the original regionalized variable:

$$z_k^*(\mathbf{u}) = \sum_{\ell=0}^L z_{\ell,k}^*(\mathbf{u}) + \mu_k^*(\mathbf{u})$$

Multivariate factorial kriging is a technique for characterizing the regionalized factors $Y_{\ell,k}$ from observations of Z_k . This technique is advantageous if the correlation between variables depends on scale and one would like to extract or filter a particular spatial structure.

Simple factorial kriging equations are equivalent to the simple kriging equations except for the right hand side covariance is the covariance of the nested structure being estimated (Hong & Deutsch, 2007):

$$\begin{cases} \sum_{j=1}^n \lambda_{\ell,j} C(\mathbf{u}_i, \mathbf{u}_j) = C_\ell(\mathbf{u}_i, \mathbf{u}) \\ i = 1, \dots, n & \ell = 1, \dots, L \end{cases}$$

$$z^*(\mathbf{u}) = \sum_{\ell=0}^L z_{\ell}^*(\mathbf{u}) + \mu$$

The sum of the estimated factors plus the mean equals the simple kriging estimate.

1.2.3 Simulating Continuous Variables

Kriging and its variants generate smooth, deterministic estimates that do not reproduce the true variability of the data (Deutsch & Journel, 1992). The smoothness of this estimate is not desirable if the transfer function is sensitive to extreme values; under-representing the variability may be consequential. Simulation methods reproduce the input variables' covariance structure and spatial variability while honouring data values at their locations (Goovaerts, 1997). Generating a set of equiprobable realizations captures uncertainty in the random variable (RV). Conventional simulation algorithms rely on the assumption that the input variables are multivariate Gaussian after a univariate normal score transform. Under the multivariate Gaussian assumption, all conditional distributions are defined by the normal equations and thus are linear combinations of the conditioning data (Chilès & Delfiner, 2012). Many algorithms are available for conditional simulation of Gaussian random functions (RFs). Pinto (2020) presents a comprehensive overview of the most common algorithms and best practices for selecting an algorithm given the problem context.

Gaussian simulation algorithms are prevalent in the mining industry, with sequential Gaussian simulation (SGS) likely being the most common (Rossi & Deutsch, 2013). SGS is a Monte Carlo simulation (MCS) technique for simulation of Gaussian RFs (Gómez-Hernández & Journel, 1993; Gómez-Hernández & Srivastava, 2021; Goovaerts, 1997; Isaaks, 1990). SGS requires (1) parameterizing a multivariate conditional CDF, and (2) drawing realizations. Given the dimensionality of the problems faced in mining, (1) is only possible if a parametric, multivariate Gaussian distribution is adopted (Leuangthong, Khan, et al., 2008). The *curse of dimensionality* (Bellman, 1961), plus the simplicity and tractability of the Gaussian distribution, precludes the use of other distributions. The following steps generalize the process of generating a realization with SGS:

1. Define a random path through the grid nodes.
2. At each location, calculate the first and second-order moments of the conditional CDF using the normal equations.
3. Randomly draw a simulated value from the conditional CDF.
4. Add the simulated value to the conditioning data.
5. Visit the next grid node in the path.

As the simulation progresses, the amount of conditioning data increases. In practice, the calculation of the conditional moments is restricted to a local neighbourhood about the location being

simulated to prevent unreasonably large systems of equations. Multiple realizations are generated by varying the random path and draws from the conditional CDFs.

1.2.4 Multiple Point Statistics

Two-point statistics summarize the relationship between points separated by a lag vector \mathbf{h} . Two-point statistics, such as the variogram or correlogram, are measures of linear continuity. Multiple point statistics (MPS) are measures of continuity between multiple spatial arrangements of points with the possibility of reproducing curvilinear or ordering patterns (Boisvert, Pyrcz, et al., 2007). There are numerous MPS including the n -point connectivity function (Journel & Alabert, 1989), the distribution of runs (Ortiz, 2003) and the multiple-point density function (MPDF) (Boisvert, Pyrcz, et al., 2007).

The foundation of multiple-point simulation algorithms (Guardiano & Srivastava, 1993; Strebelle, 2002) is the replacement of inference of two-point statistics with the MPDF, where the MPDF describes the frequency of occurrence of a particular pattern. The local conditional probabilities are derived from multiple-point configurations, allowing for the reproduction of non-linear features (Silva, 2014). A challenge of simulation with MPS is that inference of the MPDF with limited data. This challenge is overcome by extracting the MPDF from a training image (TI) (Journel, 2005). The TI acts as a substitute for a RF model and is an exhaustive image at the same support of realizations. The TI should have the expected geologic variability of the final models (Gómez-Hernández & Srivastava, 2021). Though the TI allows for inference of MPS not available from the data, one faces the challenge of selecting an appropriate TI. Boisvert, Pyrcz, et al. (2007) describes the choice of TI being analogous to variogram modeling in the two-point paradigm, and is of first-order importance. In general, the TI should represent the physics of the underlying geological process and be characteristic of the conceptual geology. TIs can be generated from outcrop data, object-based models, or process-based models (Tahmasebi, 2018). More recently, Minniakhmetov and Dimitrakopoulos (2022) present methodology for high-order simulation of categorical variables that does not rely on a training image.

Traditionally, MPS simulations focus on categorical modeling of stratigraphic deposits where object-based or process-mimicking models are applicable across geologic environments (Mariethoz & Caers, 2014). Multi-point simulation of continuous variables is approached with high-order spatial cumulants, discussed in the next section.

1.2.5 High-Order Simulation

High-order simulation methods are similar in concept to the multi-point simulation framework, with applications for both continuous and categorical variables. These methodologies are predominantly data-driven, complemented by a TI. Rather than inferring high-order conditional probabilities ex-

clusively from a TI, they are approximated by spatial cumulants calculated from the data. Dimitrakopoulos, Mustapha, et al. (2009) and Mustapha and Dimitrakopoulos (2010, 2011) propose high-order simulation based on spatial cumulants. A *cumulant* is defined as the logarithm of the moment-generating function of a RF. The idea is that spatial cumulants can generalize the covariance to orders beyond two; Dimitrakopoulos, Mustapha, et al. (2009) shows that the first and second order cumulants are the mean and variance, respectively. The high-order cumulants, similar to MPS, can characterize complex non-linear and non-Gaussian geologic features. By incorporating spatial cumulants up to order five, Mustapha and Dimitrakopoulos (2010) show the ability to capture multi-point periodicity, connectivity of extreme values and complex geometric characteristics. The key idea in the *HOSIM* approach (Mustapha & Dimitrakopoulos, 2011) is that the local conditional CDFs take the form Legendre polynomial expansions, where spatial cumulants approximate the polynomial coefficients (Mustapha & Dimitrakopoulos, 2010). Both a TI and available data are used to infer the Legendre polynomial coefficients; third- and fourth-order statistics are estimated from the data, while higher-order features come from the TI (Minniakhmetov & Dimitrakopoulos, 2022). Minniakhmetov and Dimitrakopoulos (2017) extends the high-order simulation framework to the multivariate context. Minniakhmetov, Dimitrakopoulos, et al. (2018), Yao, Dimitrakopoulos, et al. (2020) and Yao, Dimitrakopoulos, et al. (2021) present further refinements to the polynomial approximations.

Though largely data-driven, a drawback of these methodologies is that a TI is required. The inference of the spatial cumulants draws from both the TI and available data; if the high-order multi-point replicates are not available in the data, they are incorporated from the TI (Mustapha & Dimitrakopoulos, 2010; Yao, Dimitrakopoulos, et al., 2021). Selecting or generating a TI for continuous variables in mining problems is challenging without dense sampling; Minniakhmetov, Dimitrakopoulos, et al. (2018) and de Carvalho, Dimitrakopoulos, et al. (2019) use blast hole samples for TI construction. Yao, Dimitrakopoulos, et al. (2021) presents methodology for TI free simulation with aggregated kernel statistics.

1.2.6 Imputation

Imputation is a key component of probabilistic modeling of continuous heterotopic data (Barnett & Deutsch, 2015; Hadavand & Deutsch, 2023; Silva & Deutsch, 2018) and in truncated Gaussian categorical modeling techniques (Armstrong, Galli, et al., 2011; Arroyo & Emery, 2020; Madani & Bazarbekov, 2021; Silva, 2018). Simulation often considers a multiple imputation framework (Little & Rubin, 2019) where one generates realizations of missing values to transfer imputation uncertainty to the final models correctly.

The goal of imputation in the heterotopic data context is to fill in missing values. Modern geostatistical workflows necessitate the use of multivariate transforms like projection pursuit multivariate

transform (PPMT), stepwise conditional transform (SCT), and principal component analysis (PCA) which require homotopic data. Any heterotopic data observations must be either excluded or imputed. Geologic data is often “missing not at random” and simply excluding heterotopic observations can lead to biases in the final model (da Silva & Costa, 2019). The goal of any imputation algorithm is to define the distribution of the missing values conditional to the observed values. Barnett and Deutsch (2015) proposed a non-parametric methodology for imputing continuous variables based on Bayesian updating and Gibbs sampling. A simple kriging (SK) mean, and variance is merged with a collocated conditional distribution estimated with a multivariate kernel density estimate (KDE). The parameters calculated with SK account for the univariate spatial component, while the KDE accounts for the collocated multivariate component. Imputed values are drawn from the merged distribution. This methodology is computationally expensive with many data due to KDE calculations and Gibbs sampler iterations. Silva and Deutsch (2018) proposed a non-parametric imputation algorithm based on Gaussian mixture models (GMMs) to relieve the burden of KDE calculations. SK defines the local conditional spatial distribution, however the collocated multivariate density is sampled from a fitted GMM. Hadavand and Deutsch (2023) proposed another non-parametric methodology where deep learning characterizes the multivariate relationships, rather than a GMM. Two neural networks are trained to quantify the moments of the conditional missing value distribution; one for the mean and another for the second, third, and fourth-order moments. A lambda distribution is fit given the conditional moments that characterize the collocated multivariate relationship.

Latent imputation is a special scenario where all values are missing (Little & Rubin, 2019). Truncated Gaussian modeling techniques utilize the idea that categorical observations are generated by truncating underlying latent variables (Matheron, Beucher, et al., 1987). These latent variables are not observed and are a synthetic feature of the model. Imputation of latent variables subject to categorical observations is commonly approached with a Gibbs sampler (Geman & Geman, 1984), where directly sampling the multivariate truncated Gaussian distribution is not possible, but sampling the marginal conditional distributions is possible (Arroyo & Emery, 2020; Madani & Bazarbekov, 2021; Silva, 2018). Though the Gibbs approach is common, Emery, Arroyo, et al. (2014) and Silva (2018) note convergence issues with spatially correlated variables. More recently Lauzon and Marcotte (2020a, 2020b, 2023) proposed the sequential spectral turning band simulator as an alternative for Gibbs sampling, where Gaussian RFs are constructed by addition of cosine functions. The proposed spectral approach begins with Gaussian RFs that meets the inequality constraints. Then, the spatial component is introduced by sampling the spectral density. In contrast, the Gibbs sampler approach begins with Gaussian RFs with the correct spatial structure and gradually introduces constraints through Gibbs iterations. The authors show that the spectral approach is a valid alternative to the Gibbs sampler with stable convergence of many data, multiple rock types, and complex truncation rules.

1.3 Extreme Value Background

The following section provides background literature on extreme values, outliers, and their significance in geospatial and mining-related problems. Though the NMR framework does not directly incorporate extreme value theory (EVT), the statistical foundation is presented for completeness.

1.3.1 Outlier Detection

Outlier detection is relevant to all statistical modeling, and the literature is vast (Zimek & Filzmoser, 2018). Hodge and Austin (2004) note that authors may refer to outlier detection as novelty, noise, anomaly or deviation detection. In all cases, however, an outlier is an observation sufficiently dissimilar to other observations (Barnett & Lewis, 1984). Wang, Bah, et al. (2019) group outlier detection methodologies into (1) statistical methods, (2) distance-based methods, (3) density-based methods, (4) clustering-based methods, and (5) ensemble-based methods. Statistical methods, either parametric or non-parametric, compare the relationship of potential outliers with the remaining distribution. Distance-based methods (euclidean or non-euclidean) compare the distance between observations where potential outliers are “far” from other observations. Density-based methods consider outliers in low-density regions of a probability density function (PDF). Clustering-based methods classify each observation, and potential outliers are not within or near dense clusters. Finally, ensemble methods are combinations of dissimilar methodologies to create a more robust outlier detection model. Boukerche, Zheng, et al. (2021); Hodge and Austin (2004); Pimentel, Clifton, et al. (2014); Wang, Bah, et al. (2019) provide comprehensive reviews of outlier detection methodologies with applications to fraud detection, cybersecurity, sensor networks, image processing, time series and data streams, medical diagnostics and industrial monitoring. Pang, Shen, et al. (2022) presents a comprehensive review of outlier detection with deep learning, though the concepts are largely beyond the scope of this thesis.

1.3.2 Geospatial Outlier Detection

Outlier detection in the mining industry is based largely on graphical methods (Leuangthong & Nowak, 2015; Silva, 2021). These methods are necessarily subjective as the practitioner must interpret a plot and select a threshold to define an outlier. Leuangthong and Nowak (2015), Nowak, Leuangthong, et al. (2013), and Rossi and Deutsch (2013) provide practical advice on threshold selection for cumulative probability plots (CPPs) where breaks in the upper tail may represent outlier populations. Babakhani (2014) proposed a spatial bootstrap-based methodology to characterize the relationship between the naive and capped mean. The methodology aims to identify the values that cause higher mean values as potential outliers.

Practitioners often omit the spatial characteristics of sample values when identifying outliers. The spatial context of the samples is likely relevant; an extreme value surrounded by other high val-

ues may not be an outlier. The correlation structure of the variable(s) is relevant to understanding the spatial context (Filzmoser, Ruiz-Gazen, et al., 2014). Babakhani (2014) proposed a methodology for identifying spatial outliers based on the rank transform of cross-validation estimates. Outliers identified by the rank transform consider the spatial neighbourhood and are different from simply considering the univariate distribution. Many authors (Chen, Lu, et al., 2008; Ernst & Haesbroeck, 2017; Filzmoser & Gregorich, 2020; Filzmoser, Ruiz-Gazen, et al., 2014; Harris, Brunsdon, et al., 2014; Leung, Balamurali, et al., 2021) have proposed methodology for spatial multivariate geochemical outlier detection using the minimum covariance determinant (MCD) estimator of Rousseeuw and Driessen (1999). The Mahalanobis distance (Mahalanobis, 2018) is a common distance metric in multivariate space. However, it is sensitive to outliers (Filzmoser, Ruiz-Gazen, et al., 2014). The MCD is a robust measure of the global correlation structure in the presence of outliers. Given the MCD estimate, various measures of local and global multivariate distances are calculated to determine an outlier score. Chen, Lu, et al. (2008) calculate an outlying distance based on the differences between each sample value and the median value over its neighbours. Filzmoser, Ruiz-Gazen, et al. (2014) calculate a degree of isolation for each observation based on the robust Mahalanobis distance (MCD estimate) globally, as well as within a local neighbourhood; values above a defined threshold are considered potential multivariate outliers.

1.3.3 Geospatial Outlier Management

Once one identifies outliers, the practitioner must decide on a management strategy. Managing extreme values and outliers prior to resource estimation is a key component, particularly for heavy-tailed mineral deposits. The general idea is that unadjusted grades may lead to unrealistic local estimates adjacent to high-grade composites (Nowak, Leuangthong, et al., 2013). Leuangthong and Nowak (2015) provide an overview of outlier management practices in the context of mineral resource estimation. The practices are grouped into three general categories: (1) choosing appropriate domains, (2) grade capping, and (3) limiting the influence of outliers through the estimation process. The following sections touch on categories (2) and (3)—trade-craft and practitioner experience is the foundation of many suggested practices.

There are numerous practical methods to manage the influence of high-grade samples explicitly. Traditional methods published in mining technical reports largely fall into the grade-capping category. Techniques for selecting a capping threshold include decile analysis (Parrish, 1997), cumulative probability plots (Rossi & Deutsch, 2013), cutting curves (Roscoe, 1996), coefficient of variation thresholds (Parker, 1991), production reconciliation, arbitrary quantiles, metal-at-risk (Parker, 2006), and indicator correlations (Nowak & Leuangthong, 2019). More recently, Dutaut and Marcotte (2021) proposed using an error-free coefficient of variation (CV) calculated from coarse duplicate correlation to determine a capping limit.

Babakhani (2014) proposed a less conventional method of projecting outliers to an extra dimension to reduce the local influence. This method requires specifying a distance d to project the outlier; the distance could be determined by calibrating kriging results to an expected value from simulation or selecting a specific quantile. Babakhani (2014) also proposed calibrating a capping limit based on an expected value from simulation. This approach is similar to the previous; however, instead of projecting an outlier some distance, d , away, the outlier value is reduced. Within some local volume of influence, the outlier grade is reduced until the kriged grade matches the expected value of 100 simulated realizations. The idea is that simulation is more robust in the presence of outliers and this resistance is exploited to calibrate a capping grade. Rivoirard, Demange, et al. (2013) and Maleki, Madani, et al. (2014) propose the selection of an optimal capping limit through analysis of the ratios of direct and cross-indicator variograms. The goal is to identify the range $[z_{min}, z_{max}]$ in which the capping limit should fall. The minimum value is defined by the first threshold where the ratio of the cross and direct indicator variogram values are constant, independent of the lag vector. The maximum value is the threshold where the residual variogram is pure nugget.

Many methodologies have been proposed to circumvent the practice of capping. Some techniques, such as MIK (Journel, 1983), directly treat outliers through the specification of upper-class means and do not require explicit capping (Rossi & Deutsch, 2013). Costa (2003) proposed a variant of robust kriging (Hawkins & Cressie, 1984) where the weight to outlier samples differs from that of inliers. The “robust-edited” values are adjusted based on the difference between the sample value and the weighted median at the same location. Rivoirard, Demange, et al. (2013) proposed the decomposition of the grade value into a truncated grade, a weighted indicator above the top cut grade, and a residual. The residual is uncorrelated with the truncated grade and the indicator if the cutoff is sufficiently high. The final estimate is a kriged estimate of the residual plus a cokriged estimate of the indicator and truncated grade. Maleki, Madani, et al. (2014) proposed a similar decomposition approach where they suggest spatial prediction is improved by omitting outlier values from variogram calculations. Fourie, Morgan, et al. (2019) proposed a methodology that post-processes kriging weights to generate realistic estimates without grade smearing. Kriging weights are adjusted based on the ratio of the frequency of outlier samples to median samples. The methodology requires a subjective selection of bin widths after the normal score transform of the variable. More recently, Silva (2021) proposed a methodology for adjusting outlier grades based on Bayesian updating of the data distribution with a cross-validation error distribution.

Restricting the spatial range of influence of outlier values during estimation is another practical approach to circumvent capping. This restriction allows using extreme values or outliers in estimation without explicit capping. Restricting the search around extreme values allows them to influence within that range but not beyond it. There is no theory to support this restriction, though the idea is that the search should be within the range of high-grade continuity. The restricted search ellipsoid dimensions could come from understood mineralization controls, high-grade indicator vari-

ogram continuity analysis, or the p-gram (Leuangthong & Nowak, 2015). Leuangthong and Nowak (2015) provide an estimation example with a restitched search radii placed on high-grade samples in a South American gold deposit.

1.3.4 Extreme Value Theory (EVT)

Extreme value theory (EVT) is a statistical foundation for quantifying probability distributions and magnitudes of atypically high or low events. Early works by Fréchet (1927), Fisher and Tippett (1928), von Mises (1936) and later by Gnedenko (1943) systematically developed the theory and formalized the asymptotic distribution of extremes for block maximums. Fisher and Tippett (1928) show that for a sequence of independent and identically distributed (iid) random variables with a common distribution function F , the so-called “block maximum”, $M_n = \max\{Z_1, \dots, Z_n\}$, can only converge to one of three distributions:

$$G_I(z) = \exp\left(-\exp\left(-\left(\frac{z-\mu}{\sigma}\right)\right)\right), \quad -\infty < z < \infty \quad (1.1)$$

$$G_{II}(z) = \begin{cases} 0, & z \leq \mu \\ \exp\left(-\left(\frac{z-\mu}{\sigma}\right)^{-\xi}\right), & z > \mu \end{cases} \quad (1.2)$$

$$G_{III}(z) = \begin{cases} \exp\left(-\left(-\left(\frac{z-\mu}{\sigma}\right)^\xi\right)\right), & z < \mu \\ 1, & z \geq \mu \end{cases} \quad (1.3)$$

To avoid degeneracy, the variable M_n is normalized to $M_n^* = a_n^{-1}(M_n - b_n)$ for a sequence of “normalizing” constants $a_n > 0$ and b_n . Fisher and Tippett (1928) show that if these sequences of real numbers can be chosen such that M_n^* has a non-degenerate limiting distribution, it must be one of type I, II or III. These are the only possible limits for the distributions of M_n^* regardless of the population distribution F_Z . This Extremal Types Theorem is analogous to the Central Limit Theorem for extreme values (Coles, Bawa, et al., 2001).

These collectively are termed extreme value distributions with Equation 1.1 being Gumbel-type, Equation 1.2 being Fréchet-type and Equation 1.3 being Weibull-type. Each distribution has a location μ , scale σ and Equation 1.2 and Equation 1.3 have shape parameter ξ . The extreme value distributions are obtained as limiting distributions of M_n as $n \rightarrow \infty$. The limit distributions for block maximum can be grouped into a single family termed the generalized extreme value (GEV)

distribution (De Haan & Ferreira, 2007):

$$G(z) = \exp\left(-\left(1 + \xi\left(\frac{z - \mu}{\sigma}\right)^{\frac{-1}{\xi}}\right)\right) \quad (1.4)$$

The block maximum methodology may be inefficient as it ignores all but the maximum value in a given block (Davison & Huser, 2015). The second approach to identifying extremes is the so-called “peak-over-threshold” (POT) method. Consider a random variable Z with distribution function F . The CDF of the excess over some threshold u , is defined by:

$$\begin{aligned} F_u(y) &= P(Z - u \leq y | Z > u) \\ &= \frac{F(u + y) - F(u)}{1 - F(u)}, \quad 0 \leq y \leq z_F - u \end{aligned} \quad (1.5)$$

where $y = z - u$ and z_F is the right endpoint of F (Gilli & Kellezi, 2006). Pickands (1975) states that if u is large, the conditional distribution of Z given Z is much larger than u is well approximated by the generalized Pareto distribution (GPD):

$$G(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{\frac{-1}{\xi}}, & \text{if } \xi \neq 0 \\ 1 - \exp\left(\frac{-y}{\sigma}\right), & \text{if } \xi = 1 \end{cases} \quad (1.6)$$

The conditional distribution of the exceedances (Equation 1.5) can be modeled asymptotically with the GPD by estimating the scale (σ) and shape (ξ) parameters.

In a geoscience context, Caers, Beirlant, et al. (1999a, 1999b) use the GPD to model earthquake magnitudes, size distributions of diamonds and the size distributions of impact craters. Deligne, Coles, et al. (2010) use a Poisson process to model the recurrence rate of explosive volcanic eruptions, while Nguyen, Veraart, et al. (2023) use the GPD to forecast volcanic eruptions. Miniussi, Marani, et al. (2020) model the frequency of flooding events across the United States based on stream gauge measurements. Lee, Kim, et al. (2021) use a Gumbel distribution to predict the exceedance probability of extreme rainfall-induced landslides.

1.3.5 Spatial Extreme Value Theory

The assumption of independent and identically distributed (iid) observations underlies classical EVT. In many real-world applications, one must account for correlation in space or time and the multivariate nature of regionalized variables. Spatial extreme value theory represents an intersection between classical EVT and geostatistics (Neves, 2015). The primary difference between spatial EVT and geostatistics is that in the geostatistical framework, there are no observed replicates at $Z(\mathbf{u})$. Fitting of a GEV or GPD in the classical EVT sense requires multiple realizations of $Z(\mathbf{u})$ for parameter inference.

Spatial EVT builds on the concepts of max-stable distributions extending to the max-stable process. A max-stable process is the infinite dimension generalization of the max-stable distribution

where all lower-order marginal distributions are GEV distributions (Schlather & Tawn, 2003). If there exists normalizing constants $a_n(\mathbf{u}) > 0$ and $b_n(\mathbf{u})$ such that $a_n^{-1}(\mathbf{u})\{\max_{i=1,\dots,\infty} Z_i(\mathbf{u}) - b_n(\mathbf{u})\} = Y(\mathbf{u})$ then Y is a max-stable process (De Haan & Ferreira, 2007). The max-stable process applies to maximums as stable Gaussian processes with finite variance apply to the average (Chilès & Delfiner, 2012). All marginal distributions of a max-stable process are GEV distributions defined by Equation 1.4.

Unlike a Gaussian RF, which is fully defined by its correlogram $\rho(\mathbf{h})$, there is no unique model for max-stable processes (Chilès & Delfiner, 2012). Many models are found in the literature. First introduced by Smith (1990) and later modified by Schlather (2002), Gaussian storm and extremal Gaussian processes are commonly used for modeling spatial extremes. The Brown-Resnick processes (Brown & Resnick, 1977; Kabluchko, Schlather, et al., 2009) relaxes the assumption of second-order stationarity and permits the use of the variogram, which has shown to be practical in practice (Gaume, Eckert, et al., 2013). The extremal- t model (Opitz, 2013) is another popular max-stable model in the literature. A consequence of the max-stable processes is asymptotic dependence in the tails (Davison, Huser, et al., 2013).

1.4 Optimization Background

The following section provides background literature regarding geoscience and engineering-related inverse problems and the use of optimization to infer unknown model parameters.

1.4.1 Inverse Problems

Inverse problems encompass a broad class of problems where the objective is to infer a system's underlying causes or parameters from observed data or measurable outputs (Sen & Stoffa, 2013). Predicting a response is a forward problem while using a response or observed measurements to infer the properties of a model is an inverse problem (Tarantola, 2005). Inverse problems arise in various scientific disciplines, including physics, engineering, geosciences, medical imaging, and more. Geospatial inverse problems are common in both the fields of geophysics (Giraud, Lindsay, et al., 2019; Grana, Azevedo, et al., 2022; Linde, Renard, et al., 2015) and hydrogeology (Ghorbanidehno, Kokkinaki, et al., 2020; Zhou, Gómez-Hernández, et al., 2014) where the underlying geologic model is unknown, however a set of measured responses, such as hydraulic conductivities or seismic properties, are known. The inverse problem involves inferring interpretable geologic properties of the unknown model, such as lithology or porosity, that satisfy the observed measurements. Grana, Azevedo, et al. (2022) describe these problems as rock-physics inversions with seismic measurements predicting rock and fluid properties.

Solving inverse problems involves constructing a mathematical forward model that describes the relationship between the unknown parameters and the observed data and then using this model

to infer the unknown parameters. Seismic wave propagation and rock-physics models are generally well-understood forward models in geophysics (Grana, Azevedo, et al., 2022) where hydrogeological forward models consider mass conservation and Darcy’s law to predict hydraulic head, drawdown or solute concentrations (Zhou, Gómez-Hernández, et al., 2014). A challenge of inverse problems is ill-posedness, or the lack of a unique solution (Tarantola, 2005). Multiple (or infinite) solutions may be valid given the observed data. As an exact solution is rarely possible in natural, non-linear systems, one looks for solutions close to actual observations (Bárdossy & Hörning, 2016).

For this reason, many inverse problems are framed as optimization problems, minimizing an objective function relevant to the problem at hand (Athens & Caers, 2022; Giraud, Lindsay, et al., 2019; Nava-Flores, Ortiz-Alemán, et al., 2023). The objective function is minimized iteratively, which is generally computationally expensive (Zhou, Gómez-Hernández, et al., 2014). A forward modeling operator predicts an outcome for the model parameters’ current state, and the objective function evaluates the loss between this prediction and the observed measurements. The optimization algorithm updates the parameter vector until it matches the model output and observed measurements. Any iterative optimization algorithm is permissible; Athens and Caers (2022) use gradual deformation to generate a set of perturbed model realizations; Nava-Flores, Ortiz-Alemán, et al. (2023) use simulated annealing for joint inversion of gravity gradient data; Balkaya, Ekinci, et al. (2017) use differential evolution, and Dávila Rodríguez, Palafox González, et al. (2024) a general evolution strategy for inversion of magnetic anomalies.

1.4.2 Genetic Algorithms

Genetic algorithms (GAs) are metaheuristic global optimization algorithms inspired by natural processes like evolution and natural selection (Cui, Zhang, et al., 2024). This family of algorithms was first proposed by Holland (1992) and has wide-ranging applicability to engineering optimization problems as robust global optimizers that do not require differentiable objective functions (Carbas, Toktas, et al., 2021). The GA framework can efficiently explore a large solution space, handle constraints, and lend itself to parallel applications. In general, genetic algorithms are derivative-free, population-based algorithms with three main characteristics: crossover, mutation, and selection operators (Yang, 2018). Each member of the population is a feasible solution vector. The crossover operator exchanges information between members of the population and can enhance key features of the population; the mutation operator permits exploration of the solution space by increasing population diversity; the selection process drives the population towards convergence by selecting the fittest or most elite member of the population for mutation and crossover (Yang, 2018). A problem-specific objective function evaluates the fitness of each population member. Yang (2018) describes the general behaviours and components of GAs as:

- A population of agents representing solution vectors, each with an associated fitness.

- The population evolves through mutation and crossover operations. The algorithm converges when all members of the population are sufficiently similar.
- New solutions are generated with random perturbations to avoid local optima.
- GAs search locally and globally, with the local and global search ratio controlled by the genetic operators.
- GAs employ a survival of the fittest approach, where the fittest members of the population are retained for the next generation, driving the population towards convergence.

Differential evolution (DE), first proposed by Storn and Price (1997), is a GA that uses the scaled difference between population members as a mutation operator. Many mutation variants exist (Meng, Chen, et al., 2020); purely random mutation variants explore the global solution space by exploiting differences between randomly selected vectors, while others explore both local and global spaces by mutating random vectors and the current best vector. Crossover occurs after mutation, generating a trial vector. If a randomly generated uniform number $\in [0, 1]$ is less than a specified crossover probability, the element from the mutated vector transfers to the current member of the population (Price, 2013). The crossover generates a trial vector hybrid between the current population and the mutation. The selection operator evaluates the fitness of the trial vector and replaces the current population member if $f_{trial} < f_{pop}$. If the trial vector improves the solution, it is kept in the population. DE is widely used in engineering optimization problems (Georgioudakis & Plevris, 2020), geophysical inversion (Balkaya, Ekinici, et al., 2017), optimization of neural network architectures (Mirjalili, 2019; Ünal & Başçiftçi, 2022), and others including electrical power systems, image processing, chemical engineering and manufacturing (Bilal, Pant, et al., 2020). Due to the widespread use of DE, Ahmad, Isa, et al. (2022) report over 40 variants of the original algorithm.

1.5 Thesis Outline

Chapter 2 discusses outlier management in the mining industry. Though the NMR framework does not require explicit management of extreme values, it would be remiss not to discuss capping due to its ubiquitous presence and relation to outliers. A range of outlier management tools are discussed and a novel algorithm for identifying outliers in a spatial context is presented. The chapter finishes with an analytical model for predicting the frequency of intersecting extreme values.

Chapter 3 explores the core components of the NMR framework. It begins with the concepts of high-order connectivity and the relationship with non-Gaussianity. The chapter introduces the network components: (1) definition of a latent Gaussian pool, (2) non-linearity and mapping to observed space, (3) latent imputation, and (4) continuous simulation and mapping. Here, the parameterization of the NMR is posed as an inverse problem. The chapter finishes with a synthetic, non-Gaussian example to emphasize the effects of high-order connectivity on contained resources.

Chapter 4 presents details of implementing the NMR, including the network architecture, activation function, latent factor design and parameter inference via optimization. The effects of mixing latent factors are discussed in detail, and an objective function relevant to the modeling goals is formulated. The concepts of differential evolution (DE) and its application to NMR parameter optimization are presented, followed by checking and validating the network output. A three-dimensional (3-D) example is introduced and carried over into chapter 5. Finally, practical implementation details are discussed including the potential non-uniqueness of the solution.

Chapter 5 introduces a novel algorithm for imputing latent factors within the NMR framework. Imputation concepts and traditional Gibbs sampler approaches are touched on, followed by a presentation of sequential Gaussian rejection imputation (SGRI). The algorithm is an iterative, sequential imputation algorithm that uses the normal equations and rejection sampling to impute spatially correlated latent variables. Minimum acceptance criteria for checking the latent realizations are presented, followed by conditional simulation and practical checking using the example from Chapter 4.

An application of the complete NMR framework with a real dataset is shown in Chapter 6. The data comes from an operating underground mine where personnel note that multivariate Gaussian simulation algorithms do not reproduce the connected high-grade features observed in drillhole data. This scenario is the ideal application of the NMR, where it shows a 7% improvement over SGS in high-grade stopes. The NMR results are validated with a hold-out dataset.

Chapter 7 summarizes the contributions made in this thesis. Consideration is given to the limitations of the developed methodologies and avenues for future work and improvements to the NMR framework. All software developed for this research is documented in the appendices.

Chapter 2

Outlier Management

This chapter presents some underlying motivation the NMR framework. The generation of non-Gaussian spatial fields and the presence of extreme values are linked. Research initially focused on developing objective measures to identify spatial extreme values and best practices for explicitly managing these values. Then it evolved into a simulation framework with applicability beyond extreme values. A key idea of the NMR is that extreme values and outliers do not require explicit management. This idea contrasts many standard practices in the mining industry; Dutaut and Marcotte (2021) mention that methods that avoid capping are interesting but rarely applied in mining applications. The presence of extreme values and the correct characterization of their spatial distribution are important. This importance warrants methodologies beyond the standard graphical approaches (Silva, 2021) and motivates a holistic approach incorporating both the statistical and spatial components of extreme values.

The following sections present the concepts of outliers and extreme values in a mining context. Nomenclature is first defined, delineating the differences between an outlier and an extreme value. These terms have similar connotations, though different meanings when rigorously defined. An overview of outlier management practices in the mining industry is given, including commonly employed tools and methodologies, followed by a review of methodologies from a survey of 125 national instrument (NI) 43-101 reports published between 2019 and 2021. Though the NMR framework does not require explicit management of extreme values, the practice is ubiquitous in the mining industry. For this reason, a spatial outlier identification algorithm is developed that considers a data point's degree of "outlierness" from a local neighbourhood perspective and the global CDF. The final section presents an analytical approach for forecasting extreme values. Though one must make some assumptions regarding the underlying distribution, predicting the frequency of intersecting extreme values is valuable from a data collection and risk-qualified decision-making perspective.

2.1 Outliers and Extreme Values

"Outlier" is a general term for an observation sufficiently dissimilar to other observations that further investigation is warranted (Barnett & Lewis, 1984). Outliers may be random fluctuations of the data generation mechanism (noise), true anomalies, or measurement errors. An extreme value is a value in the tails of the distribution that is believed to be real but occurs rarely. Extreme values are different from outliers in that all extreme values are possible outliers, but the reverse is not always true

(Aggarwal, 2016). A key distinction here is that an outlier is not necessarily restricted to the tails of a distribution, while extreme values are. Consider the 1-D set of values: $\{1, 2, 2, 50, 98, 98, 99\}$. In the extreme value context mentioned above, 1 and 99 could (weakly) be considered extreme values, while 50 (the average) is not an extreme value. However, in the context of an outlier, the value of 50 is distant or isolated from the remaining values. Distance- or density-based outlier detection methods would likely classify 50 as an outlier, which is correct given that it is sufficiently dissimilar from the remaining values. This simple but illustrative example adapted from Aggarwal (2016) highlights the core differences between outliers and extreme values. The terms are typically synonymous in the mining industry: one is only interested in outliers if they are also extreme values. Throughout this text, the term outlier will refer to abnormal or extreme data values assumed to be in either the upper or lower tail of the distribution.

Practitioners often generate geostatistical models using widely spaced data configurations. Data collection costs prohibit exhaustive sampling and necessitate statistical inference from limited samples. Spatial prediction with widely spaced data in the presence of extreme values is a long-standing issue in the mining industry (Leuangthong & Nowak, 2015). Extreme values may have significant local influence, leading to overstated resources and the risk of production shortfalls. Practitioners face difficult decisions when limiting extreme value influence and characterizing their spatial continuity. Inputs to numerical geologic models consist of observed data, a representative histogram and spatial controls on mineralization. Each of these components presents challenges in the presence of extreme values. Extreme values are often under-sampled, making inferences about their probability of occurrence difficult. The influence of extreme values is often limited in practice through grade capping, which could significantly impact the final resource. The spatial continuity of extreme values differs from that of the barren or mineralized background. Traditional geostatistical methods are limited in capacity to adapt to both extreme values and asymmetric spatial continuity features.

2.1.1 Outlier Detection

Outlier detection is applicable in virtually all statistical modeling. Measures of “inlierness” or “outlierness” are typically based on (1) statistics of observations for the rest of the distribution (parametric or non-parametric); (2) distances (euclidean or non-euclidean) between observations with outliers being “far” from neighbours and (3) probability density-based measures where outliers have low densities (Li, Zhao, et al., 2022). A comprehensive review of outlier detection methods is beyond the scope of this chapter; the reader is referred to Aggarwal (2016); Hodge and Austin (2004); Leuangthong and Nowak (2015); Nowak and Leuangthong (2019); Pang, Shen, et al. (2022); Wang, Bah, et al. (2019) for a review of methodology in the mining industry and beyond.

A challenge of many outlier detection techniques is the choice of a threshold to delineate an abnormal measurement. Whether the technique is statistical or proximity-based, a threshold must

be chosen to classify samples based on the measure of “outlierness” or the outlier score. If the threshold is too restrictive, the algorithm may not identify true outliers, and if it is too relaxed, it will lead to false positives. For simple methodologies, selecting a threshold may have physical meaning, such as a grade distribution. For spatial (multivariate) outlier detection, selecting a proximity-based threshold in a possibly non-euclidean space is non-trivial. The following section details outlier detection methodologies specific to the mining industry; generally, these approaches are univariate and do not consider the spatial or proximity component of outliers. Threshold selection in these cases is subjective but straightforward.

2.2 Mining Industry Practices

Identifying and managing extreme values is essential, particularly concerning heavy-tailed mineral deposits and smooth deterministic estimators. A practitioner forecasting resources should understand extreme values’ statistical influence and spatial distribution. Though numerous tools exist for these purposes, there is no generally accepted workflow in the mining industry, and professional judgment guides best practices. Assessing the influence of extreme values with a variety of techniques seems reasonable. The Canadian Institute of Mining, Metallurgy and Petroleum (CIM) mineral resource and mineral reserve best practices summarizes outlier management as (CIM Mineral Resource & Mineral Reserve Committee, 2019):

“Recognition of the spatial extent of outlier values (a component of grade continuity) should be investigated and a procedure devised for incorporating such data appropriately into an estimate. Procedures including domaining, grade capping (also known as top cutting), spatially restricting the influence of high-grade assays, single and multiple indicator kriging, and Monte Carlo simulation methods all compensate in varying ways for potential overestimation. Regardless of the methodology selected, the Practitioners must provide documentation of the approach selected, along with justification and support for the decision, possibly including reconciliation of estimated block model grades with available production information. Comparisons of the outcome of the different approaches can be useful.” (pg. 18)

Leuangthong and Nowak (2015), echoed by CIM best practices, break the process of outlier management into three categories: (1) choosing appropriate domains, (2) grade capping, and (3) limiting the influence of outliers through the estimation process. Grade capping or “top-cutting” is common in the mining industry. Grades above a given threshold are reset to that threshold. The general idea is that uncapped grades may lead to unrealistic local estimates adjacent to high-grade composites (Nowak, Leuangthong, et al., 2013). The “smearing” may be significant if sparse data are estimated with kriging. Due to the normal score transform, simulation is more robust in the presence

of outliers, though some cases may still require capping. As point scale realizations reproduce the input CDF, directly capping the realizations avoids iterative re-simulation (Harding, Lagos, et al., 2023).

2.2.1 Tools for Outlier Management

This section discusses commonly used tools for outlier management in the mining industry. These tools are largely qualitative, provide general guidance, and require subjective decision-making from the practitioner. Statistical methods attempt to guide the selection of a capping limit by exploring characteristics of the distribution tails or the relationship between contained metal and high-grade restrictions. Some methodologies, such as the p-gram (Nowak & Leuangthong, 2019), attempt to characterize the spatial continuity of high grades to support restrictions on their area of influence. Simulation-based methods such as metal-at-risk (Parker, 2006) or mean uncertainty (Nowak, Leuangthong, et al., 2013) “resample” the deposit to characterize the uncertainty in metal over some number of realizations. This section does not intend to be an exhaustive list of the available tools; however, it presents some of the most frequently used methodologies.

Probability plots are ubiquitous in resource estimation. The variable value is plotted on the x-axis against the corresponding cumulative probability of the normal distribution on the y-axis. Log scaling of the x-axis results in a lognormal probability plot. Inflection points may indicate the presence of multiple populations, and gaps in the distribution are typically targeted as potential capping limits. A survey of 125 43-101 reports with gold as the primary commodity (Section 2.2.2) shows that the CPP is the most common tool for assessing and classifying outliers. This prevalence is likely due to (1) the simplicity of the technique and the ease of implementation and (2) the historical prevalence of the technique in the mining industry. Figure 2.1 illustrates a log-probability plot for a synthetic positively skewed distribution with some tail decomposition. A proposed capping limit is selected where the upper tail begins to break down. Tukey’s fences are a simple non-parametric method to identify outliers where the fence intervals are defined by (Tukey, 1977):

$$[Q_1 - k(Q_3 - Q_1), Q_3 + k(Q_3 - Q_1)] \quad (2.1)$$

where Q_1 and Q_3 are the first and third quartiles, respectively. Generally, $k = 1.5$ defines an “outlier” while $k = 3$ indicates a “far outlier”.

Decile analysis or the Parrish method (Parrish, 1997) assesses the metal content of each decile of the grade distribution. The assay population is sorted and arranged into deciles containing equal samples; the upper decile will likely contain less data than the rest. The length-weighted mean and standard deviation of each decile are calculated, as well as metal content (grade value x length). The top decile is split into percentiles, and the same summary statistics are calculated for each. Table 2.1 shows an example decile table for a synthetic positively skewed distribution. The general rules proposed by Parrish (1997) are:

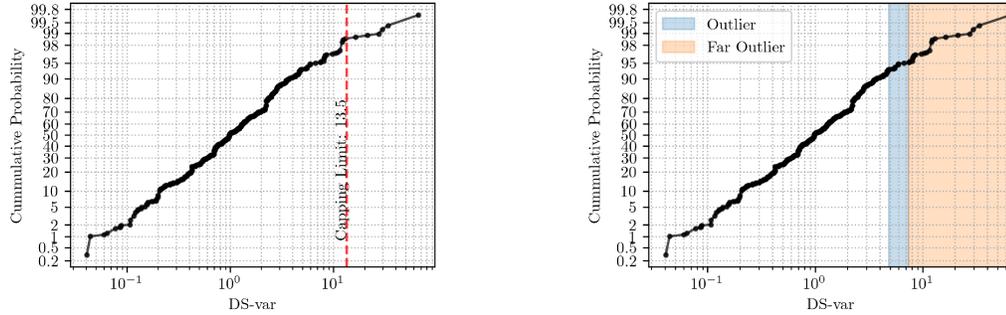


Figure 2.1: CPP showing a possible capping threshold based on upper tail decomposition (left), and the same CPP showing outlier and far outlier thresholds based on Tukey’s fences (Tukey, 1977).

1. If the top decile contains more than 40% of the metal, capping is warranted.
2. If the top decile contains more than twice the metal of the 80-90% decile, capping is warranted.
3. If the top percentile (or more) contains more than 10% of the total metal capping is warranted.

Table 2.1: Parrish decile analysis for a positively skewed distribution with 334 samples. The data is binned by deciles (0-9) and the upper decile is further split into percentiles (90-99). The Parrish methodology suggests a capping limit of 33.66.

Decile	count	mean	std	min	25%	50%	75%	max	Metal	% Total
0	34	0.14	0.05	0.04	0.11	0.15	0.19	0.23	4.5	0.7
1	34	0.34	0.06	0.23	0.29	0.35	0.40	0.43	10.0	1.5
2	34	0.56	0.07	0.43	0.51	0.55	0.62	0.68	15.9	2.4
3	34	0.78	0.07	0.70	0.72	0.76	0.84	0.90	23.3	3.5
4	34	1.06	0.08	0.92	0.99	1.02	1.12	1.19	27.0	4.0
5	34	1.32	0.09	1.21	1.26	1.29	1.37	1.51	41.4	6.2
6	34	1.82	0.23	1.51	1.59	1.77	2.02	2.17	51.0	7.6
7	34	2.57	0.23	2.22	2.36	2.53	2.75	3.05	78.3	11.6
8	34	4.02	0.61	3.06	3.48	3.87	4.62	5.30	116.0	17.3
9	28	13.59	13.27	5.47	7.50	8.44	12.49	66.00	304.5	45.3
90	3	5.51	0.04	5.47	5.48	5.48	5.52	5.57	14.6	2.2
91	3	6.20	0.36	5.93	5.94	5.95	6.33	6.71	18.6	2.8
92	3	7.53	0.05	7.44	7.48	7.52	7.55	7.57	15.1	2.2
93	3	8.06	0.02	8.04	8.05	8.06	8.08	8.11	16.9	2.5
94	3	8.33	0.17	8.18	8.24	8.29	8.44	8.59	19.3	2.9
95	3	10.47	0.41	9.74	10.06	10.39	10.66	10.93	29.0	4.3
96	3	12.02	0.29	11.58	11.84	12.10	12.20	12.31	30.0	4.5
97	3	17.11	3.13	13.03	14.69	16.35	18.79	21.23	39.4	5.9
98	3	31.38	2.69	27.47	28.55	29.64	31.65	33.66	55.5	8.3
99	1	66.00	0.00	66.00	66.00	66.00	66.00	66.00	66.0	9.8
TOTAL	334	2.42	5.09	0.04	0.55	1.17	2.46	66.00	671.9	100.0

If the practitioner decides that capping is appropriate, one could reduce the high values in each percentile exceeding 10% of the total metal to the maximum value of the next lowest percentile. For example, in Table 2.1, all values in P99 could be reset to the maximum value of 33.66 in P98. However, P98 contains 8.3% of the total metal, suggesting further capping may be required based

on practitioner judgment.

Cutting curves (Leuangthong & Nowak, 2015; Roscoe, 1996) is another simple tool to assess the average capped grade versus the capping limit. The idea is that the average grade will eventually stabilize as the threshold increases and a reasonable capping limit exists near the inflection point. Figure 2.2 (left) shows an example of a cutting curve for a synthetic positively skewed data set. The proposed capping limit is from the decile analysis.

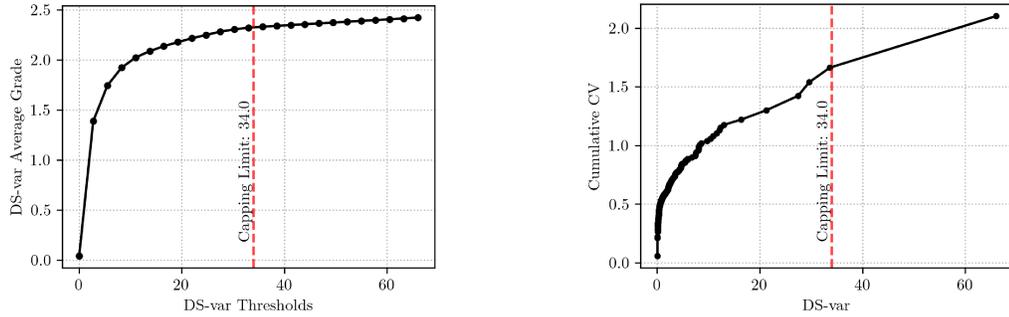


Figure 2.2: Cutting curve showing the relationship between the top cut grade and the mean of the adjusted data (left) and cumulative CV versus sample grade for a lognormal-like dataset (right). The cumulative CV plot shows a rapid increase at approximately $CV=1.25$. The capping limit corresponds to the value determined with the Parrish method.

Parker (1991) proposed plotting the cumulative CV against the grade distribution to identify a point where the influence of high values in the upper tail becomes strong. The data are sorted in descending order, and the cumulative CV is calculated considering all samples with grades less than or equal to the current sample. A rapid increase in cumulative CV characterizes this point where the influence of the upper tail is significant. Parker (1991) then breaks the distribution into two parts and fits the upper distribution with a truncated lognormal distribution; estimation is performed without capping the distribution. Figure 2.2 (right) shows an example cumulative CV plot of a synthetic log-normal-like data set.

The influence of suspected outlier values can be investigated by assessing uncertainty in the average grade of the domain through the spatial bootstrap (Solow, 1985). The procedure proposed by Nowak, Leuangthong, et al. (2013) involves bootstrapping the grade distribution while leaving out some percentage of the highest data (arbitrarily 2%) and comparing the expected mean to the uncapped mean. If the expected mean with the top 2% of samples removed is significantly lower than the uncapped mean, capping may be warranted within the domain. Figure 2.3 shows an example of 1000 bootstrapped means capped at P98 of the data distribution and the declustered uncapped mean (left) and the bootstrapped distributions for reference (right). The difference between the expected and declustered actual mean suggests that the grade distribution within the domain is sensitive to high values in the upper tail. If the expected and uncapped means are similar, this suggests that the upper tail of the distribution has little influence, and capping may not be necessary.

Metal-at-risk, summarized in Parker (2006), is a procedure that establishes uncertainty related

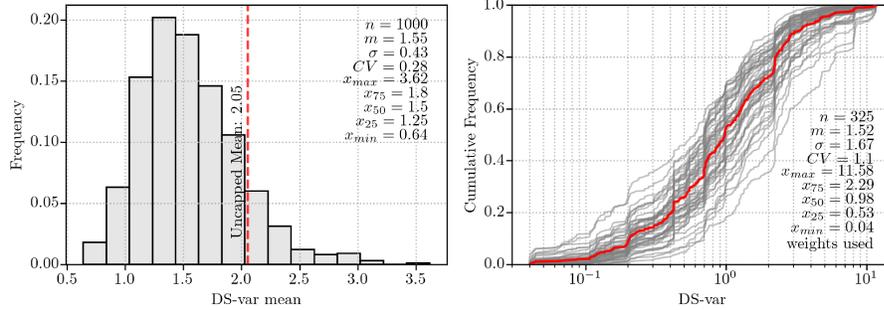


Figure 2.3: Distribution of 1000 bootstrapped mean values with the top 2% of values removed (left), and the CDFs of 50 of those realizations (right).

to the amount of high-grade metal within annual or global production volumes. The deposit can be “re-drilled” utilizing Monte Carlo simulation, where the total number of samples drawn equals the annual production tonnage divided by the number of tonnes per assay in the domain. The assay distribution is sampled randomly with replacement, assuming that high-grade can occur anywhere in the domain. The total metal above a given high-grade cutoff is calculated for each resampled realization. The P20 value of this distribution is added to the metal content below the cutoff to calculate a “risk-adjusted” metal content. Parker (2006) states that theoretically, the mine should exceed the risk-adjusted metal in four out of five periods; however, “there is additional and largely unquantifiable uncertainty related to the representivity of the sample-grade frequency distribution input to the simulation”. Metal-at-risk can be used as a guide to restrict the influence of high-grade samples during estimation or to calibrate a capping limit that removes this metal content. An advantage of the metal-at-risk approach versus others is that it accounts for data density and production volume; as data density increases, metal-at-risk decreases, and larger production volumes have less risk than small ones.

P-grams are a spatial statistic to characterize the continuity of high grades: the average probability that the tail and head of a lag vector \mathbf{h} are both above a cutoff grade (Leuangthong & Nowak, 2015; Nowak & Leuangthong, 2019). The p-gram value for each lag is the ratio of the number of pairs where both ends of \mathbf{h} are above the cutoff to the number where only the tail is above the cutoff. This average probability decreases as the lag distance increases. The p-gram can highlight a range of continuity for a specified cutoff value; this range guides restrictions to the search parameters for high-grade samples during estimation.

Regardless of the methodology, restricting extreme values aims to mitigate downside risk in resource estimates. The following section presents a survey of outlier management strategies published in publicly available technical reports. The survey provides an overview of current practices in the mining industry.

2.2.2 Current State of the Art

This section summarizes 125 NI 43-101 reports published by companies traded on Canadian securities exchanges between May 2019 and May 2021. For each deposit, the decision to cap or not, the capping methodology if so, and the decision to cap either assays or composites is recorded. For producing operations, whether underground or open-pit, is also recorded. Numerous reports documented multiple outlier management strategies. Due to many possible combinations of strategies, each is recorded separately. If one project uses cumulative probability plots and decile analysis, it contributes to the total methodology fraction twice. The left panel of Figure 2.4 shows the results for capping methodology, and the right panel shows a summary of assays versus composites for those who capped. Unknown refers to reports that stated a grade cap was applied, though it provides no methodology.

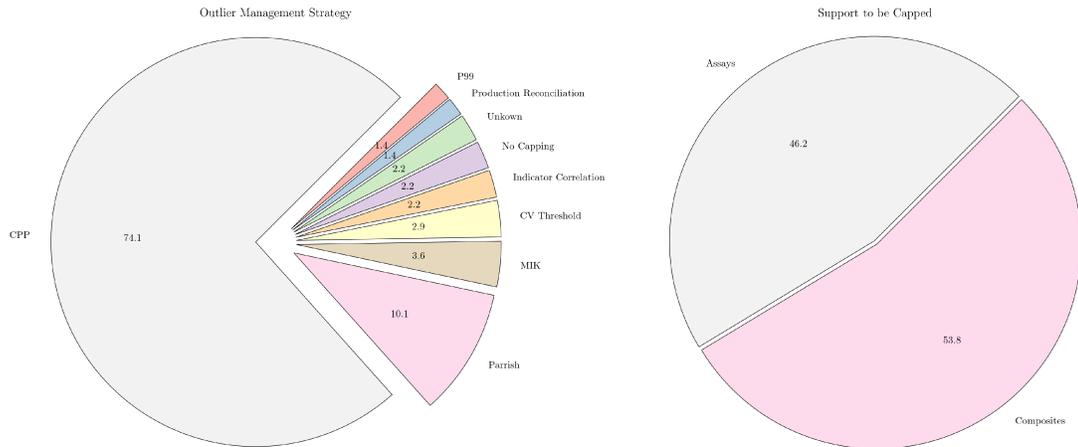


Figure 2.4: Summary of capping methodology (left) and, if capped, the data support to which the cap is applied (right) from 125 NI 43-101 reports. CPP - cumulative probability plot; P99 - 99th percentile; CV - coefficient of variation; MIK - multiple indicator kriging; Parrish - decile analysis after Parrish (1997).

The most common method is the analysis of CPPs. Practitioners look for inflection points to identify multiple populations or look for a point where the upper tail “degrades”. The choice of threshold is necessarily subjective, though gaps and inflection points guide selection. The second most common methodology is decile analysis developed by Parrish (1997).

MIK is a risk-qualified estimation methodology that can manage highly skewed distributions without needing grade capping (Journel, 1983). The non-linear transform of the original variable to indicators reduces the influence of extreme values in the upper tail, and practitioners suggest this removes the need for explicit grade capping (Cardinal Resources, 2019; Nevada Gold Mines LLC, 2020; Pretium Resources Inc., 2020; TriStar Gold Inc., 2021). Pretium Resources Inc. (2020) describe the logic of not applying a grade cap as:

“The positive tail of the grade distribution does not break down (tail decomposition method) until well into the multi-kilogram per tonne range, and even then, the more data that is collected, the higher the value before tail decomposition. Using a percentile-based approach results in an arbitrary and unjustifiable capping of extreme gold grades.” (pg. 14-11)

Though MIK is more robust concerning extreme values than traditional kriging algorithms, Carvalho and Deutsch (2017) suggests outlier values should still be managed in the usual (industry best practices) way. Artemis Gold Inc. (2020) is an example of a project where both an explicit grade cap (from CPP analysis) and MIK are employed.

A small proportion of the projects report no grade capping (Eldorado Gold Corporation, 2020; Medgold Resources Corp., 2021; Pasofino Gold Ltd., 2020). The justification of no grade capping is based primarily on a relatively low coefficient of variation and no “tail decomposition” in the CPPs.

Indicator correlation (analogous to p-grams) is another commonly employed practice. This approach considers the degradation of the spatial correlation of grades above a threshold. For many increasing thresholds, the spatial correlation decreases. The sill of the indicator variogram identifies the range of high-grade influence. This approach was employed at the Cariboo Gold Project (Osisko Gold Royalties Ltd, 2020) where the range corresponding to 99% of the indicator variogram sill is the maximum range of influence.

The remaining outlier management strategies are relatively straightforward. Some projects utilize an experience-based CV or percentile threshold. The CV threshold, commonly 2.0, selects a grade cap such that the remaining population has a CV equal to or less than the threshold. A percentile threshold, commonly the 98th or 99th, sets all grade values greater than $F^{-1}(0.99)$ to that value. Production reconciliation involves an iterative estimation process with a range of grade caps. The estimated metal content is reconciled to available production data for each capping threshold. The threshold that reconciles best is selected. Reconciliation is a reasonable approach, but it assumes that past production is characteristic of future production.

Another point of general indecision is whether to cap before or after compositing. In this survey, it is more or less a 50-50 split. One approach is not more correct than the other. The author generally believes it is more appropriate to apply a capping grade after assays have been brought to the same support. It seems logical to compare values that are effectively equally weighted. If much of the assay data is the same length, one could cap before compositing.

2.3 Spatial Outlier Detection

Most outlier detection methodologies employed in the mining industry neglect the spatial component of outliers and focus solely on the univariate distribution (quantiles, decile analysis, CPPs). Assessing the univariate distribution is considered a density-based approach; if a statistical model is

fit to the empirical distribution, samples in the tails of the distribution have low density relative to the underlying PDF. Inliers exist in high-density regions and outliers in low-density regions (Géron, 2019). While this is a justifiable line of reasoning, one must also consider the local spatial component of outliers. An extreme value located in a neighbourhood of other high values may be an outlier from a density perspective but not in the context of its local spatial arrangement.

Additionally, one must understand the volume of influence of a potential outlier. Isolated extreme values, or ones with large areas of influence, pose a risk of overestimation (Leuangthong & Nowak, 2015). The spatial context of a sample motivates the development of an outlier identification algorithm that considers both the spatial neighbourhood of the sample and its position within the global distribution. The spatial dimension characterizes the relationships within the local neighbourhood, while the density dimension characterizes the global relationship with the remaining data. The following sections describe the proposed algorithm components and provides examples of its use first on a synthetic 1-D example and then on a real two-dimensional (2-D) dataset.

2.3.1 Methodology

Consider a dataset $\{z(\mathbf{u}_i), i = 1, \dots, n\}$ where n is the total number of data and \mathbf{u}_i is a coordinate vector at the i^{th} location. About location \mathbf{u}_i , there exists a neighbourhood of k nearest locations, $NN(\mathbf{u}_i; k)$. This neighbourhood could be defined by a fixed search radius and maximum k or a fixed number k . The data $\{z(\mathbf{u}_j), j = 1, \dots, k, j \neq i\}$ define the neighbourhood of samples about location \mathbf{u}_i . Next, consider a function $m(\mathbf{u}_i)$ that returns a summary statistic, such as the mean or median, for all data values within the neighbourhood $NN(\mathbf{u}_i; k)$. The spatial component of the algorithm compares the data value $z(\mathbf{u}_i)$ to the value returned from $m(\mathbf{u}_i)$: $h(\mathbf{u}_i) = |z(\mathbf{u}_i) - m(\mathbf{u}_i)|$. The function $m(\mathbf{u}_i)$ is chosen to be the median value of all samples in the neighbourhood (excluding location \mathbf{u}_i), weighted by distance from \mathbf{u}_i . The median is chosen as it is a more robust measure of central tendency in the presence of outliers than the mean. The vector $\{h_1, h_2, \dots, h_n\}$ contains the absolute differences between each data value and the median of the surrounding neighbourhood. The h_n values are scaled $\in [0, 1]$ where values closer to 1 differ most from their neighbourhood. This comparison accounts for the local spatial relationship between data values. The choice of k is problem-specific. If k is too small, the median values will be noisy and not represent the true local variation. If k is too large, the median values will be smooth and not representative of the true local variation.

The neighbourhood $NN(\mathbf{u}_i; k)$ accounts for the area of influence of each sample. The area (in 2-D) or volume (in 3-D) of influence is calculated using the maximum distance to any sample in the neighbourhood (r_{max} in Figure 2.5):

$$A_{2D}(\mathbf{u}_i) = \pi \cdot r_{max}(\mathbf{u}_i)^2 \quad (2.2)$$

$$A_{3D}(\mathbf{u}_i) = 4 \cdot \pi \cdot r_{max}(\mathbf{u}_i)^2 \quad (2.3)$$

This measure is only applicable if $NN(\mathbf{u}_i; k)$ considers a fixed number k rather than a fixed search radius. A fixed search radius amounts to equal area weighting of each sample. The area directly accounts for the sparseness of the data configuration within the neighbourhood. The anisotropic area values are scaled $\in [0, 1]$. Samples located in sparse regions have values closer to one. A sample being geographically isolated does not constitute an outlier, and thus, one must consider the area of influence in conjunction with $h(\mathbf{u}_i)$.

The spatial neighbourhood approach is similar to the ‘‘Median Algorithm’’ proposed by Chen, Lu, et al. (2008); however, some key differences exist. Firstly, the neighbourhood $NN(\mathbf{u}_i; k)$ is determined using covariance-based distances rather than Euclidean distance. The covariance distance comes from the anisotropy ratios of a robust measure of spatial correlation. As the traditional semi-variogram is sensitive to the presence of outliers, one should use robust measures of correlation such as the correlogram, normal score variogram, or pairwise relative variogram (Babakhani, 2014; Drumond, Rolo, et al., 2019). Covariance-based distance ensures that the local neighbourhoods align with relevant geologic features. Secondly, the function $f(\mathbf{u}_i)$ is a weighted statistic, incorporating information about the data configuration and sparseness. The weight given to each sample in the neighbourhood is $w(\mathbf{u}_j) = \frac{1}{d(\mathbf{u}_i, \mathbf{u}_j)^p}$ ($d(i, j)$ in Figure 2.5). Accounting for the area of influence further incorporates spatial information.

The second component of the algorithm considers the relationship of each data value within the global distribution. This relationship is quantified by fitting a GMM to the univariate distribution to approximate the underlying PDF. GMM models are commonly used for outlier or anomaly detection (Géron, 2019; Qu, Du, et al., 2021), where data values falling in low-density regions of the fitted GMM are potential outliers. The details of fitting the GMM with the expectation-maximization (EM) algorithm are not given here; the reader is referred to McLachlan, Lee, et al. (2019) for more details. After the GMM is fit to the univariate data, estimating the density at any location is straightforward. The log of the PDF is calculated as:

$$\log p(z_i) = \log \left(\sum_{j=1}^J \pi_j \mathcal{N}(z_i | \mu_j, \Sigma_j) \right) \quad (2.4)$$

where z_i is the i^{th} sample of z , J is the number of fitted GMM components, π_j is the weight to the j^{th} component, and $\mathcal{N}(z_i | \mu_j, \Sigma_j)$ is the PDF of a multivariate Gaussian distribution with mean μ_j and covariance Σ_j . Exponentiation of Equation 2.4 results in an estimate of the PDF for each sample. The higher the probability density, the more likely the sample belongs to the fitted distribution. As this calculation results in an estimate of the PDF and not a true probability, the values are scaled to sum to one. The vector $\{p_1, p_2, \dots, p_n\}$ then contains an estimate of the probability that the i^{th} sample belongs to the fitted distribution. Low-probability samples that come from low-density regions are possible outliers. This result depends on the number of components, J , which must be chosen. Practice shows that 2-3 components are reasonable for a 1-D distribution,

and fitting is generally straightforward. The final outlier measure is then a weighted combination of h_i , A_i , and p_i for each data value:

$$g_i = w_h * h_i + w_A * A_i + w_p * (1 - p_i), \quad i = 1, \dots, n \quad (2.5)$$

where w_h , w_A , and w_p weights to the neighbourhood, radius and density components, respectively, and $w_h + w_A + w_p = 1.0$. Figure 2.5 shows a sketch of the local and global components to calculate a final outlier measure. The first two components of g_i come from the spatial neighbourhood (1) in Figure 2.5 while the third component comes from the fitted probability density (2). Thresholding the final outlier measure in Equation 2.5 results in the binary classification of each sample as an inlier or outlier.

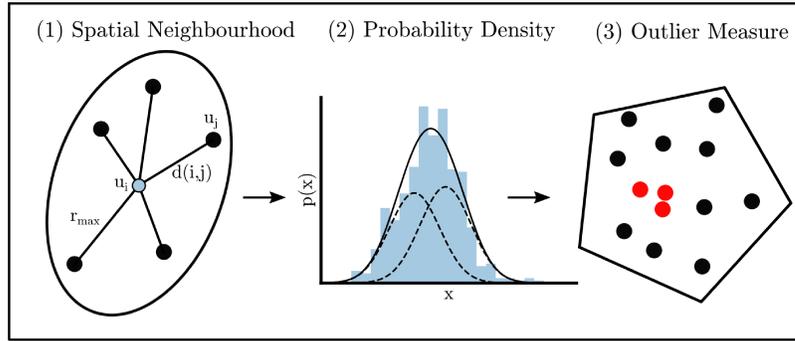


Figure 2.5: A sketch illustrating the spatial and density components of the spatial outlier algorithm. (1) shows an example search neighbourhood with $k = 5$ where $d(i, j)$ is the distance between location \mathbf{u}_i and \mathbf{u}_j , and r_{max} is the distance to the furthest neighbour. (2) shows the PDF of all samples fitted by a GMM with $J = 2$ components. (3) shows an example domain with samples classified as inliers (black) and outliers (red); thresholding the outlier measure in Equation 2.5 results in the binary classification of each sample.

Consider the 1-D synthetic data in Figure 2.6a. The data show cyclicity and a trend with increasing x , which are common properties of earth science data. Random noise drawn from a Gaussian distribution ($\mu = 1.0$, $\sigma = 3.0$) is added to ten samples in the sequence to simulate the presence of outliers. The degree of “outlierness” varies across the samples. A variogram model is fit to the original data without noise for a robust measure of spatial continuity. Access to the true data values is not possible with real data; alternatively, a correlogram or pairwise relative variogram could be fit to the noisy data. The spatial neighbourhood considers $k = 10$ data after scaling the coordinates according to the fitted variogram model.

Figure 2.6b shows the difference between the data and weighted neighbourhood median values. The red x’s indicate the data with added noise. Except for one noisy value, the suspected outliers show moderate to significant differences with their neighbourhoods. Figure 2.6c shows the $J = 3$ GMM components fit to the univariate distribution $f(x)$. The solid line is the fitted model, and the dashed lines are the individual components. The number of components is chosen based on trial and error to minimize false positives. Again, this is not possible with real data, but this process allows exploration of the bounds of reasonable parameters. Figure 2.6d shows the outlier

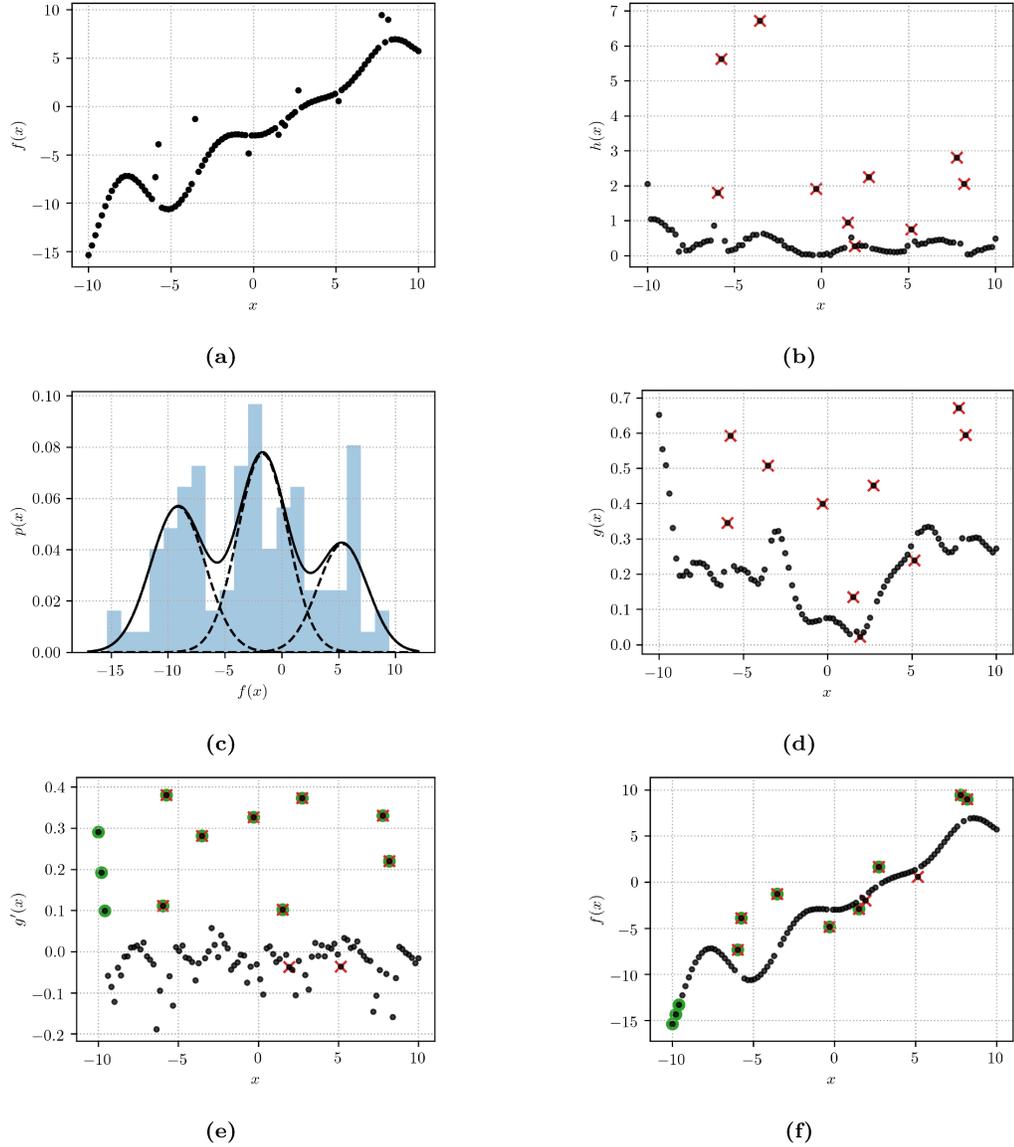


Figure 2.6: (a) 1-D function of the coordinate position x ; (b) the difference between each sample the neighbourhood median, $h(x)$; (c) fitted GMM model with $J = 3$ components; (d) the combined outlier measure $g(x)$; (e) detrended $g(x)$ measure with flagged outliers; (f) the 1-D function from (a) with outliers identified. The red x's denote the samples with added noise, and the green circles are the samples classified as outliers by the algorithm.

measure of Equation 2.5 considering $w_h = 0.5$, $w_A = 0.0$ and $w_p = 0.5$. $w_A = 0.0$ as all samples are equidistant and thus equally weighted. Values closer to one are more likely to be outliers. Compared to Figure 2.6b, the algorithm identifies additional values (near $x = -10$) as discordant with the rest of the population. Removing the trend from $g(x)$ using a moving window average (MWA) further elucidates the potential division between inliers and outliers. Figure 2.6e shows the detrended distribution $g'(x)$. At this point, a threshold must be selected to define the division. The green circles in Figure 2.6e identify all data above the threshold of $g'(x) = 0.08$. Figure 2.6f shows the final outliers identified by the algorithm (green circles) with the original data configuration.

The algorithm effectively identifies the noisy data points. Two points remain undetected, though the magnitude of the noise is small, and there are no drastic intra-neighbourhood changes in these areas. In addition to the noisy data, the algorithm identifies three additional outliers near $x = -10$. It is possible that a traditional univariate method such as visual examination of a CPP would identify outliers in the upper and lower tails of the distribution (i.e. the values flagged in the lower left and upper right corners of Figure 2.6f); however, the outliers in the middle of the distribution would go undetected without considering the spatial component.

2.3.2 Application

Consider the 2-D spatial distribution of platinum group elements (PGE) in Figure 2.7. There are scattered high values throughout the northwest portion of the domain where the data density is highest; there are also isolated high values in the more sparse regions to the south. These samples are potential outliers in a spatial context. Figure 2.8 shows the CDF and CPP for the PGE distribution. The empirical CDF shows a degree of tail decomposition for high values - this is evident in the CPP. Tukey's fences (Tukey, 1977) identify potential outliers in both distribution tails.

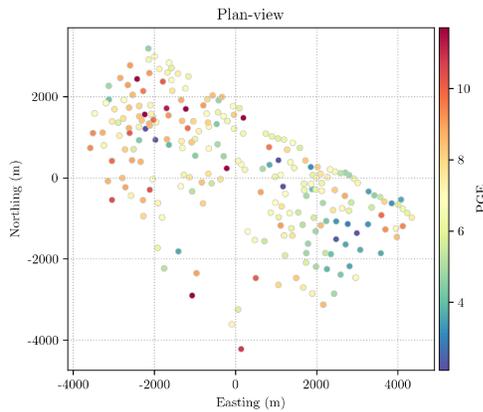


Figure 2.7: Location map of PGE sample locations.

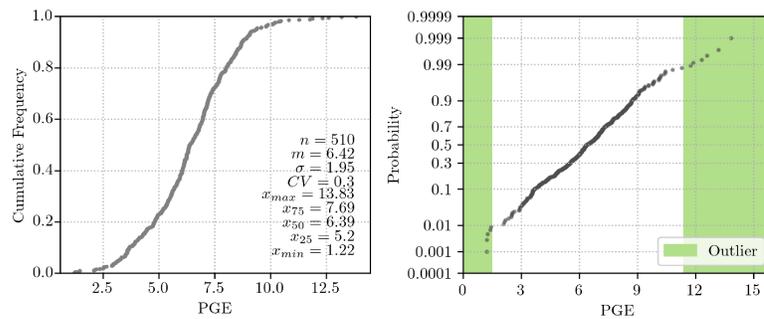


Figure 2.8: CDF plot and CPP for the PGE distribution. The green shaded area in the CPP are Tukey's fences (Tukey, 1977), identifying potential outliers.

The methodology from Section 2.3.1 is employed to identify spatial outliers. An experimental correlogram is calculated and fitted with an exponential model to characterize the spatial relationship between sample locations. The neighbourhood $NN(\mathbf{u}_i; k)$ is characterized by $k = 25$ neighbours and an inverse distance weighting exponent of $p = 1.0$. The area of influence for each sample is calculated in 2-D (Equation 2.3). The GMM is fitted to the histogram using $J = 2$ components. The final outlier measure is calculated using $w_h = 0.4$, $w_A = 0.2$, and $w_p = 0.4$. The weight to the area of influence is less than the other measures to not overly weight all sparse samples, or samples near the edges of the domain.

The top row of Figure 2.9 shows the absolute difference between each sample value and its corresponding neighbourhood. The left column shows values along the east-west direction, and the right column shows the north-south direction. There are some clear outlying samples, though $h(\mathbf{u}_i)$ does not show a clear delineation between two populations. The second row of Figure 2.9 shows each sample’s area of influence measure. Row three shows $1 - p(\mathbf{u}_i)$, where values closer to one are in lower-density regions of the distribution. The fourth row shows the final outlier measure. The division between inliers and outliers is not immediately clear, but one could argue there is a grouping of points above the $g(\mathbf{u}_i) = 0.6$ threshold, highlighted by the red dashed lines.

Figure 2.10 shows the identified outliers in the spatial context. The algorithm identifies high and low values: ten above the mean and three below. Potential outliers must be aligned with geologic intuition. There is a clustering of extreme value outliers in the northern part of the domain, where neighbouring samples are medium-grade. An extreme low-grade sample adjacent to an extreme high-grade sample is also flagged. Given the position of each sample in the CDF and spatial arrangement, these samples appear to be appropriate candidates for outliers. Three high-grade samples in the sparsely sampled southern portion of the domain are also flagged as outliers. These samples are also appropriate outlier candidates due to their high grades relative to adjacent samples and their potential influence area for each sample. In an estimation paradigm, these samples could influence the grades in many blocks and lead to overestimation. One could derive a capping limit as the minimum grade of the high-grade samples flagged as outliers.

The spatial outlier detection algorithm effectively identifies potential outliers. Measures of “outlierness” in spatially correlated data must consider both the univariate distribution and the spatial context of the data. The proposed algorithm directly incorporates features of the grade distribution, robust measures of geologic distance, and features of the data configuration, including sparseness. The final outlier measure is practical in the sense that it includes multiple sources of relevant information, being a weighted blend of a proximity-based algorithm and a parametric statistical model (GMM) (Li, Zhao, et al., 2022). The algorithm could be extended to the multivariate context by considering the Mahalanobis distance between a K dimensional vector of medians and the center of the data (Chen, Lu, et al., 2008). A challenge of the outlier detection algorithm is that an arbitrary threshold must be chosen. The final measure may not have a clear demarcation between inlier and

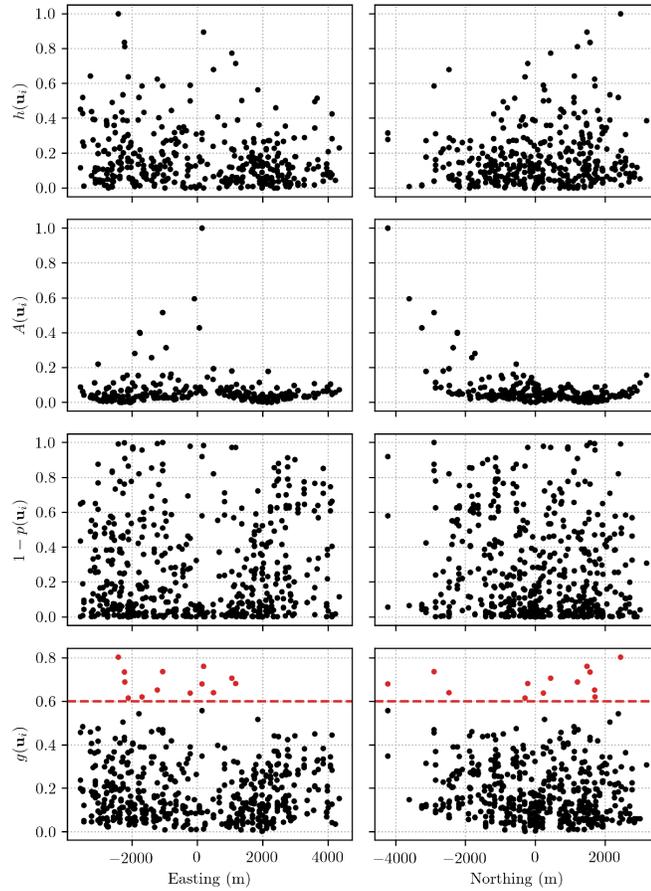


Figure 2.9: Absolute difference between the sample value and corresponding neighbourhood (1^{st} row); area of influence for each sample (2^{nd} row); density measure (3^{rd} row); final outlier measure (4^{th} row) with flagged outliers above the $g(u_i) = 0.6$ threshold. The left column shows values in the east-west direction, and the right column shows the north-south direction.

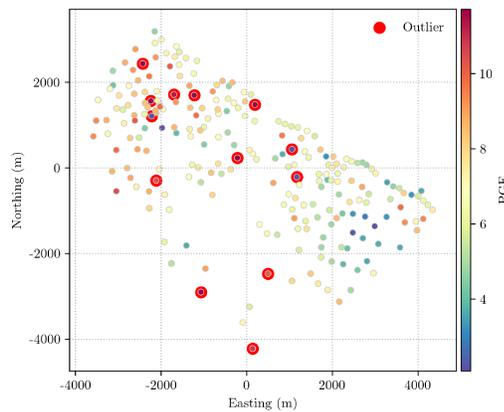


Figure 2.10: Location map of PGE sample locations with flagged outliers.

outlier. Using additional outlier detection measures, such as CPP analysis, as an ensemble may aid threshold selection. For instance, the outliers identified by Tukey’s fences in Figure 2.8 roughly correspond to the spatial outliers, suggesting appropriate parameters. The identified outliers also

appear “reasonable” from a geologic perspective. Immediately adjacent extreme high and low values warrant further investigation. One could employ an unsupervised clustering method on the final outlier measure to remove subjectivity from the choice of threshold. Outlier detection methodologies should generally form an ensemble, with higher confidence given to samples identified by multiple methods (Zimek & Filzmoser, 2018). The spatial outlier detection algorithm is another tool in conjunction with those in Section 2.2.1 and provides further justification for the choice of capping limit or outlier management strategy.

2.4 Analytical Extreme Value Models

Many techniques have evolved to restrict the influence of outliers or extreme high grades. No statistical or geostatistical model exists to understand and manage the resource contributions of extreme high-grade (EHG). The validity of such a model could only be established based on many data or bulk mining. The idea is to develop a simple and intuitive model that accommodates the resource contributions of EHG. This model could be applied to understand and explain historical mining and to project the possibility of EHG mineralization in unmined areas.

2.4.1 Methodology

The geological processes that led to the precipitation and preservation of the grades under consideration in a particular deposit are complex and defy a simple deterministic assessment. The processes influence our understanding, but we adopt a statistical model since there is no way to understand the initial and boundary conditions of the non-linear and chaotic processes that led to the deposit under consideration. This section describes a trimodal model for mineralization: (1) mineralized (M), (2) high-grade (HG), and (3) extreme high-grade (EHG). Figure 2.11 illustrates this. The three populations overlap, mix and are not exclusive, but we could reasonably define a range that represents them, for example, 0.1 to 1.0 g/t for mineralized (M), 5 to 20 g/t for high-grade (HG), and 500+ g/t for EHG.

The concept of three populations is reasonable. The illustration in Figure 2.11 appears discrete as three populations; however, the data distribution from this model would appear highly skewed. Considering one highly skewed population may be possible; however, a flexible parametric distribution is not available to satisfy observed data, explain outliers, and avoid unrealistically high grades. Considering more than three populations would be possible; however, it seems reasonable to have M, HG and EHG. Additional intermediate populations would complicate the model and could be grouped into one of the three.

An assumption is that the M and HG are more pervasive while EHG is encountered rarely. However, the EHG is assumed to have some reasonable thickness within geologic structures. A drill hole intersecting M, HG or EHG would be identified as such. We do not expect many EHG intersections.

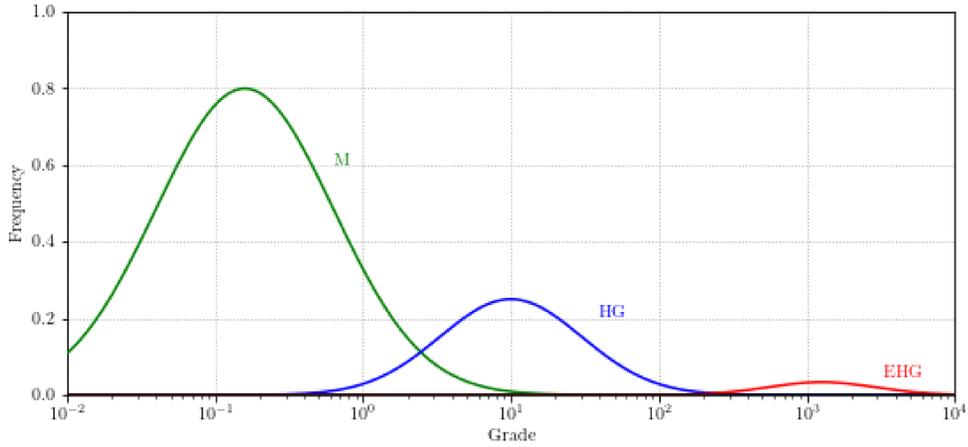


Figure 2.11: Trimodal distribution model of mineralized (M), high-grade (HG) and extreme high-grade (EHG).

The model is parameterized by the probability of each population (P_M , P_{HG} , and P_{EHG}) and three lognormal distributions defined by mean values (m_M , m_{HG} , and m_{EHG}) and variance or standard deviation parameters for each (σ_M , σ_{HG} , and σ_{EHG}). The sum of the proportions is one:

$$P_M + P_{HG} + P_{EHG} = 1$$

The overall mean is defined as:

$$m_{Overall} = P_M \cdot m_M + P_{HG} \cdot m_{HG} + P_{EHG} \cdot m_{EHG}$$

The variation of each distribution (M , HG , and EHG) is important and relevant, but the metal and resource are defined mainly by the proportions and mean values. There is a great challenge in inferring the parameters of this model. A key parameter is establishing the metal in the HG population versus the EHG population. A straightforward way to parameterize this is to assume that the metal in the EHG population is a fraction of that in the HG population. The fraction is essential to understand the probability of encountering extreme high grades and, ultimately, for spatial prediction. Historical mining or external information guides the fraction of the total metal from EHG . Considering this fractional model leads to the following:

$$P_{EHG} \cdot m_{EHG} = f \cdot P_{HG} \cdot m_{HG}$$

$$P_{EHG} = \frac{f \cdot P_{HG} \cdot m_{HG}}{m_{EHG}}$$

If there is enough data, the precise value of the fractional metal in EHG versus HG (the f parameter in the equations above) will be inferred from the data. If there are too few data, it could be assumed. For example, assuming $f = 1$ as a reasonable value, a sensitivity study could be performed.

Combining the equations above for the overall mean:

$$\begin{aligned}
 m_{Overall} &= P_M \cdot m_M + P_{HG} \cdot m_{HG} + P_{EHG} \cdot m_{EHG} \\
 &= \left(1 - P_{HG} - \frac{f \cdot P_{HG} \cdot m_{HG}}{m_{EHG}}\right) \cdot m_M + P_{HG} \cdot m_{HG} + \frac{f \cdot P_{HG} \cdot m_{HG}}{m_{EHG}} \cdot m_{EHG} \\
 &= m_M + P_{HG} \cdot \left(-m_M - \frac{f \cdot m_M}{m_{EHG}} + m_{HG} + f \cdot m_{HG}\right)
 \end{aligned}$$

The proportions of the populations are defined in sequence by the following:

$$P_{HG} = \frac{m_{Overall} - m_M}{-m_M \cdot \left(1 + \frac{f \cdot m_{HG}}{m_{EHG}}\right) + m_{HG}(1 + f)}$$

$$P_{EHG} = \frac{f \cdot P_{HG} \cdot m_{HG}}{m_{EHG}}$$

$$P_M = 1 - P_{HG} - P_{EHG}$$

The mean values of the three populations could be estimated with reasonable confidence. The overall mean could be estimated from historical mining. The fraction of metal in the EHG population relative to the HG population (the f value) is a model parameter that could be inferred from available drilling if enough intersections are available. Given the mean values and f , we could infer the proportions of the populations and the contribution to metal from each population.

2.4.2 Probability of Drilling EHG

The probability of drilling n successive drill holes without encountering EHG could be computed by:

$$(1 - P_{EHG})^n$$

This approach assumes the drill holes are independent, which may or may not be reasonable, given drillhole spacing. It also assumes the EHG will be seen in a drill hole with a significant thickness; that is, the EHG is not distributed in very small nuggets. This assumption is reasonable since if the EHG were at a very small scale, it would be composited with other rock and end up as mineralized (M) or high grade (HG).

Consider $m_{Overall} = 10g/t$, $m_M = 0.1g/t$, $m_{HG} = 10g/t$, $m_{EHG} = 1000g/t$, and $f = 1$ that is, there is the same metal in the HG and the EHG. These numbers appear reasonable given the intersections encountered at epithermal vein systems. Following the calculations described above:

$$P_{HG} = \frac{10 - 0.1}{-0.1 \cdot \left(1 + \frac{1 \cdot 10}{1000}\right) + 10(1 + 1)} = 0.4975$$

$$P_{EHG} = \frac{1 \cdot 0.4975 \cdot 10}{1000} = 0.0050$$

$$P_M = 1 - 0.4975 - 0.0050 = 0.4975$$

The overall mean of the model is checked: $m_{Overall} = 0.4975 \cdot 0.1 + 0.4975 \cdot 10 + 0.0050 \cdot 1000 = 10$ as it

must. So, for thirty ($n = 30$) drill holes there is an $0.995^{30} = 0.86 = 86\%$ chance of *not* intersecting EHG. To get to a 50% chance of encountering an EHG drill hole $\log(0.5)/\log(0.995) = 138$ drill holes would be required.

2.4.3 Application

The following demonstrates the conceptual trimodal model for extreme high grades applied to an actual data set. The data comprises 61,027 channel samples across a gold-bearing Witwatersrand reef structure with cumulative grade (cmg/t) and thickness (cm) measurements. Gold values (g/t) are back-calculated from the other two measurements. A significant proportion of the samples are considered “high grade”. The data set is sufficiently dense such that some *valid* extreme values are likely observed, and we can infer the parameters of the extreme high-grade model with reasonable confidence.

Figure 2.12 (left) shows the Au distribution. The overall distribution is high grade with multiple oz/t assays. Log probability plots are a useful tool for identifying multiple populations and visualizing the upper tail of the distribution. Figure 2.12 (right) shows the log probability plots for Au. Au plot shows a distinct inflection point near 0.1 g/t and tail decomposition around 30 g/t. These may be reasonable thresholds for the initial assessment of mineralized, high grade and extreme high grade.

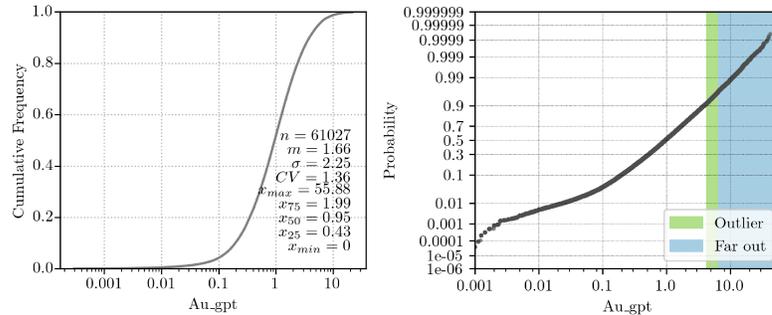


Figure 2.12: Au CDF with log scaling (left) and log probability plot with Tukey’s fences (Tukey, 1977) (right).

Parameterization of the extreme high-grade model requires four mean values (overall, mineralized, high grade and extreme high grade) and the fraction of extreme high-grade metal contributing to the overall high-grade metal. Thresholds must be selected to delineate sub-populations. A subsample of the full data set is used to calculate the proportions of each population. 80% of the data is withheld, and the distribution of the remaining 20% is used for inference to simulate a sparse sampling regime. Thresholds are defined based on distribution quantiles. Mineralized is defined by the 0.1-0.9 interquantile range, high grade from 0.95-0.9995, and extreme high grade 0.9999 and above. Though subjective, threshold selection is informed by log probability plots. The 0.1-

0.9 quantile range roughly defines the “inlier” range based on Tukey’s fences (Tukey, 1977). 0.95 roughly defines the break between “outlier” and “far outlier”. 0.9999 roughly defines the point of tail decomposition, suggesting extreme values. The thresholds consider the number of samples within each distribution; we do not expect to encounter many extreme values. Figure 2.13 shows a location plot of all samples with the corresponding EHG locations highlighted in red.

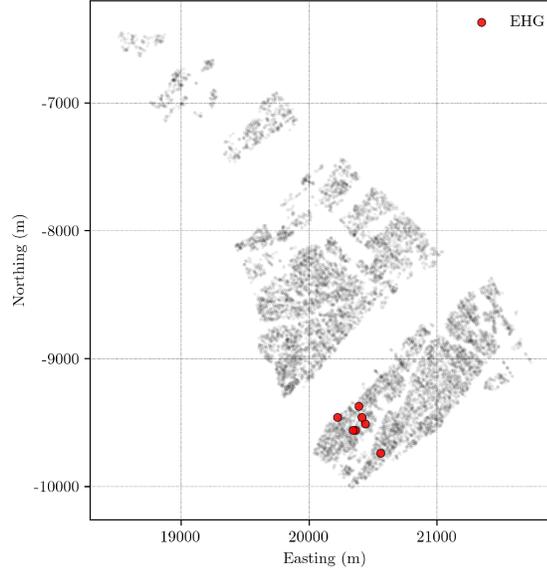


Figure 2.13: Location plot of all samples with EHG locations highlighted in red.

A sub-sample of the full data set is used to calculate the proportions of each population, summarized in Table 2.2. Table 2.3 summarizes the proportion of metal contributed by each population. The overall mean of the model can be checked:

$$\begin{aligned}
 m_{Overall} &= P_M \cdot m_M + P_{HG} \cdot m_{HG} + P_{EHG} \cdot m_{EHG} \\
 1.45 &= 0.9601 \cdot 1.15 + 0.0398 \cdot 8.55 + 0.0001 \cdot 48.35 \\
 1.45 &= 1.45
 \end{aligned}$$

In this scenario, the fractional EHG component, f , is quite small due to a significant number of samples in the HG population. The channel samples are narrow (tens of cm) and likely only sample mineralized material. If the samples were drill core, one would expect more internal dilution, lower grades, and a higher f value. Given the proportions, we can calculate the probability of sampling extreme high grade with n successive channel samples or the required number of samples for a P_n probability of sampling EHG. For example, the probability of *not* sampling EHG in 100 channel samples is $(1 - 0.0001)^{100} = 98.9\%$. In order to get a 50% chance of observing an EHG sample, $\log(0.5)/\log(1 - 0.0001) = 6301$ additional channels are required.

The reasonableness of this model can be checked with a simple simulation study. For a given

Table 2.2: Inferred and calculated model parameters.

	Value
Overall mean (g/t)	1.4500
Mineralized mean (g/t)	1.1500
High grade mean (g/t)	8.5500
Extreme high grade mean (g/t)	48.3500
f	0.0200
Mineralized proportion	0.9601
High grade proportion	0.0398
Extreme high grade proportion	0.0001

Table 2.3: Overall metal proportions by category.

	Proportion
Mineralized	0.7611
High grade	0.2352
Extreme high grade	0.0037

Table 2.4: Analytical versus simulated EHG probability.

EHG Prob.	Samples	Expected Prob.
0.1	957	0.1012
0.5	6301	0.4873
0.9	20931	0.8876

probability, say 0.1, 0.5 or 0.9, the number of additional samples required to intersect an EHG value with that probability is calculated analytically as above. High-resolution simulated realizations of gold using all available data are considered the truth. If we randomly sample the realizations with the calculated number of samples for some number of trials, we can directly observe how many EHG intersections occur. Ten realizations are used for numerical stability. Each realization is sampled 1000 times with the calculated number of samples. Table 2.4 summarizes the predicted number of samples required to have a 0.1, 0.5 and 0.9 probability of intersecting EHG and the corresponding expected probabilities from resampling. The simulation results closely reproduce the analytical predictions.

Access to a high-resolution “true” model is rarely possible in practice. Often, when data is sufficiently dense to be considered the truth, mining has already occurred. One does not have the luxury of calibrating their analytical model. Though applying the proposed analytical model is unverifiable in a practical scenario, the example presented highlights that the analytical model *could* be reasonable. The simulation study shows that the model can accurately predict the probability of intersecting EHG. Determining how much data is required to infer model parameters is a topic of future research.

2.5 Discussion

Outlier management is an important component of the traditional resource estimation paradigm, particularly with precious metals. Using smooth estimators, such as kriging, in the presence of unadjusted extreme values poses a risk of overestimation. High grades in sparse data configurations exacerbate this risk. The selection of grade caps is a long-standing issue; high-grade values are important from an economic perspective, though they are problematic to resource estimation. There is no definitive recipe for outlier management. There is a necessarily subjective threshold choice if one decides to cap. The capping strategies outlined in this chapter provide general guidance; however, no definitive metric exists. One must first assess that the extreme values are valid. If possible, one should address high grades with domain boundaries. A capping strategy may be considered if there is little continuity between high grades. Best practice suggests considering multiple strategies as an ensemble and determining a consistent grade cap across multiple methods.

Many outlier detection methodologies do not consider the spatial context of the samples. To overcome this, a spatial outlier detection algorithm is proposed, which considers the spatial configuration and the probability density of each sample in the distribution. The idea that outliers fall in low-density regions is only sometimes true. The context of each sample within its local neighbourhood should influence the decision to classify an outlier. If an extreme value is discordant within a geologically-driven neighbourhood, it should be flagged as an outlier. An extreme value near other high values is likely a true feature of the underlying distribution. High-grade samples with large areas of influence should also be flagged. The algorithm effectively identifies outliers in a real 2-D example with variable data density across the domain. Multiple features of the sample distribution are combined and projected into a feature space where the threshold selection may be more intuitive than directly selecting a grade value. High-grade samples flagged as outliers permit the inference of a capping limit: the minimum grade of the high-grade outliers. The algorithm provides the practitioner with another tool in the ensemble to guide and detect outliers.

An analytical extreme value model is developed to understand the contributions of EHG on model resources. The effect of extreme values is significant from an economic perspective. The EHG model allows the practitioner to predict the frequency of intersecting extreme values, which is powerful in the context of drillhole planning or designing data collection schemes. The analytical approach is limited in that historic mining or dense drilling is required to parameterize the model. However, a simulation study with the Witwatersrand data set shows the model can correctly predict the occurrence of EHG.

The remaining chapters in this thesis are dedicated to developing the NMR framework. The NMR explicitly accounts for the spatial features of extreme values through high-order connectivity metrics. If the high-order features are correct, the data does not require explicit capping or outlier management. For this reason, outlier management is not discussed beyond this chapter.

Chapter 3

Network Model of Regionalization Framework

This chapter introduces the NMR paradigm and the framework for generating non-Gaussian spatial fields. The main idea of the NMR is that capturing non-Gaussian spatial features requires reproducing high-order statistics. Reproduction of high-order statistics should also reproduce two-point statistics such as the variogram. Multi-point connectivity measures can be easily calculated from drillholes by considering them as sequences. The NMR is ultimately a function that maps an unknown latent space to an interpretable observed space. The mapping of the latent space generates a non-Gaussian spatial distribution with the correct high-order statistics. Inference of this mapping function is an inverse problem as we only observe the system's output. The chapter begins by framing the problem setting and addressing known issues with the multivariate Gaussian assumption. Next, the relationship between connectivity and non-Gaussianity is discussed. Multi-point measures of spatial connectivity, like distributions of runs, are a core component of the NMR framework. Connectivity measures are critical in the context of extreme values; the spatial arrangement of these values is likely significant concerning a transfer function. The connectivity of extremes likely drives the project economics in many mining scenarios. Finally, an overview of the NMR methodology is presented with a 2-D synthetic example highlighting improved resources relative to a traditional SGS model.

3.1 Problem Overview

Geostatistical problems are often high dimensional, considering multiple variables at millions of locations. The appeal of the Gaussian distribution is its mathematical tractability in any dimensions, where a mean vector and variance-covariance matrix fully parameterizes it; thus, it is pervasive in geostatistics. Many algorithms take advantage of the fact that under the multivariate Gaussian assumption, all conditional distributions are Gaussian and calculated by linear combinations of the conditioning data. Multivariate geostatistical problems necessitate a parametric distribution as there are typically only hundreds to thousands of data available. The curse of dimensionality (Bellman, 1961) precludes the use of non-parametric distributions.

Connectivity of extreme values is commonly discussed as a shortcoming of the Gaussian RF model (Guthke & Bárdossy, 2017; Journel & Alabert, 1989; Journel & Deutsch, 1993; Kerrou, Renard, et al., 2008; Yan, Jeong, et al., 2020). The maximum entropy characteristic of the Gaussian RF model leads to maximum disorder for a given covariance structure; it does not allow for spatial correlation

or connectivity of extreme quantile indicators (Kerrou, Renard, et al., 2008). Figure 3.1 emphasizes these disconnected extreme values. The left panel is a reference image showing the volume of shale for interbedded sands and shale of the McMurray Formation. The middle panel shows a Gaussian realization generated with the normal score variogram calculated from the reference. The right panel shows the connection probability for subsequent steps in the vertical direction. The Gaussian realization is more disorganized and shows less connectivity than the true image. Furthermore, the nature of the bivariate Gaussian relationship leads to symmetric destructuring of the indicator variogram about the median. In practice, natural non-Gaussian distributions show asymmetric destructuring of indicator variograms (Vincent, 2021) and correlation need not approach zero as z_k approaches an extreme quantile (Journel & Alabert, 1989).

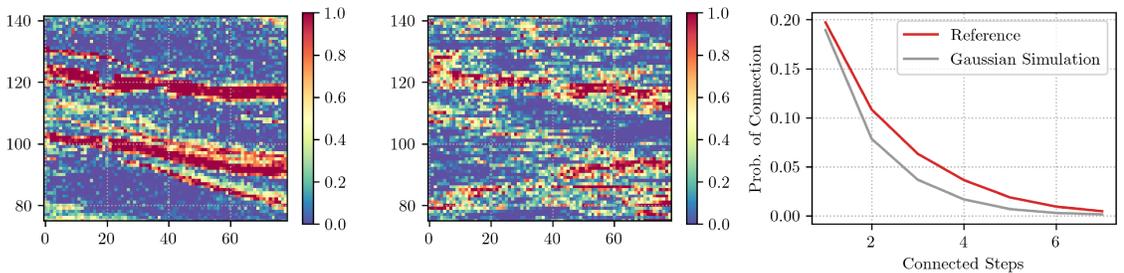


Figure 3.1: Reference image showing the volume of shale for interbedded sands and shale of the McMurray Formation (left), a Gaussian realization generated with SGS (middle), and the corresponding n -point connectivity functions (right). The reference image is part of a core photograph.

High-order connectivity metrics can characterize non-Gaussianity. This concept is explored in detail in Section 3.2. These multi-point measures include the frequencies of runs and the n -point connectivity function. As properties of the Gaussian RF model lead to maximum disorder, one can use a connectivity measure to assess non-Gaussianity from sequences. The idea is that a Gaussian RV would show different connectivity than a non-Gaussian RV. The multi-point spatial arrangement would differ depending on the underlying RV. RVs with more connectivity, will have fewer but longer runs. This connectivity concept applies to sequences with structured extreme values not suited by a Gaussian distribution. There is some degree of connectivity of extremes in mineral systems; the underlying spatial phenomena are not entirely disorganized.

The challenges associated with MIK and traditional outlier management motivate the need for spatial models that can characterize the non-Gaussian continuity of extreme values. Though the goal is to achieve non-Gaussian features, the NMR framework remains based on Gaussian distributions. The idea is to extend the concept of the linear model of regionalization (LMR) and map a collection of regionalized, independent, standard normal factors $\{Y_m, m = 0, \dots, M\}$ through a weighted network consisting of non-linear activations. The NMR framework poses aspects of the generation of stochastic realizations as an optimization problem rather than the traditional random function approach. The parameters that map the factors are unknown, though we have observed

measurements at the data locations. This model-based inverse problem is approached through stochastic optimization. This approach allows the incorporation of multi-point geologic information from various sources into the final model while honouring the statistics of the observed data. A network-based mixing architecture and non-linearities allow the final models to capture richer and higher-order spatial features; this is particularly useful when constructing models in the presence of extreme values.

The following components summarize the overall structure of the NMR framework:

1. Objective targets: specifying the two- and multi-point statistics for the final model to reproduce.
2. Latent factor design: selection of the base pool of Gaussian distributions for mixing.
3. Parameter inference: determining the parameters of the mixing function that result in a distribution with the correct spatial features.
4. Latent factor imputation: imputation of the latent factors such that mixing reproduces the observed data values.
5. Latent factor simulation and mapping: conditional simulation of the imputed latent factors from (4) and mapping to observed space with parameters from (3).

The first component is specifying the goals of the model. These goals, or objectives, are quantified by two-point spatial statistics and higher-order statistics. The second component involves choosing the covariance structure of each latent factor to mix. These choices depend strongly on the modeling goals. The third component involves inferring the parameters of the mixing function, which results in the model meeting the objective targets. The fourth component involves generating synthetic realizations of the factors. The fifth component is conditionally simulating the factors on a modeling grid and mapping them to observed space. This mapping results in gridded realizations that reproduce the observed data and the objective targets away from the data. Figure 3.2 shows the complete NMR framework as a flow chart. There are three junctions in the workflow where the practitioner must decide if there is acceptable reproduction of desired statistics. Given the interconnected nature of the framework, if acceptable reproduction is not achieved, one may have to revisit one or multiple previous steps. This interconnectedness is highlighted by reverse-flow arrows leading to multiple actions in Figure 3.2.

The remaining sections of this chapter discuss the calculation of high-order statistics and their use for measures of non-Gaussianity, details regarding the components of the NMR framework, and finally, a small synthetic example of the complete NMR workflow with highly structured extreme values.

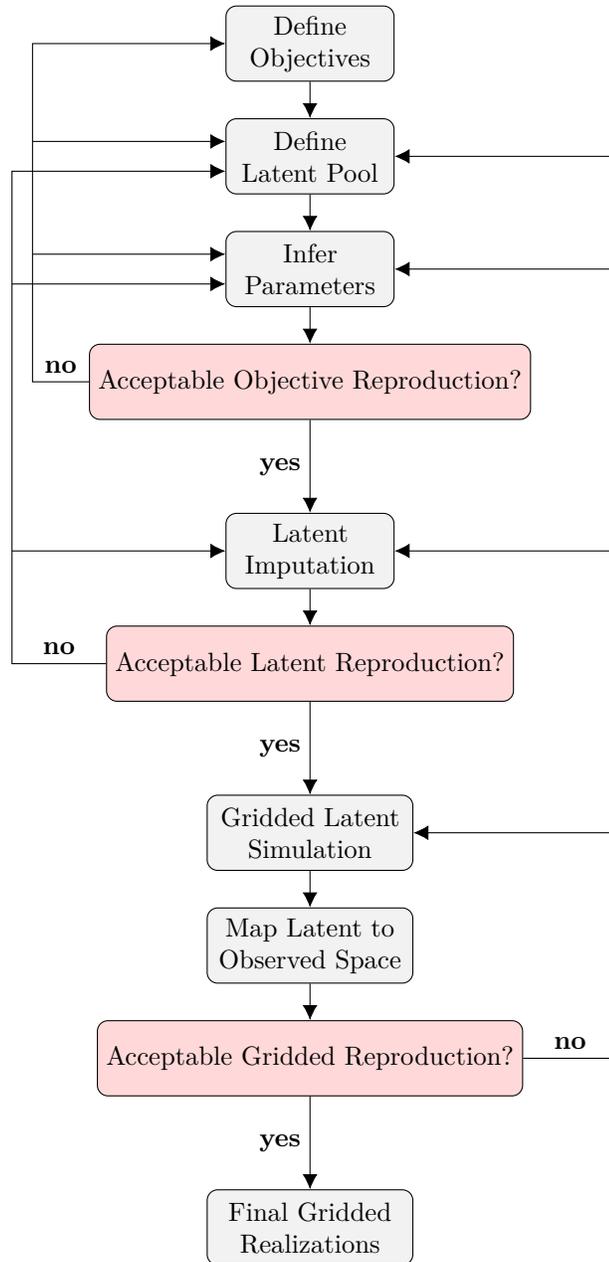


Figure 3.2: High level flow chart illustrating the key components of the NMR framework. There are multiple junctions where previous steps can be revisited and refined. Reverse-flow arrows leading to multiple actions emphasize the interconnected nature of each component.

3.2 Connectivity and Non-Gaussianity

A sequence is a collection of elements where the order of the elements matters. The arrangement or order of the elements can be used to characterize the connectivity within the sequence. Grades assayed on intervals down a drillhole represent a one-dimensional sequence of real numbers in a geo-statistical context. The connectivity of high and low values is often of practical importance, mainly when the transfer function is sensitive to extreme values. A one-dimensional sequence provides access to data-driven multiple-point configurations or patterns that would be difficult to infer in two- or three-dimensions. Connectivity is a different way of measuring correlation within a sequence. Each drillhole can be considered an exhaustive, one-dimensional training image from which n -point statistics can be inferred.

A natural extension of this concept is the analysis of runs of binary sequences from linear strings of data (Ortiz, 2003). A binary sequence is either 0 or 1 and computed through the indicator transform of a continuous RV. For a given threshold z_k , $k = 1, \dots, K$:

$$I(\mathbf{u}_i; z_k) = \begin{cases} 1, & \text{if } z(\mathbf{u}_i) \leq z_k \\ 0, & \text{otherwise} \end{cases} \quad (3.1)$$

It is common to consider multiple thresholds resulting in multiple binary sequences. A *run* of length L is defined as L identical values bound on either end by an opposite value. Runs of consecutive values above or below the threshold can be assessed. The theory of the distributions of runs for random uniform sequences is well documented by Fu and Lou (2003); the moments of the distribution of runs have analytical expressions, and they show that the limit distributions are normal. Though useful in many applications like cryptography and random number generation (Rukhin, Soto, et al., 2010), the assumption of independence between elements in the sequence is limiting in the spatially correlated scenario.

Ortiz (2003) shows that the analytical derivation of multi-point events is only possible when the multivariate spatial law is known. Practically, this is either the random or multivariate Gaussian case. A run of length L above a threshold z_k consists of $L + 2$ elements where the first and last elements are below the threshold. In the general case, the probability of a run of length L is defined as:

$$\begin{aligned} P\{\text{run } L\} = & \\ & P\{Z(\mathbf{u}) \leq z_k | Z(\mathbf{u} + \mathbf{h}) > z_k, \dots, Z(\mathbf{u} + L \cdot \mathbf{h}) > z_k, Z(\mathbf{u} + (L + 1)) \leq z_k\}, \dots, \\ & P\{Z(\mathbf{u} + L \cdot \mathbf{h}) > z_k | Z(\mathbf{u} + (L + 1)) \leq z_k\} \cdot \\ & P\{Z(\mathbf{u} + (L + 1)) \leq z_k\} \end{aligned} \quad (3.2)$$

The separation vector between elements is \mathbf{h} . This definition amounts to a recursive application of Bayes' Law to determine the joint probability of the multiple-point event. In the multivariate Gaussian case, the conditional probabilities are calculated using simple indicator kriging (Journal

& Alabert, 1989). With correlated sequences, the range of correlation influences the frequencies of runs. As correlation increases, there are fewer short and more long runs relative to random (Ortiz, 2003).

Correlated Gaussian sequences with an arbitrary covariance structure can be easily generated using any stochastic simulation algorithm and the indicator formalism of Equation 3.1. The total number of runs and frequencies of run lengths can be determined experimentally for the multivariate Gaussian scenario. If the covariance structure of the Gaussian realizations matches that of the true one-dimensional drillhole sequence, run-based statistics can measure non-Gaussianity. Section 3.2.3 presents this idea further. If the number or frequency of runs in the data deviates significantly from the Gaussian case, one may investigate the sequence as non-Gaussian. These measures can provide insight into domain sub-regions that exhibit non-Gaussian behaviour and warrant further investigation. Run-based connectivity measures in the following sections provide access to high-order statistics to characterize non-Gaussianity.

3.2.1 Distribution of Runs

High-order statistics characterize the spatial relationship between multiple points. In contrast to two-point statistics, such as the variogram, multiple-point statistics can reproduce curvilinear features and more complex ordering (Guardiano & Srivastava, 1993). Runs are one-dimensional patterns or a type of multi-point statistic calculated from data sequences like drillholes (Boisvert, Pyrcz, et al., 2007).

The indicator formalism (Equation 3.1) applied to linear, one-dimensional drill strings (Figure 3.3) generates binary sequences. This transform characterizes each element in the sequence as a binary event relative to thresholds z_k , $k = 1, \dots, K$. It is implicit that the elements in the sequence are equidistant; that is, the drillhole is composited. Multiple-point configurations, such as runs, can then be extracted from the sequences. In practice, one may have to adapt tolerances to infer statistics from approximately linear or equidistant sequence elements (Ortiz, 2003).

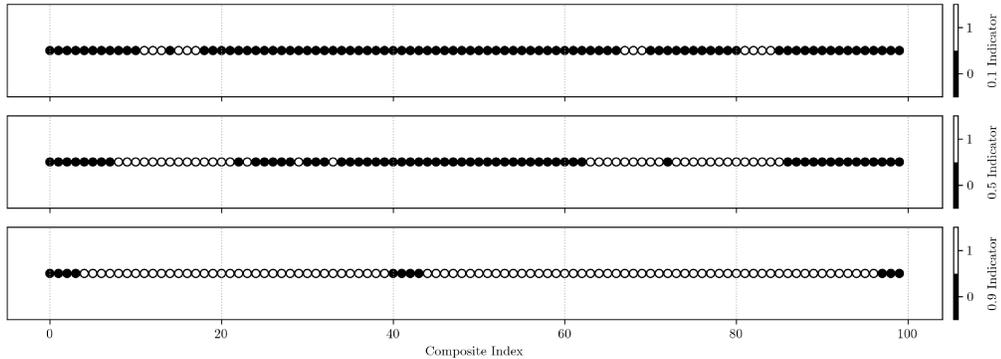


Figure 3.3: Indicator transform of a Gaussian RV using the 0.1, 0.5, and 0.9 quantiles as thresholds.

Calculating runs considers a cumulative or “overlapping” approach. That is, one run of 3 is

also two runs of 2 and three runs of 1; Figure 3.4 shows this nesting. This generalizes to a run of length L being i runs of length $L - i + 1$, $i = 1, \dots, L$. This cumulative approach accounts for the dependence between overlapping runs within the sequence (Fu & Lou, 2003). Considering cumulative runs generates a histogram of run lengths that decreases as run lengths increase. This consideration controls long runs as a long run contains elements for all shorter runs (Ortiz, 2003).

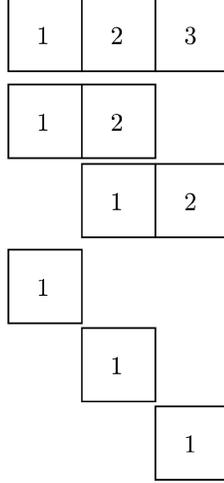


Figure 3.4: Example of lower-order runs for a single run of 3. One run of 3 is also two runs of 2 and three runs of 1.

It is interesting to note that the indicator transform results in nested sequences. Sequences that are above the current threshold are also above any lower threshold. Figure 3.3 illustrates this concept where all sequences above the 0.9 quantile (zeros) are nested within a sequence above the 0.5 quantile, which is nested within a sequence above the 0.1 quantile. Ortiz (2003) uses this property of sequences to simulate hierarchically as a run at one threshold can be used to condition the next threshold.

3.2.2 N -point Connectivity Function

A connectivity function quantifies the connectedness of a sequence (Renard, Straubhaar, et al., 2011). Two-point connectivity is characterized by the expected value of the product of indicators $I(\mathbf{u}; z_k)$ and $I(\mathbf{u} + \mathbf{h}; z_k)$. The n -point connectivity function, proposed by Journel and Alabert (1989), in a given direction can be generalized as the expected value of the product of n indicators where the lag distance \mathbf{h} is the distance between sequence elements:

$$\psi_{z_k}(n; \mathbf{h}) = E \left\{ \prod_{j=1}^n I(\mathbf{u} + (j-1)\mathbf{h}; z_k) \right\} \quad (3.3)$$

The n -point connectivity function describes the probability of n successive elements in the sequence being jointly *below* the threshold z_k . To consider the probability of being jointly *above* the

threshold z_k , Equation 3.3 becomes:

$$\psi_{z_k}(n; \mathbf{h}) = E \left\{ \prod_{j=1}^n 1 - I(\mathbf{u} + (j-1)\mathbf{h}; z_k) \right\} \quad (3.4)$$

Journel and Alabert (1989) show that the n -point connectivity function for a RV with connected extreme values simulated with a Gaussian RF model is much less than the true connectivity. Figure 3.5 shows an example of an n -point connectivity function for a non-Gaussian one-dimensional string data (red) and a Gaussian realization of the same string (gray). The Gaussian realizations are generated with LU matrix simulation (Davis, 1987) and have the same covariance structure as the non-Gaussian string. The n -point connectivity function shows that the probability of connected steps within the Gaussian realizations is significantly less than the non-Gaussian data for the first 20 steps.

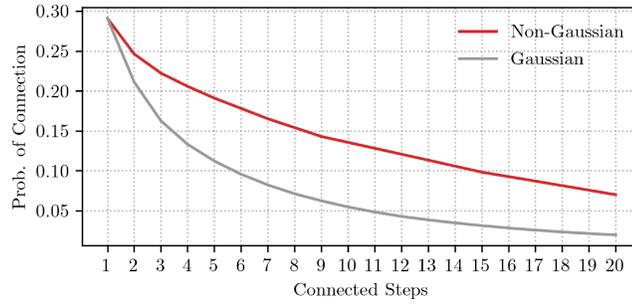


Figure 3.5: n -Point connectivity function for a string of non-Gaussian data (red) and the expected value of Gaussian realizations (gray) with the same covariance structure.

Notably, the n -point connectivity function considering runs *above* the threshold (Equation 3.4) is analogous to cumulative run-length frequencies of runs above the threshold scaled to the fraction of total elements in the sequence. Considering the n -point connectivity function above a quantile is the same as considering the cumulative run-length frequency above that quantile.

3.2.3 Measures of Non-Gaussianity

Four measures of non-Gaussianity are proposed. Three based on the concepts of runs and connectivity within drill string sequences, and one based on the volume-variance relation:

1. Total Runs: the number of unique runs above or below a threshold.
2. Run Length Frequencies: the frequency of cumulative run lengths above or below a threshold.
3. N -Point connectivity: the number of connected steps above or below a threshold.
4. Change of Support: the relationship between continuous variance and scale.

The first measure calculates the total cumulative runs within the binary sequence. As connectivity within the sequence increases, fewer but longer *total* runs exist. This connectivity translates to more *cumulative* runs as each longer run contains $\{L - i + 1, i = 1, \dots, L\}$ lower-order runs. A

non-Gaussian sequence with connectivity of extremes is expected to exhibit a greater number of cumulative runs than a Gaussian sequence. The second measure calculates the cumulative run-length frequencies within the binary sequence. Increased connectivity within the sequence leads to fewer but longer run lengths. Similar to the first measure of total runs, a non-Gaussian sequence with connectivity of extremes is expected to exhibit a greater frequency of longer run lengths. The nature of cumulative run lengths leads to a histogram of lengths that decreases as run length increases, facilitating a more straightforward comparison of distributions. The third measure is the binary sequence's n -point connectivity. The n -point connectivity function quantifies the probability of n successive elements in the sequence being below (or optionally above) the indicator thresholds. Only elements below the threshold contribute to the probability in Equation 3.3. As connectivity within the sequence increases, the probability of successive elements being jointly below the given threshold increases. A highly structured non-Gaussian sequence is expected to have a greater probability of n connected steps compared to a maximum entropy Gaussian sequence as n increases. Figure 3.5 highlights this characteristic.

The fourth measure is not sequence-based but calculates the change of support for the original continuous variable. This measure is quantified by averaging n consecutive elements within the sequence and calculating the change in variance relative to the original. As volume increases, the variance of the elements within the sequence decreases. The idea is that if the sequence has structured or connected extreme values, the variance of the sequence should be less sensitive to scale. A non-Gaussian sequence with connected extreme values is expected to show a less drastic reduction in variance as scale increases compared to a maximum entropy Gaussian sequence.

The covariance structure of each sequence is calculated from the exhaustive drill string to measure non-Gaussianity. This covariance is used to generate multivariate Gaussian realizations with the exact covariance of the original sequence. The Gaussian realizations are back-transformed to original units, and then indicator transformed with Equation 3.1. Each metric is calculated for each one-dimensional data string and compared to the distribution of metrics observed from the Gaussian realizations. The deviation between the original sequence and the Gaussian distribution measures non-Gaussianity. The general workflow for calculating the proposed measures of non-Gaussianity on a drillhole-by-drillhole basis is as follows:

1. Indicator transform the grades of all drillholes for quantiles of interest.
2. Calculate cumulative runs and run-length frequencies for all thresholds for all drillholes.
3. Calculate the n -point connectivity function for all thresholds for all drillholes.
4. Composite the continuous variable by a number of length factors to calculate the relationship between variance and scale.
5. Normal score transform the grades of all drillholes.
6. Calculate the autocovariance matrix for each drillhole and simulate $\ell = 1, \dots, L$ unconditional

- Gaussian realizations of each drillhole.
7. Back transform Gaussian realizations to original units.
 8. Repeat steps 1-4 for each realization of each drillhole.
 9. Calculate the score for each measure as $y = \frac{z_{dh} - \mu_r}{\sigma_r}$ where z_{dh} is the measure from the original drillhole and μ_r and σ_r are the mean and standard deviation of the distribution of measures from the realizations, respectively.
 10. The final y score is taken as $|y|$ such that only the magnitude, not the sign, is considered.

The y score measures how many standard deviations away from the Gaussian distribution the original drillhole is. Figure 3.6 illustrates y -score calculation. This score indicates the appropriateness of a Gaussian RF model, as the Gaussian realizations have the exact covariance structure of the original drillhole. Any drillhole with a y value greater than 2.5 is considered highly non-Gaussian.

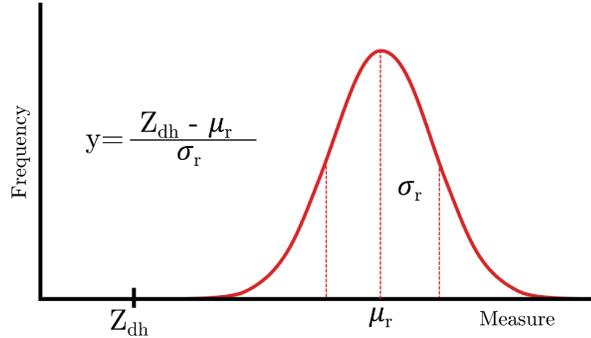


Figure 3.6: Schematic illustrating the calculation of the non-Gaussian Score y where z_{dh} is the metric from the drillhole, and μ_r and σ_r are the mean and standard deviation of the Gaussian distribution.

Consider a 2-D synthetic example generated from an image of a meandering river system. The red-green-blue (RGB) colour channels are averaged to generate a greyscale image, which is then normalized $\in [0, 1]$. The main river channels show structured high values relative to the background flood plane and abandoned channels (Figure 3.7). These features are a natural example of non-Gaussian characteristics in geospatial data. The image is “drilled”, resulting in 10 drill strings or sequences used to calculate non-Gaussianity measures.

Table 3.1 shows the expected non-Gaussian scores for each sequence across 100 realizations for the 0.1 and 0.9 quantile indicator transforms. The structured regions are predominantly high-grade, so the 0.1 quantile indicators do not show significant non-Gaussianity. Drillholes 5 and 7 show strong non-Gaussianity for all connectivity metrics, while holes 8 and 9 are moderately non-Gaussian. These drillholes intersect structured high-grade river channels in the eastern portion of Figure 3.7 and exhibit connectivity features that cannot be reproduced by a multivariate Gaussian RF. Abrupt transitions between high- and low-grade regions are difficult to capture with a maximum entropy RF that leads to increased connectivity of median values and disconnected extremes.

There is likely a spectrum of non-Gaussianity, and the practitioner must decide on appropriate thresholds. The proposed measures indicate that some regions within the domain have non-Gaussian

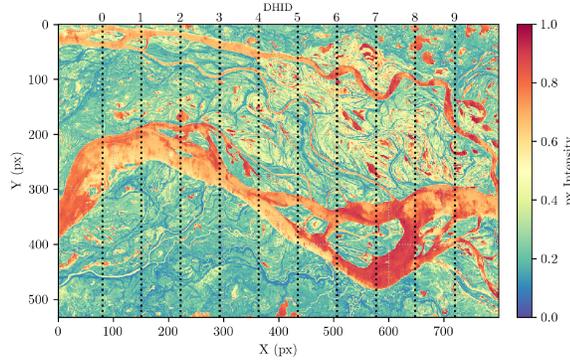


Figure 3.7: Image of a meandering river system sampled in ten locations to generate sequences of pixel data. DHID = drillhole ID, px = pixel.

Table 3.1: Non-Gaussian metrics calculated for ten drillholes considering the 0.1 and 0.9 quantile indicator transforms. y -scores ≥ 2.5 are considered strongly non-Gaussian. DHID corresponds to Figure 3.7.

dhid	Total Runs		Cumulative Runs		N-point Connectivity	
	0.1 Indicator	0.9 Indicator	0.1 Indicator	0.9 Indicator	0.1 Indicator	0.9 Indicator
0	1.18	0.23	1.48	0.17	1.23	0.17
1	0.55	0.10	0.35	0.05	0.32	0.05
2	0.68	1.18	0.72	1.81	0.68	1.81
3	0.41	0.58	0.68	1.08	0.64	1.08
4	0.75	0.05	0.63	0.14	0.72	0.14
5	1.16	2.44	1.15	5.18	0.96	5.17
6	1.34	0.76	1.25	0.97	0.89	0.96
7	1.18	5.15	1.24	4.73	1.16	3.73
8	0.89	1.41	1.20	1.50	1.03	1.48
9	1.04	1.26	0.77	1.39	0.61	1.50

features. These measures can identify sub-regions that warrant further investigation. The NMR is designed to capture these non-Gaussian connectivity measures. The goal of NMR is to develop a framework for incorporating two- and n -point statistics where two-point statistics come from the continuous variogram and indicator variograms, and the n -point statistics come from drillhole sequences.

3.3 Network Methodology

Sections 3.2.1 and 3.2.2 present methodology for accessing higher-order statistics from drillhole sequences. Section 3.2.3 shows that these statistics can effectively characterize non-Gaussianity. This section presents the overall NMR framework and the generation of non-Gaussian spatial fields by integrating these statistics. Achieving higher-order connectivity is a key component of the NMR framework. Two-point, variogram-based statistics alone cannot capture the complex multi-point relationships we seek in a non-Gaussian RF. The indicator transform of drill strings provides access to binary sequences that permit the calculation of multi-point connectivity statistics.

3.3.1 Notation

This section shows an overview of the NMR mathematical notation and definitions, while Chapters 4 and 5 present complete details. Consider a continuous RF:

$$\{Z(\mathbf{u}), \forall \mathbf{u} \in \mathcal{D}\} \quad (3.5)$$

where \mathcal{D} is a domain of interest. The location vector \mathbf{u} could be data or grid node locations. Next, consider a set of $M + 1$ latent variables, each characterized by a Gaussian RF:

$$\{\mathbf{Y}(\mathbf{u}) = (Y_0(\mathbf{u}), \dots, Y_M(\mathbf{u})), \forall \mathbf{u} \in \mathcal{D}\} \quad (3.6)$$

Finally, consider a forward mapping function \mathcal{F}_θ that defines the mapping from the real-valued latent space to the real-valued observed space:

$$\{\mathcal{F}_\theta : \mathbb{R}^M \mapsto \mathbb{R}\} \quad (3.7)$$

where θ is a parameter vector that characterizes \mathcal{F} , and \mathcal{F}_θ is such that the mapping of the latent space to the observed space reproduces the observed data values:

$$\{\mathcal{F}_\theta(\mathbf{Y}(\mathbf{u})) = Z(\mathbf{u}), \forall \mathbf{u} \in \mathcal{D}\} \quad (3.8)$$

The NMR approximates the forward mapping function \mathcal{F}_θ . Chapter 4 discusses inference of the parameter vector, θ , and Chapter 5 presents imputation of the Gaussian RFs such that the equality in Equation 3.8 holds.

3.3.2 The NMR Inverse Problem

The NMR approximates the forward mapping function from latent to observed space and can be considered a model-based inversion problem (Sen & Stoffa, 2013). The parameters of this function (or model) are unknown and must be inferred from the observed measurements or data. The goal is to find the unknown parameters, θ , so the model output has the desired spatial characteristics. In this context, the observed measurements are drillhole data or empirical statistics calculated from these data. Synthetic data are generated by mapping latent Gaussian variables through \mathcal{F}_θ for a given state of θ . If the match between the spatial characteristics of the synthetic data and those of the observed data is acceptable, then the parameters are retained. Otherwise, θ is perturbed until the match is acceptable. An objective function quantifies the mismatch between the model output and observed data. Figure 3.8 shows a high-level flowchart of this optimization process. The iterative parameter search, or minimization of the objective function, is an optimization problem approached with DE, a directed Monte Carlo search method; Chapter 4 presents details on the specific formulation of this problem.

Inverse problems are typically ill-posed; that is, the solution is non-unique. To help mitigate the non-uniqueness of the solution, constraints in the form of (1) prior information and (2) parameter constraints are imposed on the solution. Prior geologic information is incorporated through models

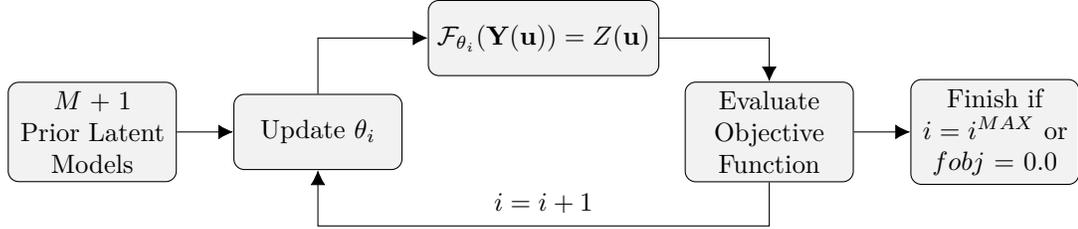


Figure 3.8: Flowchart showing an overview of the optimization workflow for determining the parameter vector θ .

of latent Gaussian variables (Section 3.3.3) and forms the basis of the NMR output. Careful design of these latent variables ensures the optimization algorithm explores the appropriate solution space from a geologic perspective and acts as a regularization element (Zhou, Gómez-Hernández, et al., 2014). The model output is a mixture of these latent variables; the observed statistics are reproduced by integrating spatial features from the prior models. Geologically reasonable prior information ensures a solution for θ is feasible. Limiting parameter values based on their physical meaning can further constrain the problem. As the NMR is a positive, non-linear combination of the latent variables, each latent variable’s relative contribution cannot be negative; these values are always constrained to be ≥ 0 . Uncertainty in the parameter vector θ is captured by considering multiple realizations of the prior models during optimization. The objective function is minimized in expectation; that is, it considers the mismatch error across all realizations. Considering a space of uncertainty in the prior models ensures the parameters are not overly sensitive to the features of a particular realization, acting as an additional regularization element.

Determining the parameters of the NMR does not directly consider the reproduction of the observed data values but rather the observed two- and multi-point statistics. The goal is to learn how to map from the latent to the observed space, which results in the desired spatial structure. As discussed above, multiple distributions could reproduce these statistics, so explicit data matching is unnecessary, simplifying the optimization. Additionally, by relaxing this data-matching constraint, the prior latent models can be generated through unconditional simulation. Exact reproduction of the observed data values is ultimately required, and Chapter 5 addresses this imputation problem in detail.

3.3.3 Latent Spatial Structure

The set of latent Gaussian variables of Equation 3.6 is the foundation of the NMR. This set is referred to as the Gaussian “pool” throughout this text; the components of the pool are referred to as latent “factors”. It is a pool in that there is a collection of Gaussian RFs to be shared with the goal of reproducing high-order statistics. This pooling is analogous to a GMM in the spatial context. The idea of a GMM is that a finite mixture of Gaussian densities can approximate a continuous distribution (as in Equation 3.5) (McLachlan, Lee, et al., 2019). Furthermore, Silva and

Deutsch (2018) shows that a mixture of Gaussian components can fit complex non-Gaussian uni- and multivariate distributions. The NMR is an extension of this concept; rather than a mixture of Gaussian densities, $Z(\mathbf{u})$ is approximated by a mixture of Gaussian RFs. Rather than using the expectation-maximization algorithm to fit the parameters of each Gaussian component (McLachlan, Lee, et al., 2019), M standard Gaussian RFs, with covariance structures specified by $\gamma_m(\mathbf{h})$, are explicitly chosen. The factors are chosen such that:

1. Latent factors are standard normal: $E \{Y_m(\mathbf{u})\} = 0, E \{Y_m(\mathbf{u})^2\} = 1, \forall \mathbf{u}$
2. Latent factors are independent: $E \{Y_m(\mathbf{u}_i)Y_n(\mathbf{u}_j)\} = 0, \forall m \neq n, \forall i \neq j$
3. Latent factors reproduce their respective variogram model $\gamma_m(\mathbf{h}), \forall \mathbf{h}$

The output of the mapping function \mathcal{F}_θ is a spatial mixture of the latent factor components of $\mathbf{Y}(\mathbf{u})$.

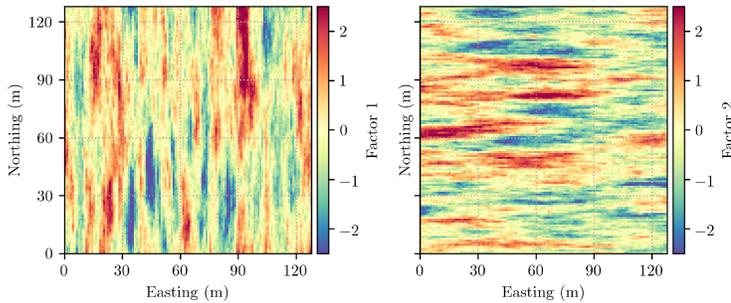


Figure 3.9: Unconditional realizations of Gaussian RFs with orthogonal directions of continuity.

Consider a hypothetical scenario where the modeling goal is to have highly continuous low grades in the east-west direction and highly continuous high-grade continuity in the north-south direction. This orthogonal continuity is challenging for a multivariate Gaussian simulation algorithm, as a normal score variogram cannot simultaneously capture both directions of continuity. The NMR can draw different spatial structures from different latent factors within the pool; the sharing of structures is a core concept of the NMR approach. Consider the 2-D realizations of two unconditional Gaussian RFs in Figure 3.9. It is possible to achieve the modeling goals if we can draw high-grade continuity from factor 1 and low-grade continuity from factor 2. A simple linear combination of the factors, analogous to the linear model of co-regionalization (LMC), cannot capture the orthogonal continuities as high and low values cancel when combined. Before mixing, we must emphasize the high values in factor 1 (and mute the low values) and the low values in factor 2 (and mute the high values). This “extraction” of specific factor features requires the introduction of a non-linearity.

3.3.4 Non-Linearity and Mapping

Emphasizing a portion of a range of values requires the definition of a threshold. Given item (1) from the list in the previous section, zero is the logical threshold value of a Gaussian distribution. A straightforward approach to introducing non-linearity is applying a power law type function. The

non-linear value is the original value raised to a power, ω . If we want to emphasize the high values, all values > 0 are raised to the exponent $\omega \geq 1$ while the values < 0 are raised to the exponent $\frac{1}{\omega}$. The non-linearity effectively mutes the magnitude of values below zero and exponentially increases values above zero. The opposite is true if we wish to emphasize low values; all values < 0 are raised to the exponent $\frac{1}{\omega}$ while the values > 0 are raised to the exponent ω , where $\omega < 1$. Figure 3.10 shows this power law relationship graphically for the two realizations shown in Figure 3.9 using $\omega_1 = 4$ and $\omega_2 = 0.25$. Where factor 1 is greater than zero, the function significantly emphasizes values > 1 . Where factor 1 is less than zero, the function mutes or dampens values < -1 . This non-linearity advantageously isolates the high-grade structure from factor 1. The opposite is true for factor 2. Chapter 4 presents the complete details of the non-linearity $\phi(\mathbf{y}, \boldsymbol{\omega})$.

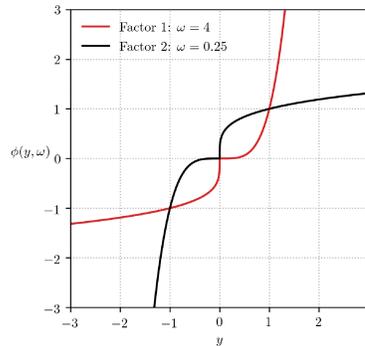


Figure 3.10: Power law relationship between the linear input y and the non-linear output $\phi(y, \omega)$, where $\phi(\dots)$ is the power law function.

Figure 3.11 (left, center) shows the spatial distribution of factors 1 and 2 after applying the power law function. Note the magnitude of the colour bars between the plots. The power law function can isolate the desired components of each factor and can be thought of as an “activation” function. Figure 3.12 (left, center) shows the histograms of factors 1 and 2 after transformation. As expected $\phi(y_1, \omega_1)$ is strongly positively skewed, and $\phi(y_2, \omega_2)$ is strongly negatively skewed. $\phi(\dots)$ results in a distribution of arbitrary activation units. These units are not interpretable and require combination and mapping to an observed space in Gaussian units. Combining the “activated” realizations is a weighted, linear combination, followed by a normal score transform. For the sake of the example, the weights to each factor are $\mathbf{a} = [1, 1]$. The final model is then $\mathbf{z} = G^{-1}(\phi(\mathbf{Y}, \boldsymbol{\omega}) \cdot \mathbf{a}^T)$, where G^{-1} is the inverse of the Gaussian CDF.

Figure 3.11 (right) shows the spatial distribution of the final model, while Figure 3.12 (right) shows the histogram. Note that the spatial distribution is complex and non-multivariate Gaussian. However, the univariate histogram is perfectly Gaussian. The high values show north-south continuity, while the lows show east-west continuity. The overprinting of the factors appears natural, resulting in a spatial structure that a single variogram model cannot generate. These steps provide an overview of the forward pass through the NMR. In summary, they are:

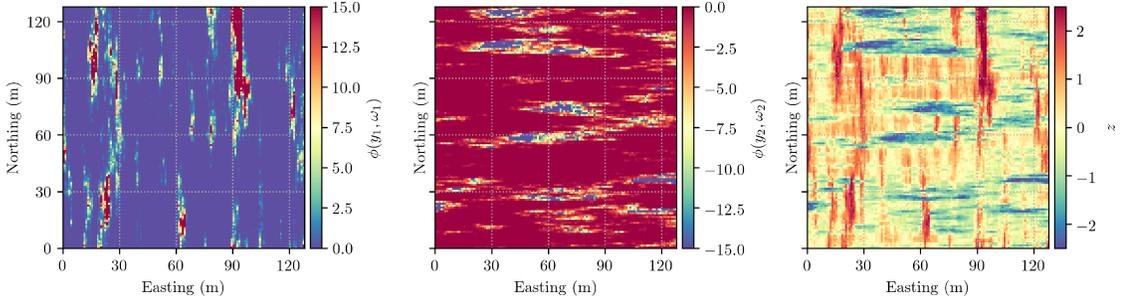


Figure 3.11: Factors 1 (left) and 2 (center) after application of the power law function, and the final mapped, normal score transformed realization (right).

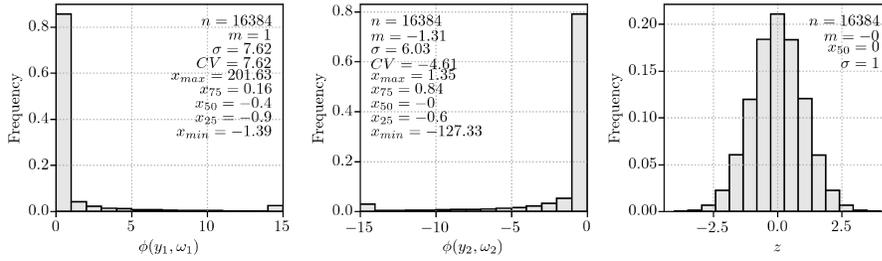


Figure 3.12: Histograms of factors 1 (left) and 2 (center) after application of the power law function, and the final mapped, normal score transformed realization (right).

1. Unconditionally simulate latent factors to form the Gaussian pool
2. Activate the factors with ω to emphasize certain features
3. Linearly combine the activated factors with weights \mathbf{a}
4. Normal score transform the combination

In practice, parameters ω and \mathbf{a} are unknown and must be inferred. As Section 3.3.2 discusses, inverse problems are commonly approached by optimization. The optimization problem consists of determining ω and \mathbf{a} for a given Gaussian pool to minimize an objective function. The objective function in this context is the reproduction of desired two- and multi-point statistics in the final model produced by the steps above.

3.3.5 Two-Point Statistics

Two-point statistics are the experimental normal score and indicator variograms. The experimental variogram is the expected squared difference between data pairs separated by lag vector \mathbf{h} . Figure 3.13 shows the normal score variogram (blue), the 0.1 quantile indicator variogram (black), and the 0.9 indicator variogram (red) of the mapped model in the north-south (left) and east-west (right) directions, respectively. The final model achieves its goals from the perspective of two-point statistics. The indicator variograms show increased low-grade continuity in the east-west direction and high-grade continuity in the north-south direction. The normal score variogram is a blend of the indicator variograms; in the north-south direction, it has the continuity of the highs with the

cyclicality of the lows and the range of the lows and cyclicality of the highs in the east-west direction. This feature blending further emphasizes how a single variogram cannot capture the complexity of the NMR.

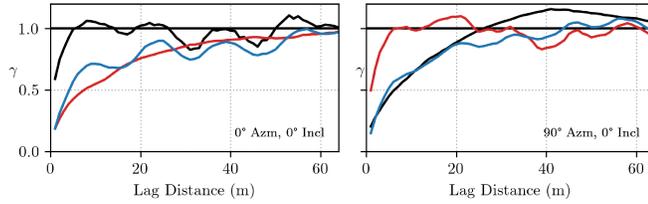


Figure 3.13: Normal score experimental variogram (blue), the 0.1 quantile indicator experimental variogram (black), and the 0.9 indicator experimental variogram (red) of the mapped model in the north-south (left) and east-west (right) directions, respectively.

An interesting experiment to confirm this behaviour is to generate a realization with a multivariate Gaussian simulation algorithm using the normal score variogram from Figure 3.13 and compare it to the final NMR model. Figure 3.14 shows the same NMR realization from above (left) and a realization generated with SGS (right). The SGS model uses the variogram model fitted to the blue experimental variogram in Figure 3.13. The NMR and SGS models are significantly different, though they share the same two-point covariance structure. The SGS model shows less connectivity in low and high values and generally more disorder. The normal score variogram shows a mixing between the orthogonal structures and cannot adequately capture multiple anisotropies when considering the full range of values. In contrast, the NMR isolates a particular covariance structure to values above and below zero.

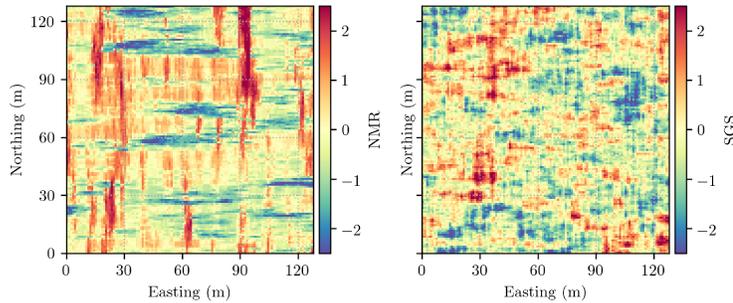


Figure 3.14: The final NMR model (left) and a SGS model generated with the same normal score variogram model.

Beyond two-point statistics, the network approach also considers the multi-point statistics discussed in Section 3.2. A non-Gaussian RF should show increased multi-point connectivity of extreme values over a Gaussian RF. Whether it is connectivity of high or lows depends on the structure of the latent pool and power law exponents, ω .

3.3.6 Multi-Point Statistics

Consider the 2-D models above as sections rather than plan view. We can “drill” the realizations to generate sequences of data. Sampling the realization results in 32 synthetic drillholes, each with 64 samples. Figure 3.15 shows the drillhole configuration and the corresponding 0.9 quantile indicator transform for the NMR (left) and SGS models (right). The direction of continuity of the high-grade factor corresponds with the direction of drilling.

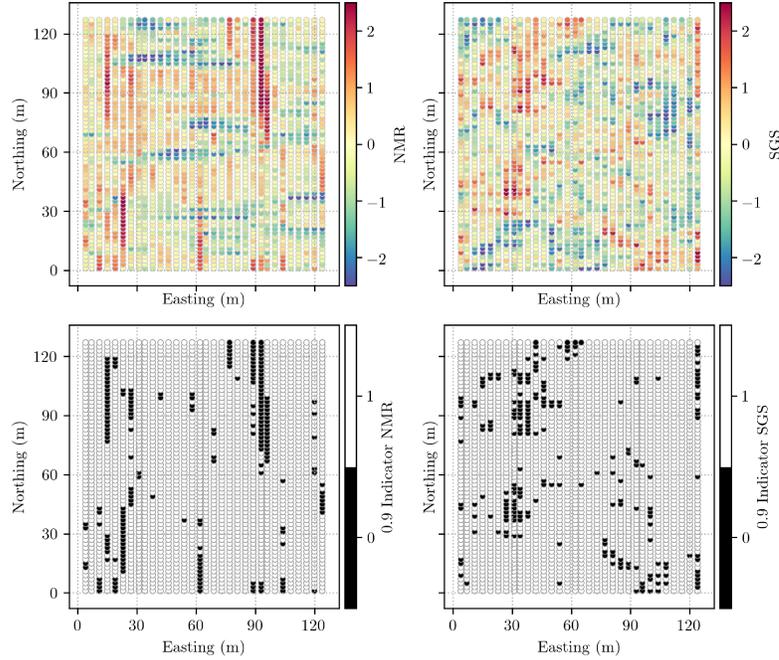


Figure 3.15: Synthetic drillhole samples and corresponding 0.9 quantile indicator transform for the NMR model (left) and the SGS model (right).

As expected, the NMR model visually shows increased high-grade connectivity. The more structured high-features result in fewer total but longer connected sequences. The SGS model is more disorganized, resulting in shorter connected sequences. The n -point connectivity function in Figure 3.16 quantifies this visual discrepancy. The probability of connection above the 0.9 quantile for the multivariate Gaussian model steeply declines in the first five steps and is effectively zero at seven steps. The probability of connection in the non-Gaussian model decreases notably slower and remains > 0 at 20 connected steps.

Connectivity measures are calculated globally, that is, considering all drillholes. The primary orientation of drilling is an important consideration when designing latent factors with the goal of high-order connectivity. In practice, high-grade structures are generally orthogonal to the primary orientation of drilling, so specific factors aligned with the tertiary variogram direction may be required.

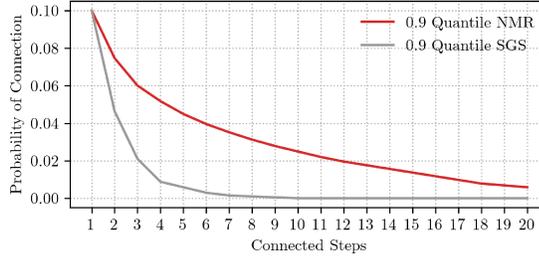


Figure 3.16: N -point connectivity function for the 0.9 quantile indicator transform of the drillholes in Figure 3.15.

3.3.7 Latent Imputation

The previous sections consider using unconditional realizations to infer the mapping function \mathcal{F}_θ . The ultimate goal is to generate gridded realizations conditional to the observed data with non-Gaussian spatial features characterized by θ . This process requires gridded conditional realizations of each latent factor in the pool. As the latent factors are a synthetic construct of the NMR model and not directly observed, they must be imputed (Little & Rubin, 2019). The imputed factors have the conditions listed in Section 3.3.3 plus the additional constraint that they must reproduce the observed data (within a specified tolerance) when mapped through \mathcal{F}_θ :

$$\mathcal{F}_\theta(\mathbf{y}(\mathbf{u})) = \mathbf{z}(\mathbf{u}) \pm \alpha, \forall \mathbf{u} \quad (3.9)$$

where α is a data matching tolerance. A straightforward approach to satisfying Equation 3.9 is to assign random Gaussian values to the vector $\mathbf{y}(\mathbf{u})$ until the equality is met (Silva, 2018). This approach ensures the correct collocated multivariate relationship between the latent factors. However, it does not ensure that each regionalized factor has the correct spatial variability. If the imputed latent factors do not have the correct spatial variability defined by the pool, the mapping function \mathcal{F}_θ is no longer valid. Directly sampling the high-dimensional multivariate distribution is difficult, though sampling the marginal conditional distributions is possible. This problem is typically approached by a Gibbs sampler (Geman & Geman, 1984) and commonly employed in the truncated Gaussian simulation paradigm (Arroyo & Emery, 2020; Madani & Bazarbekov, 2021). Noted Gibbs sampler convergence issues with spatially correlated data motivates the development of a novel imputation algorithm presented in Chapter 5. The algorithm combines sequential simulation and rejection sampling components to iteratively sample the marginal distributions, resulting in latent factors that satisfy Equation 3.9 and have the correct covariance structure.

The solution to Equation 3.9 is non-unique. Multiple combinations of latent variables can reproduce the observed value when mapped. Figure 3.17 shows the relationship between two imputed factors and the observed value, $z(\mathbf{u})$. In this example, if $z(\mathbf{u})$ is high, say 1.5, factor 2 is constrained to be high. However, factor 1 can take any value from the range of $[-2.5, 2.5]$. The opposite is true for low values, where factor 1 is constrained. The NMR framework utilizes multiple imputation

to transfer uncertainty related to non-uniqueness to the final gridded realizations. Multiple latent factor realizations are imputed, and a unique data realization conditions each gridded realization.

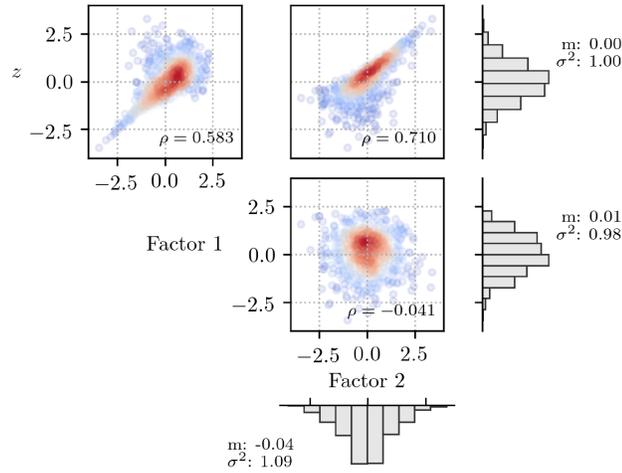


Figure 3.17: Bivariate scatter plots comparing the observed value, z , and factors 1 and 2. The histograms are the marginal distributions of each variable.

Chapter 5 presents the details of the imputation algorithm and the of checking latent data realizations. The algorithm shows stable convergence for spatially correlated variables and correctly reproduces the latent pool’s collocated multivariate relationships and covariance structure.

3.3.8 Simulation and Mapping

After imputing all latent factors at the data locations, they are conditionally simulated at grid node locations. Any conditional simulation algorithm is valid; different algorithms could be used for different structures depending on the range of correlation, structure type and anisotropies (Pinto, 2020). Once the latent factors are defined at all grid nodes, the gridded realizations are mapped from latent to observed space with \mathcal{F}_θ . The mapping function includes a normal score transform based on a reference distribution of activation values output by the network. The corresponding normal score values of these activations are used as a transform table to transform the gridded realizations into Gaussian units.

After transformation to observed space, the gridded NMR realizations reproduce all specified two- and multi-point spatial features while being univariate Gaussian. Transforming the realizations to original units is simply the inverse of the normal score transform. Chapter 5 presents further details regarding the checking and validating of both latent and mapped gridded realizations.

3.4 Effects of High-Order Continuity

The NMR framework permits a flexible approach to continuous variable simulation in the presence of non-Gaussian geologic domains. One can design latent factors to account for differences between

background and high-grade mineralization or changes in the orientation of continuity if sub-domains cannot be defined. The flexibility of the latent pool is advantageous, providing $M \cdot 7$ orientation, range and covariance structure parameters, compared to the 7 of a single variogram model in a traditional multivariate Gaussian simulation algorithm. The following section presents a small synthetic, non-Gaussian example to highlight the effect of multi-point connectivity on expected mineral resources. NMR resources are contrasted against those from SGS.

The reference truth comes from a natural image of fork lightning rotated 90 degrees. The image is chosen as it exhibits narrow, highly connected, high-value, dendritic features with abrupt changes in grade between the lightning and the background features. It is reasonable to suggest similarities between the image and narrow vein-type mineral deposits. The image is rotated 90 degrees, so the synthetic drillholes are roughly orthogonal to the lightning structures. The RGB colour channels are averaged, generating a grayscale image. The grayscale image is normalized $\in [0, 1]$ and then transformed to a log-normal distribution with a $\mu = 1$ and $\sigma = 2$. Data with a CV of 2 is common in mineral systems with positively skewed distributions, such as precious metals. Twenty synthetic drillholes are sampled from the image, extracting every third pixel, resulting in 1560 samples. Figure 3.18 shows the reference image (left) and the synthetic drillhole configuration (right). The drillhole data is strongly non-stationary; the high-grade features are vertically continuous in some areas and horizontally in others. The dip of the structures also changes locally. The goal of the NMR model is to capture these non-stationary features and correctly characterize the connectivity of high and extreme values observed in the true image. Though synthetic, this scenario provides an excellent opportunity to highlight the capabilities of the NMR.

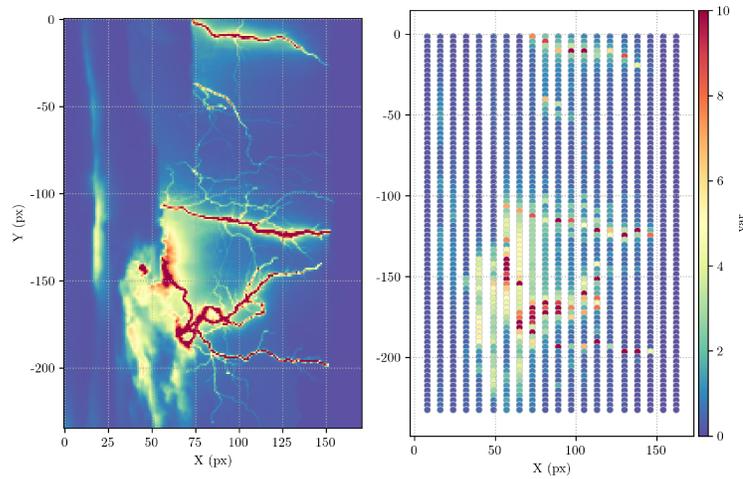


Figure 3.18: The non-stationary reference truth image of fork lightning (left) and the drillhole samples extracted from the image (right).

The normal score variogram model plus two additional highly anisotropic factors comprise the Gaussian pool. The normal score variogram model exhibits strong zonal anisotropy where the range in the direction parallel to the sampling is larger than the domain size. The first of the highly

anisotropic factors is oriented 100 degrees with a 10 : 1 anisotropy ratio, and the second is oriented 45 degrees with a 5 : 1 anisotropy ratio. The normal score variogram factor can influence any grade range with $\omega \in [0.25, 4.0]$ while the additional factors are constrained to influence high values with $\omega \in [2.0, 4.0]$. The first factor can capture the orientation of medium to high-grade values, while factors two and three capture the orientation of the highest-grade structures. The mapping function, \mathcal{F}_θ , is inferred by minimizing the sum of squared errors between the objective function components and the two and multi-point statistics extracted from the synthetic data. Chapter 4 gives the complete details of the objective function and optimization algorithm. One hundred realizations of each latent factor are imputed at the data locations such that Equation 3.9 holds using $\alpha = 0.01$. Once the latent factors are defined at the data locations, they are conditionally simulated on a 1 x 1 pixel point scale grid using SGS. The point scale latent realizations are mapped through the same function, resulting in a univariate Gaussian spatial mixture with features characterized by θ .

The point scale realizations are back-transformed from Gaussian to original units and block averaged to a 5 x 5 pixel selective mining unit (SMU) scale grid. The SMU scale realizations are post-processed to calculate the e-type mean. Figure 3.19 shows the block averaged reference truth (top left) and corresponding 0.9 quantile indicator transform (bottom left), with the SMU scale NMR e-type model and indicators (top middle, bottom middle), and the SMU scale SGS e-type model and indicators (top right, bottom right). An outcome of correctly characterizing the point scale, high-grade continuity is that the SMU scale realizations should show more connectivity. As discussed in Section 3.2.3, disordered realizations are more sensitive to changes in scale; connected or organized features should remain as scale increases (to an extent). This connectivity is evident in the 0.9 quantile indicator transform of the NMR e-type model. The indicator model exhibits increased east-west high-grade continuity over the SGS model, particularly in the central and top portions of the grid. The SGS model effectively captures the vertical high-grade continuity; however, the east-west structures are visibly more disconnected.

Table 3.2 summarizes the resources of the SMU scale NMR and SGS models above the 0.1, 0.5, and 0.9 quantiles as a fraction of true resources. The resources assume the synthetic variable is measured in grams per metric tonne, priced in troy ounces, and density is a constant value of $2.6g/cm^3$. The NMR and SGS models show similar resources for the 0.1 and 0.5 quantile cutoffs. However, the NMR model shows improvement in both tonnes and grade above the 0.9 quantile cutoff, leading to a 9% increase in contained metal ounces relative to the SGS model. The increase in tonnes above the true 0.9 quantile is attributed to the increase in east-west continuity imparted by factors two and three. A single covariance model considering the complete range of grade values cannot effectively capture the non-stationary features of the true image.

Drill hole data may exhibit non-Gaussian features that are difficult to capture with a single covariance model based on two-point statistics. The NMR model permits generating multiple realizations that consider both two- and multi-point statistics from the observed data. These considerations al-

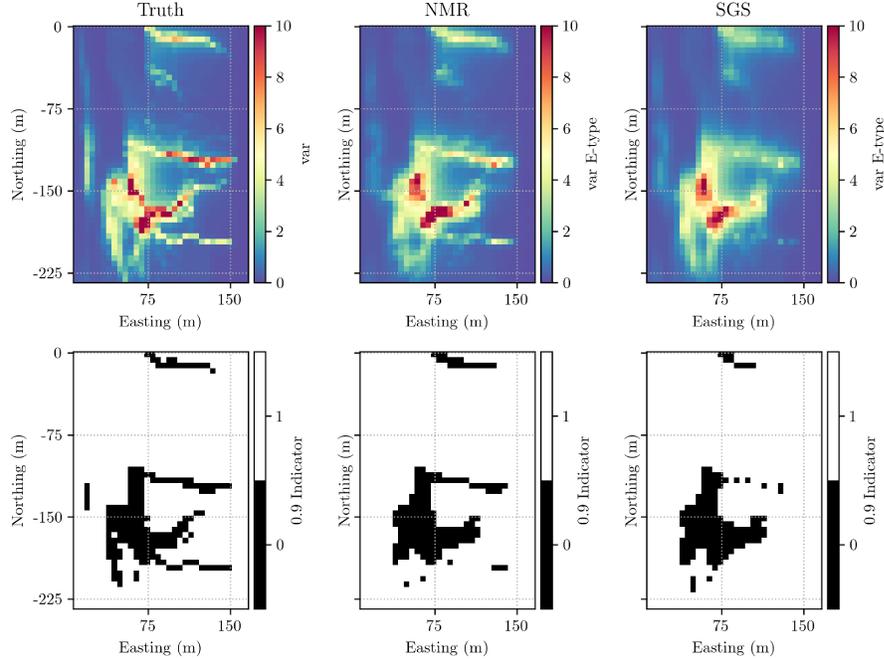


Figure 3.19: The reference truth block averaged to a 5 x 5 pixel SMU (top left) and corresponding 0.9 quantile indicator transform (bottom left), with the SMU scale NMR model and indicators (top middle, bottom middle), and the SMU scale SGS model and indicators (top right, bottom right)

Table 3.2: SMU scale resources above the 0.1, 0.5, and 0.9 quantiles as a fraction of the true resources. Cutoff values are calculated from the true image. g/t=grams per tonne.

Cutoff	0.1			0.5			0.9		
	Truth	SGS	NMR	Truth	SGS	NMR	Truth	SGS	NMR
Tonnes	1.00	1.00	1.00	1.00	1.07	1.04	1.00	0.82	0.88
Grade (g/t)	1.00	0.95	0.97	1.00	0.89	0.93	1.00	0.95	0.99
Ounces	1.00	0.94	0.96	1.00	0.95	0.97	1.00	0.78	0.87

low for extreme values that are more structured than what is possible with a multivariate Gaussian simulation algorithm. This small example emphasizes the importance of the connectivity of extreme values concerning SMU scale resources. Correctly characterizing the connectivity of extremes significantly impacts the contained metal of a resource estimate.

3.5 Discussion

The NMR is a framework for generating non-Gaussian spatial fields. Characterizing non-Gaussian spatial features, like the connectivity of extreme values, requires statistics above the second order. It is shown that these higher-order statistics, such as the distribution of runs and the n -point connectivity function, can differentiate non-Gaussian from Gaussian sequences. The NMR is designed to overcome some shortcomings of the multivariate Gaussian RF model, particularly with strongly positively skewed distributions. This advancement is achieved by considering both two- and multi-point

statistics in generating a probabilistic model. Examples highlight spatial structures that cannot be reproduced with a two-point covariance structure and a multivariate Gaussian simulation algorithm. A core concept of the NMR is that it is a spatial GMM, where the mixing of Gaussian components results in a non-Gaussian output.

A unique covariance structure defines each component of the mixture; this “pool” of components is designed such that certain components impart certain spatial features in the final model. A power law activation function applies non-linearity to the input components. This non-linear activation function allows features of the latent factors to be emphasized in particular regions of the continuous grade range. That is, the low values and high values can have different spatial structures and different multi-point connectivity. This difference in continuity is something that cannot be easily achieved with two-point statistics alone. The spatial features of the final model are defined at the beginning of the modeling process. The practitioner specifies the model’s goals using two-point and high-order statistics. These goals include the normal score variogram model, indicator variogram models, cumulative run-length frequencies, and the n -point connectivity function. The parameters of the mapping function are inferred with stochastic optimization. These parameters result in a spatial mixture reproducing the statistics outlined in the modeling goals. The generation of conditioning data for latent factor realizations is an imputation problem, and these imputed factors become conditioning data for gridded factor realizations. Finally, the gridded factors are mapped to observed space with the inferred parameters, resulting in the final gridded realizations with the correct high-order statistics.

A small example highlights the ability of the NMR to capture connected extreme values and non-stationary features like orientation changes. A well-designed Gaussian pool can accommodate multiple orientations and anisotropies. The model significantly improves contained metal relative to an SGS model characterized by a two-point covariance structure. The following chapter presents the details of network components of the NMR framework, including latent factor design and parameter inference.

Chapter 4

Network Implementation

This chapter presents the implementation details of the NMR introduced in Chapter 3. The network is the first component of NMR framework and operates in conjunction with sequential Gaussian rejection imputation (SGRI), introduced in Chapter 5. Inference of network parameters, θ , is an inverse problem as only the true data values are known. Therefore, a mapping function, \mathcal{F}_θ , is required to map the unknown latent space to the known observed space. The function is parameterized such that a pool of unconditional, latent Gaussian factors have the desired spatial features when mapped with \mathcal{F}_θ . As the latent space is a synthetic feature of the NMR, it is free to contain any number of components with arbitrary covariance structure. This flexibility permits creativity with latent factor design; combining latent covariance structures in unique ways allows a mixture of univariate Gaussian distributions to possess non-multivariate Gaussian spatial features. Practical implementation details include considerations for the design of the network and the latent Gaussian pool, as well as sensitivities associated with these model parameters. A synthetic example demonstrates latent factor design, network parameter inference and non-uniqueness properties of the NMR. This example highlights the network component of the NMR workflow and is carried forward into Chapter 5 to illustrate imputation.

4.1 Network Design

The NMR is not a true neural network but rather a model of regionalization inspired by neural network structure. The network consists of an input layer where the number of latent factors determines the number of input nodes, a single “hidden layer” of the same dimension, and an output layer with a single node. The single output node makes the NMR univariate (analogous to the LMR); however, there is no reason it could not be extended to the multivariate case (analogous to the LMC). However, with geospatial data, one commonly considers extreme values in the univariate context; it is not immediately clear what constitutes a multivariate extreme.

4.1.1 Architecture

The number of latent factors in the Gaussian pool controls network architecture. The current NMR implementation restricts the network architecture to an input layer, a single hidden layer and a univariate output layer. The input and hidden layer both contain $M + 1$ nodes. Figure 4.1 shows an example network configuration with $M = 3$ input latent factors, $\{Y_0, \dots, Y_3\}$, where the nugget effect is the 0^{th} factor by convention. The network’s hidden layer is a transformation layer, transforming

the latent Gaussian values to arbitrary activation units. The network activation function $\phi(\dots)$ is a modified power-law (MPL) function; the following subsection presents the details. The network output Z_1 is a weighted linear combination of the activated latent factors.

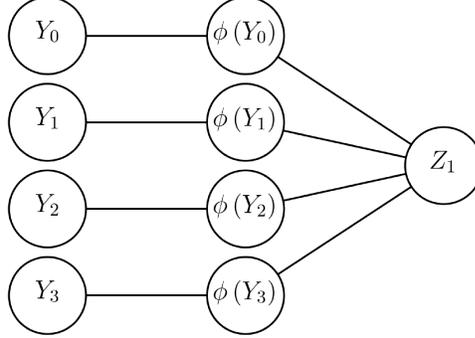


Figure 4.1: Schematic representation of the NMR with $M = 3$ latent factors. $\phi(\dots)$ is the MPL activation function. For clarity the ω parameter is excluded.

The following operators define the general forward pass through the network:

$$\mathbf{x} = \sum_{m=0}^M a_m \cdot \phi(\mathbf{y}_m, \omega_m) \quad (4.1)$$

$$\mathbf{z} = G^{-1}(F_X(\mathbf{x})) \quad (4.2)$$

where \mathbf{x} is an intermediary activation vector, a_m is a weight applied to factor m , $\phi(\mathbf{y}_m, \omega_m)$ is the MPL activation function with exponent ω applied to factor m , F_X is the CDF of \mathbf{x} and G^{-1} is the inverse of the Gaussian CDF. The weights $\{a_0, \dots, a_M\}$ are constrained to be greater than or equal to zero and reflect the relative importance of each factor in the mapping function. Alternatively, Equation 4.1 can be rewritten in matrix notation as:

$$\mathbf{x} = \phi(\mathbf{Y}, \boldsymbol{\omega}) \mathbf{a}^T \quad (4.3)$$

where \mathbf{a} is a row vector of factor weights and $\phi(\mathbf{Y}, \boldsymbol{\omega})$ is an $n_{data} \times M + 1$ matrix of activated latent factors. With the given architecture, the NMR requires inference of $2 \cdot (M + 1)$ parameters: a $1 \times M + 1$ dimensional vector of factor weights and a $1 \times M + 1$ dimensional vector of MPL exponents ω . These $2 \cdot (M + 1)$ network parameters are inferred through stochastic optimization and discussed in Section 4.3. The raw output of the network is activation units that must be transformed to Gaussian space, necessitating the normal score transform of \mathbf{x} to \mathbf{z} . The univariate normal score transform from activation units to Gaussian units in Equation 4.2 requires a representative distribution unique to the network's given state. Each iteration of the stochastic optimization algorithm generates an updated parameter vector, resulting in a new mapping. The iteration-specific reference distribution is created by drawing a $1e^4 \times M + 1$ dimensional matrix of independent, standard normal Gaussian values and mapping them to a $1e^4 \times 1$ vector of activation values. This mapping establishes the relationship between the latent and the activation space for the given parameter state. These

activation values are then normal score transformed, resulting in a temporary, iteration-specific transformation table. This transformation table establishes the relationship between the activation and the observed space for the given parameter state. Using this transformation table (acting as the G^{-1} operator in Equation 4.2), any activation value can be transformed to a Gaussian value.

4.1.2 Activation Function

The goal of the NMR is to parameterize the arbitrary mapping function, \mathcal{F}_θ , between the latent and observed spaces. The form of this function is not obvious; hence, a network-based approach is used as a function approximation. Given the complex spatial features we wish to capture in the final models, a polynomial of degree greater than one is useful. Suppose the activation function is linear. As the forward pass through the network is a weighted, linear combination of the inputs, a linear activation function (or simply $c \cdot \mathbf{y}$ where c is a constant) results in a linear output or single-order polynomial (Sharma, Sharma, et al., 2020). To achieve a non-linear network output, one must introduce a non-linearity as an activation function. Real data commonly contain non-linearly separable features, and a non-linear activation function permits the projection of these features onto a non-linear feature space (Dubey, Singh, et al., 2022).

In the traditional machine learning (ML) context, the constraint of differentiability is placed on neural network activation functions due to the use of the back-propagation algorithm (Rojas, 1996). The NMR structure is only inspired by a neural network, and is not subject to this constraint. NMR parameters are “learned” through gradient-free stochastic optimization. This gradient-free approach, and a strength of the NMR, permits using virtually any activation function, differentiable or not. The NMR activation function considered here is the modified power-law (MPL) function with the form:

$$\phi(\mathbf{y}, \omega) = \begin{cases} \mathbf{y}^\omega, & \text{if } \mathbf{y} > 0 \\ \mathbf{y}^{\frac{1}{\omega}}, & \text{if } \mathbf{y} < 0 \end{cases} \quad (4.4)$$

where ω is a trainable parameter. The magnitude of ω allows the activation function to emphasize certain regions of the latent distribution. If $\omega = 1$, the activation is linear and $\phi(\mathbf{y}, 1) = \mathbf{y}$. If ω is less than one, the function takes on a concave shape that emphasizes low values and mutes the influence of high values. When ω is greater than one, the opposite is true. The function takes on a convex shape that emphasizes high values and mutes the influence of low values. Figure 4.2 (left) shows the relationship between $\phi(\mathbf{y}, \omega)$ and \mathbf{y} for various values of ω . As the magnitude of ω increases, the activation function becomes steeper above zero and flatter below zero. The high values’ magnitude increases exponentially, and low values are muted significantly. This non-linear amplification allows the network to embed high-grade features of latent factors in the mapping function \mathcal{F}_θ . Low-grade features are embedded in opposite fashion.

Notably, the MPL activation has three inflection points: -1, 0, and 1. The inflections exist as

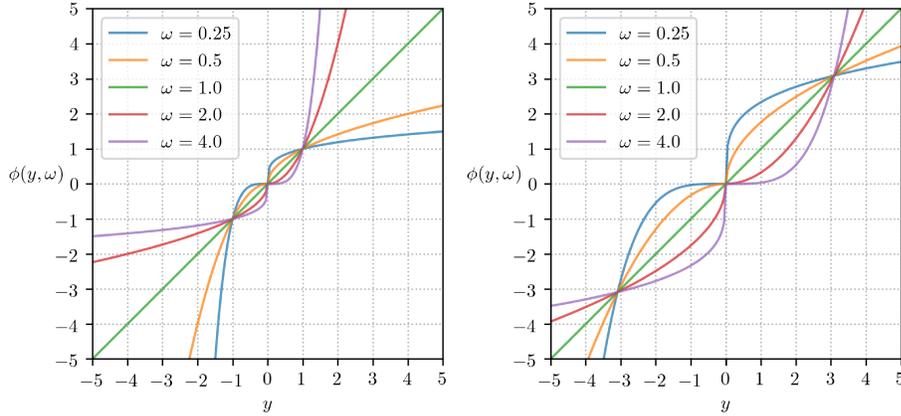


Figure 4.2: The MPL (left) and scaled MPL (right) activation function for various values of ω and input $\mathbf{y} \in [-5, 5]$. The scaled activation uses $\xi = G^{-1}(0.999) \approx 3.09$.

$\pm 1^\omega = \pm 1$ and $0^\omega = 0$, regardless of ω . No negative effects have been observed related to these inflections; however, the points at -1 and 1 may be adjusted by introducing a scaling factor, ξ , to the MPL activation function:

$$\phi(\mathbf{y}, \omega, \xi) = \begin{cases} \xi \cdot \left(\frac{\mathbf{y}}{\xi}\right)^\omega, & \text{if } \mathbf{y} > 0 \\ \xi \cdot \left(\frac{\mathbf{y}}{\xi}\right)^{\frac{1}{\omega}}, & \text{if } \mathbf{y} < 0 \end{cases} \quad (4.5)$$

The scaling parameter ξ allows setting a predetermined inflection point while retaining the linear nature of the function when $\omega = 1$. Figure 4.2 (right) shows the scaled MPL activation function where $\xi = G^{-1}(0.999) \approx 3.09$ and G^{-1} is the inverse of the Gaussian CDF. ξ effectively controls the threshold for dampening or emphasizing a factor:

- When $\omega > 1$ and $0 < y < \xi$, the MPL reduces the value of y slightly, similar to an opportunity-seeking risk perspective (Eidsvik, Mukerji, et al., 2015). More emphasis is placed on values of $y > \xi$.
- When $\omega < 1$ and $0 < y < \xi$, the MPL increases the value of y slightly, similar to a risk-averse perspective. More emphasis is placed on values of $y < \xi$.

The opposite relationships hold when $y < 0$. It is not immediately clear when the MPL should be scaled; however, if the practitioner notices artifacts related to the activation function, ξ can be tuned. ξ could be introduced as a trainable parameter. The range of ω for each latent factor is an important component of latent factor design, discussed in Section 4.2. Constraining $\omega_m > 1$ embeds higher-grade features of the m^{th} factor in the mapping. Conversely, constraining $\omega_m < 1$ embeds lower-grade features of the m^{th} factor in the mapping. Values of ω are practically within $[0.25, 4.0]$ and reflect the relative influence of each latent factor’s high and low values.

4.2 Latent Factor Design

Latent factor design is critical to the workflow of the NMR. The latent factors form the Gaussian “pool” from which the NMR draws structure. The univariate Gaussian factors are mixed to generate a spatial distribution that is univariate Gaussian but not multivariate Gaussian. The mixture distribution can only contain features, or a combination of features, present in the Gaussian pool. Therefore, any feature required in the final model must be in the pool. The Gaussian pool is similar to the concept of factorial kriging (Goovaerts, 1997), where the nested variogram model is a composition of L basic variogram structures, and the regionalized variable is a composition of L independent, standard normal, spatial components operating at different geologic scales:

$$\gamma(\mathbf{h}) = \sum_{\ell=1}^L \gamma_{\ell}(\mathbf{h}) \quad (4.6)$$

$$Z(\mathbf{u}) = \sum_{\ell=1}^L b_{\ell} Y_{\ell}(\mathbf{u}) + \mu(\mathbf{u}) \quad (4.7)$$

where the coefficient b_{ℓ} is the square root of the variance contribution of γ_{ℓ} . The key difference between the NMR and factorial kriging approaches is that the regionalized NMR variable is a weighted, linear combination of non-linear, independent spatial components as Equation 4.1 highlights. With the LMC or factorial kriging, the Y_{ℓ} independent factors are synthetic features of the model and are not directly observed; only characterized by γ_{ℓ} . The NMR requires explicit definition of a pool of independent Gaussian factors. The pool may contain any structures. It is straightforward to generate realizations of the independent Gaussian factors with any unconditional simulation algorithm such as SGS (Gómez-Hernández & Journel, 1993), LU simulation (Davis, 1987) or turning-bands (Journel, 1974).

The choice of covariance structures for the Gaussian pool must consider the end goal of the spatial mixture. The latent factors must be reasonable in the sense that achieving the objective is possible. For example, if the final goal is short-range features, the pool must contain short-range features. Long-range and short-range structures can produce a final mixture with medium-range features. The mixing of factors causes destructive interference; that is, feature ranges in the final mixture are shorter than their corresponding ranges in the pool. A pool of long-range structures could conceivably generate short-range features through destructive interference and noise from the nugget effect, though it is recommended the ranges of structures in the pool be slightly longer than the ranges of the model objectives.

Pool considerations include (1) the conceptual geological model, (2) the L nested components of the normal score variogram, (3) the L nested components of each indicator variogram model and their potential asymmetry, (4) the downhole connectivity measures from the observed data and the potential connectivity of extreme values, and (5) the composition of the objective function, discussed in more detail in Section 4.3. The factorial kriging concept provides a reasonable starting point for

the design of the Gaussian pool. An initial pool can be inferred by decomposing all variogram models into their basic components (as in Equation 4.6). Each basic variogram component is a single licit structure with three orientation parameters, three range parameters, a nugget of zero and a sill of one. By convention, a pure nugget latent factor is added as the 0^{th} factor.

Consider a small 2-D synthetic example where the goal of the NMR is to generate gridded realizations with strongly asymmetric 0.1 and 0.9 quantile indicator variograms. That is, the low-grade two-point spatial continuity differs drastically from the high-grade. This scenario is challenging for multivariate Gaussian simulation algorithms and has been discussed at length by many practitioners (Gómez-Hernández & Wen, 1998; Guthke, 2013; Journel, 1983; Journel & Deutsch, 1993; Renard, Straubhaar, et al., 2011). The maximum entropy characteristic of the multivariate Gaussian distribution tends towards disconnected extremes and maximum connectivity of intermediate values. The destructuring of indicators away from the median is symmetric in the multivariate Gaussian case. The NMR framework can overcome this challenge using a well-designed Gaussian pool and ω bounds. Consider two factors with the goal of asymmetric indicator variograms. Factor 1 is a highly anisotropic factor, oriented in a north-south direction, and factor 2 is isotropic with a range off $\approx \frac{1}{2}$ the domain size. The factors are activated using the MPL with $\omega_1 = 4$ emphasizing the high values and $\omega_2 = 1/\omega_1 = 0.25$ emphasizing the low values, followed by linear combination (Equation 4.1) and normal score transform. The top row of Figure 4.3 shows the factors in activated units (left and center) and the normal score transform of the mixture (right). The mixture model is univariate Gaussian but not multivariate Gaussian. Note the difference in activation units between the factors. The bottom row shows the 0.1 (black) and 0.9 (red) indicator variograms of the final NMR mixture in the north-south and east-west directions. The longer range, more-isotropic, low-grade continuity is preserved through factor 2 and $\omega < 1$, and the highly-anisotropic, higher-grade continuity is preserved through factor 1 and $\omega > 1$. This small example is illustrative of three key concepts concerning latent factor design:

1. Latent factors should be designed in conjunction with ω constraints, targeting continuity in specific grade ranges and the final model goals.
2. The ranges of factors are reduced through mixing. Mixing of factors cannot increase continuity beyond the longest-range structures in the pool. Factor 2 in Figure 4.3 (top row, center) has an isotropic variogram range of 64 meters. The 0.1 quantile indicator variogram range is reduced to roughly 40 and 25 meters in the north-south and east-west directions, respectively.
3. Anisotropy ratios of factors are affected through mixing. The anisotropy ratio of factor 1 is roughly unchanged in Figure 4.3 (top row, left); however, the anisotropy ratio of factor two is increased from 1:1 to $\approx 1.6:1$. In general, isotropic factors will tend to become more anisotropic, and anisotropic factors will become more isotropic.

The Gaussian pool must be reasonable concerning the optimization targets discussed in Section

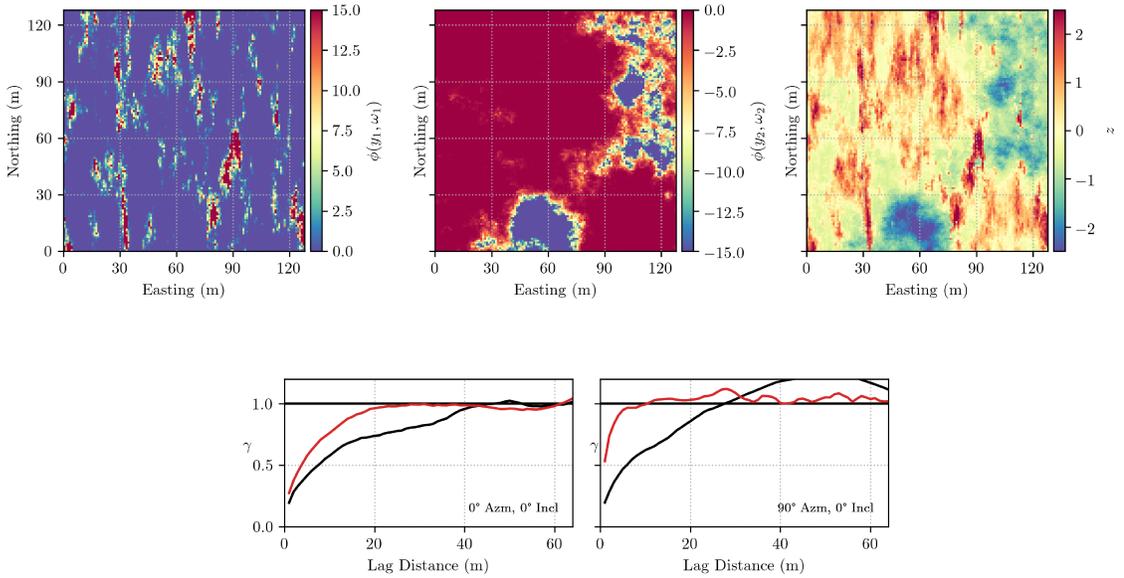


Figure 4.3: The top row shows factor 1 activated with $\omega_1 = 4$ (left), factor 2 activated with $\omega_2 = 1/\omega_1$ (center), and the normal score transform the mixture (right). The bottom row shows the gridded indicator variograms for the 0.1 quantile (black) and 0.9 quantile (red) in the north-south (left) and east-west (right) directions.

4.3. The inferred mapping function is non-unique, so there is no “correct” Gaussian pool; however, some latent factor combinations may be geologically unreasonable. For example, one cannot expect a Gaussian pool of long-range structures to well reproduce short-range features in the NMR model, and vice-versa. This is similarly true for factor orientations. If the final model requires spatial features in certain orientations, factor design should address this requirement. Factors aimed at generating high-order connectivity features should consider the major orientation of drilling. Factors with different orientations may be combined; for example, a north-south and an east-west factor can generate north-east striking features in the final model. This base-case requirement of “reasonableness” motivates variogram decomposition as the initial approach to latent factor design. There is no definitive recipe for latent factor design, though the following rules of thumb apply:

1. Decompose the nested variogram models into their basic components.
2. Check objective component reproduction with a traditional simulation algorithm (such as SGS) to provide insight into any “missing” features. Add factors to capture the missing features.
3. Add factors based on the conceptual geologic model and end goals of the NMR model.
4. Prune any redundant factors. That is, the pool is only as complex as required.
5. Anisotropy of factors is reduced through mixing. The NMR model is only as anisotropic as the most anisotropic factor. Increase the anisotropy ratio of factors if required.
6. The range of factors is reduced through mixing. That is, the range of NMR model is only as

long as the longest range factor. Increase the variogram ranges of factors if required.

The next step is initializing the objective function components after establishing a reasonable Gaussian pool. Optimizing the parameter vector θ occurs after this initialization. The following section discusses details of the objective function, optimization algorithm and NMR parameter inference.

4.3 Parameter Inference

The NMR approximates the mapping function, \mathcal{F}_θ , between latent and observed space. This function is parameterized by the unknown vector θ . The unknown parameters are inferred from known features: (1) the observed data, (2) the specified latent pool, and (3) the objective function components. Features (2) and (3) are specified by the user and guide the parameter optimization process. θ is a vector of $2 \cdot (M + 1)$ real values of factor weights and MPL exponents:

$$\theta = \{a_0, \dots, a_M, \omega_0, \dots, \omega_M\} \quad (4.8)$$

where θ is optimized with the heuristic, genetic algorithm DE (Price, 2013). A population of candidate solutions is initialized and then evolved to mimic natural selection. Each member of the population has an associated “fitness” value, and the fittest members of the population are carried over to subsequent generations. Through multiple generations of mutation and crossover operations, the candidate solutions converge towards a solution that minimizes the objective function. Parameter inference begins by simulating a set of $L \cdot (M + 1)$ unconditional realizations at the input data locations, where L is the number of realizations and $M + 1$ is the number of independent factors, including the nugget. These unconditional realizations permit the evaluation of the objective function and evolution of θ .

4.3.1 Differential Evolution (DE)

The optimization of θ uses a gradient-free, heuristic genetic algorithm. Gradient-free methods are typically employed when information about the derivative of the objective function is either costly to obtain or unreliable and noisy (Conn, Scheinberg, et al., 2009). Differential evolution (DE) is a global stochastic search algorithm that is practical for non-linear, non-differentiable objective functions with a necessarily large search space (Rios & Sahinidis, 2013). Any objective function is permissible, an advantage of DE.

DE is based on natural evolutionary processes where the fittest members of the population survive through a “natural selection” process. An initial population of size $NP \times D$ is generated by randomly sampling the objective function space within the defined constraints. NP is the number of individuals in the population, and D is the problem’s dimensionality. Each vector from the initial population is passed through the objective function to evaluate its evolutionary “fitness”. The basic

DE algorithm generates a mutant vector from the population by adding the scaled difference between two randomly selected vectors to a third randomly selected vector (Price, 2013). The algorithm then generates a trial vector by recombining the mutant vector with the initial population’s current row vector, considering a user-defined crossover probability. The trial vector’s fitness is compared to the current population vector’s fitness in an evolutionary sense. If the trial vector’s fitness exceeds the current population vector’s, it replaces it. Each iteration compares all population vectors to a randomly generated trial vector and accepts the trial vector if its fitness exceeds the current vector. Each algorithm iteration’s “surviving” vectors become the parent vectors of the next iteration population. The following general steps summarize the DE algorithm:

1. Initialize counter $i = 0$
2. Initialize a random population of size $j = 1, \dots, NP$
3. If $i < i_{MAX}$:
 - a) $i = i + 1$
 - b) For each member of the population, \mathbf{x}_j :
 - i. Select r other individuals where $\mathbf{x}_r \neq \mathbf{x}_j$ and r is unique
 - ii. Generate mutant \mathbf{v}_j
 - iii. Generate trial \mathbf{t}_j through crossover between \mathbf{x}_j and \mathbf{v}_j
 - iv. If $f(\mathbf{t}_j) < f(\mathbf{x}_j)$ replace \mathbf{x}_j with \mathbf{t}_j
 - c) Finish if $i = i_{MAX}$
4. Return $\text{argmin}_j f(\mathbf{t}_j)$

The algorithm begins by randomly initializing a population of candidate vectors (Equation 4.8) and evaluating each member’s fitness. A population size of 30-50 is a reasonable balance between sufficient diversity and algorithm runtime, though there is no definitive guideline for population size (Balkaya, Ekinici, et al., 2017; Piotrowski, 2017). Over the specified number of iterations, each member of the population undergoes mutation. The NMR implementation employs a DE/current-to-best/1 mutation strategy. Georgioudakis and Plevris (2020) describes the mutated vector as:

$$\mathbf{v}_j = \mathbf{x}_j + F(\mathbf{x}_{best} - \mathbf{x}_j) + F(\mathbf{x}_{r1} - \mathbf{x}_{r2}) \quad (4.9)$$

where $r1 \neq r2 \neq j$, F is the scaling factor controlling the amplification of the difference vectors and \mathbf{x}_{best} is the current best vector in the population. The mutation process learns from the current best vector (local searching of the solution space) while exploring the global search space through the randomly selected difference vector. The parameter F is analogous to the learning rate in machine learning problems. A smaller F value will lead to smaller mutation step sizes, and the algorithm will take longer to converge. Larger F values increase the degree of solution exploration but may lead to divergence. Price (2013) suggests there is no upper limit for F , however effective values are almost always $F < 1.0$.

This new mutated “genetic information” is crossed over to other population members based on a crossover probability, CR , and a binomial crossover scheme. The trial vector is built from two vectors: the mutant and another member of the population. Price (2013) describes the crossover as:

$$\mathbf{t}_j = \begin{cases} \mathbf{v}_{j,k}, & \text{if } rand_k(0,1) \leq CR \text{ or } k = k_{rand} \\ \mathbf{x}_{j,k}, & \text{otherwise} \end{cases} \quad (4.10)$$

where \mathbf{t}_j is the trial vector, $\mathbf{x}_{j,k}$ is the target vector, k denotes the 1-D vector index of the j^{th} member of the population, $rand_k(0,1)$ is a uniform random number $\in [0,1]$, and k_{rand} is a random index. For each element in the target vector, if the uniform random number is less than CR , the element is copied from the mutant vector $\mathbf{v}_{j,k}$; otherwise, the target element remains. The random index k_{rand} ensures the target vector is not copied completely; at least one element of the mutant vector passes to the trial vector. CR influences the diversity of the evolving population. Larger CR values introduce more variation in the population, resulting in a greater search of the global solution space, while smaller values may lead to stagnation (Georgioudakis & Plevris, 2020). Optimization employs a non-linear crossover scheme following the work of Mohamed (2014):

$$CR = CR_{hi} + (CR_{lo} - CR_{hi}) \cdot (1 - i/N)^4, \quad i = 1, \dots, N \quad (4.11)$$

The idea behind the non-linear crossover scheme is that the crossover rate is lower when population variance is high in early generations. This crossover scheme prevents extreme diversity or potential divergence in the initial iterations. In subsequent generations, the population variance decreases as the vectors become similar, approaching the solution. In order to thoroughly explore this more local search space, the crossover should be high, encouraging a diverse population of “good” solutions.

4.3.2 Objective Function

The parameter vector θ is optimized heuristically through an objective function and DE. This approach is highly flexible as the objective function can take any form with any number of components. The practitioner provides target spatial features of the final mixture from which a loss or objective value can be calculated for the candidate mixture. The objective value is minimized through successive iterations reproducing the target spatial features. The possible objective function components are (1) the continuous variogram, (2) indicator variograms, (3) cumulative run length frequencies, and (4) the n -point connectivity function. Components may have multiple sub-components for each indicator threshold. Run length frequencies and the n -point connectivity function are higher-order multi-point statistics that better characterize non-Gaussian features than the two-point variogram. The objective function comprises a weighted combination of $C \leq 4$ objective components:

$$O = \sum_{c=1}^C w_c O_c \quad (4.12)$$

where w_c is the component weight and O_c is the component objective value. The objective function quantifies how different the desired feature is from the target feature. Weighting the objective function components is required as each component may exist in widely different units. For example a variogram component is expressed in units of variance squared ($\sum[\gamma_{\text{target}} - \gamma_{\text{realization}}]^2$) while cumulative runs is expressed as the number of runs squared ($\sum[R_{\text{target}} - R_{\text{realization}}]^2$), and may be orders of magnitude different. A weighting scheme to prevent one component from dominating the objective value follows the work of Deutsch (1992). The goal is to have each component contribute equally to the overall objective value where each weight is inversely proportional to the average change of that objective component (Deutsch, 1992):

$$w_c = \frac{1}{|\Delta\bar{O}_c|} \quad (4.13)$$

The average change of the objective component is approximated numerically by averaging the change of $J = 1000$ independent forward passes through the network:

$$|\Delta\bar{O}_c| = \frac{1}{J} \sum_{j=1}^J |O_c^j - O_c|, \quad c = 1, \dots, C \quad (4.14)$$

where $|\Delta\bar{O}_c|$ is the average change of component c , O_c^j is the updated objective value for iteration j and O_c is the initial objective value of that component. The objective value is the sum of squared errors between the experimental values calculated on the distribution derived from Equation 4.1 and the initialized target values. The objective function (Equation 4.12) is evaluated across all simulated realizations and minimized in expectation. Though many components can enter the objective function, one should consider contrasting objectives. For example, one cannot expect to reproduce continuous and indicator variograms closely if the principal directions of continuity are orthogonal. It is less clear how the specification of run frequencies or connectivity functions affect variograms, though there are likely confounding factors.

4.3.3 Precedence

The application of precedence is an optional component of parameter inference. Precedence allows the spatial features of a certain latent factor to take priority in the mapping function. A particular factor can be given precedence during optimization if one desires its spatial features in certain portions of the grade range. The algorithm enforces precedence by employing a sigmoid weighting function where $s(y) = \frac{1}{1+e^{-y}}$. The weighting function modifies the forward pass (Equation 4.1) through the network to:

$$\mathbf{x} = a_p \cdot \phi(\mathbf{y}_p, \omega_p) + \sum_{m=0}^{M-1} \phi(\mathbf{y}_m, \omega_m) \cdot s(\mathbf{y}_p \cdot x), \quad m \neq p \quad (4.15)$$

$$\mathbf{z} = G^{-1}(F_X(\mathbf{x})) \quad (4.16)$$

where a is the weight to the factor, subscript p is the factor index with precedence, $\phi(\dots)$ is

the MPL activation function, $s(\dots)$ is the sigmoid function, and x is a constant $\in [-10, 10]$. The sign and magnitude of x controls what part of the grade range factor \mathbf{y}_p influences. If $x < 0$, \mathbf{y}_p influences high values and, $x > 0$, \mathbf{y}_p influences low values. Figure 4.4 shows a weighting function where $x = -1.5$. The y-axis is the weight to the remaining $M - 1$ factors for the range of values of \mathbf{y}_p . In this scenario, \mathbf{y}_p receives $\approx 95\%$ of the weight when equal to 2.0, $\approx 80\%$ of the weight when equal to 1.0 and so on. As the magnitude of x increases, $s(\dots)$ tends towards a binary step function centered on zero.

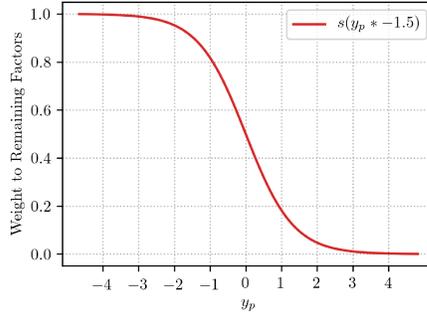


Figure 4.4: Sigmoid weighting function where $x = -1.5$. The x-axis is the range of \mathbf{y}_p , and the y-axis is the weight to the remaining M-1 factors.

The sigmoid weighting function allows a certain factor to take precedence where the observed data takes on either high or low values. This increased control is particularly useful for capturing local high- or low-grade continuity with well-designed latent factors.

4.3.4 Complete Algorithm

Algorithm 1 summarizes all components of the NMR parameter inference workflow. The first step is the unconditional simulation of the latent Gaussian pool at the data locations (lines 2-6). The objective function components are scaled using the unconditional realizations (lines 7-10) prior to heuristically optimizing θ with DE (lines 11-35).

Inference of θ is relatively straightforward with only a handful of DE parameters. However, care must be given to align the design of the latent pool and objective components with the final goals of the NMR model. Depending on the modeling scenario, this is the most challenging portion of the workflow.

4.3.5 Checks

The quality of the NMR parameter inference is measured based on reproducing the objective function components. Generally, these components come directly from the observed data, so the inferred parameters reproduce observed geologic features. A small 3-D synthetic example with 746 data is generated to highlight the concepts of parameter checking. The data set is simulated so that the

Algorithm 1 NMR parameter inference pseudocode.

```

1: Simulate unconditional realizations (LU or SGS):
2: for  $\ell = 1, \dots, L$  do
3:   for  $m = 1, \dots, M$  do
4:     Simulate  $\mathbf{Y}_m^{(\ell)}$ 
5:   end for
6: end for
7: Scale objective components:
8: for  $c = 1, \dots, C$  do
9:    $w_c = \frac{1}{|\Delta O_c|}$ 
10: end for
11: Initialize population:  $pop(popdim, popsize)$  ▷ Candidate solutions
12: Calculate fitness of population:  $fpop$ 
13:  $best = minloc(fpop)$  ▷ Index of initial best solution
14: for  $i=1, \dots, its$  do ▷ Begin DE
15:   for  $j=1, \dots, popsize$  do
16:      $mutant = MUTATION(pop(j))$  ▷ Best-to-current mutation
17:      $\theta = CROSSOVER(mutant)$  ▷ Binomial crossover
18:      $a, \omega = VECTOR\_TO\_MATRICES(\theta)$  ▷ Reshape  $\theta$  to  $a, \omega$  vectors
19:      $O = 0.0$ 
20:     for  $\ell = 1, \dots, L$  do
21:        $z = NETWORK\_FORWARD(a, \omega)$  ▷ Forward pass through network
22:        $iz = INDICATOR\_TRANSFORM(z, thresholds)$ 
23:       for  $c = 1, \dots, C$  do
24:          $O_c = \sum [\gamma_c^{target} - \gamma_c^{realization}]^2$ 
25:       end for
26:        $O = \sum_{c=1}^C w_c \cdot O_c$ 
27:     end for
28:      $O = O/L$ 
29:     if  $O < fpop(best)$  then
30:        $fpop(best) = O$  ▷ Update objective value
31:        $pop(best) = \theta$  ▷ Retain the trial
32:        $best = j$  ▷ Track the new best
33:     end if
34:   end for
35: end for ▷ End DE

```

indicator variograms are asymmetric about the median and have connectivity of high-grade values. These features provide a reasonable test case for applying the NMR workflow. Imputation concepts are presented using the same data set in Chapter 5. Figure 4.5 shows a plan-view and long-section through the synthetic data in Gaussian units. Figure 4.6 shows experimental variogram points and fitted models for the normal score variable (a), the 0.1 indicator (b) and the 0.9 indicator (c). For brevity, the median indicator variogram is not shown; however, the indicator variograms show asymmetric destructuring about the median. Figure 4.7 shows the n -point connectivity function, which approaches zero at five connected steps ($\approx 50\text{m}$). The objective function components are the normal score variogram model, the 0.1 indicator variogram model and the 0.9 indicator variograms (Figure 4.6 (a), (b), and (c)), and the n -point connectivity function (Figure 4.7).

The latent factor pool contains two structures: the long-range structure from the 0.1 quantile

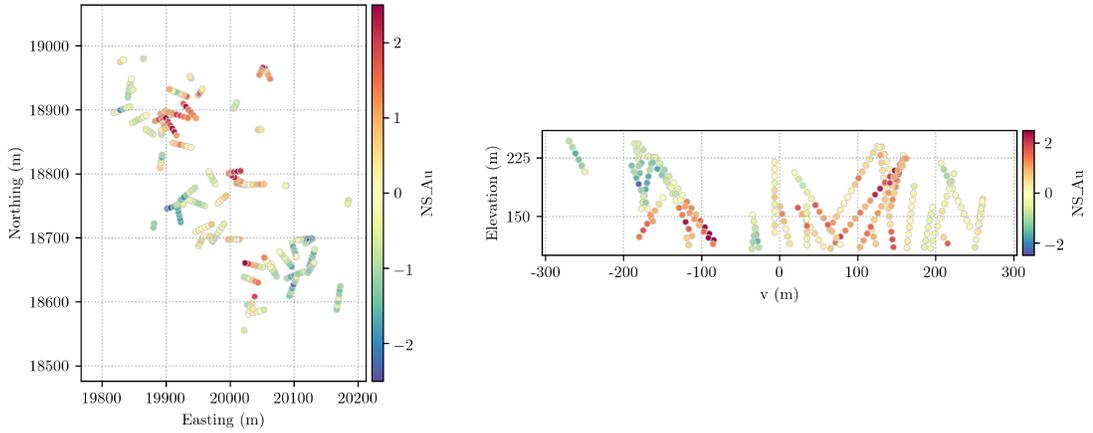


Figure 4.5: Plan-view (left) and long-section oriented 340° (right) through the synthetic data set. Both sections have a tolerance of ± 25 meters.

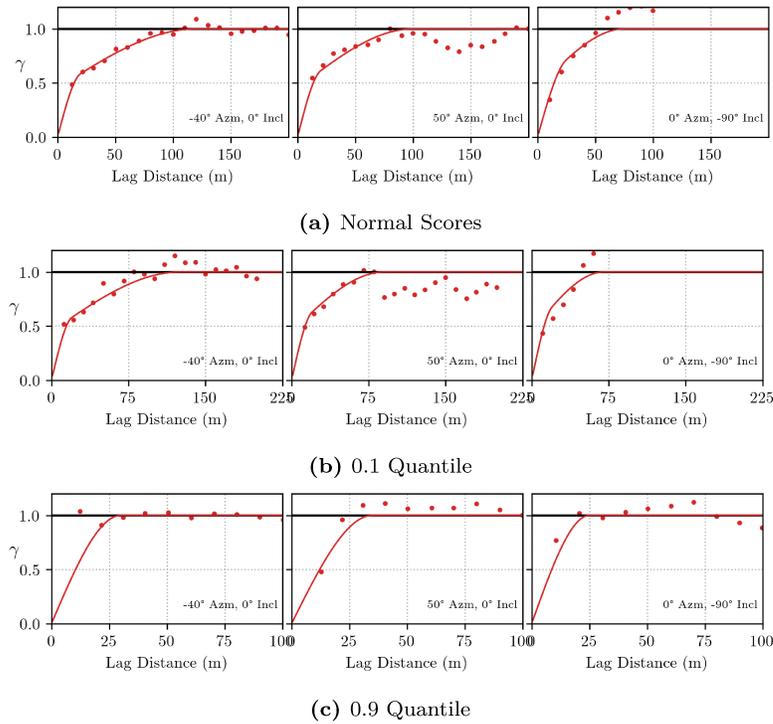


Figure 4.6: Fitted experimental variograms for the normal score variable (a), the 0.1 indicator (b) and 0.9 indicator (c). Note the asymmetric destructuring of the indicator variograms.

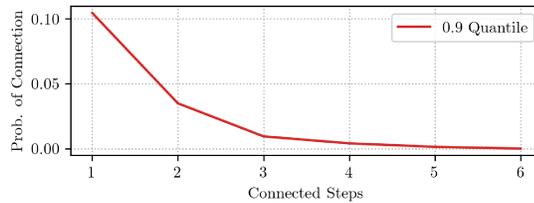


Figure 4.7: N -point connectivity function above the 0.9 quantile indicator for the synthetic data set.

indicator model and the single structure from the 0.9 quantile indicator model. The goal of the pool is to have the long-range structure spatially control the lower grades and the short-range structure spatially control the highs. Some degree of mixing between the structures permits the reproduction of the normal score variogram model. The omega parameters are constrained such that $0.25 \leq \omega_1 \leq 1.0$ and $2.0 \leq \omega_2 \leq 4.0$. No precedence is given to either factor. The algorithm is run for 1500 DE iterations. The primary NMR parameter check is the reproduction of objective function components. The mapped unconditional realizations, $\mathcal{F}_\theta(\mathbf{Y})$, must reproduce, on average, the input objective targets. Figure 4.9 shows normal score variogram reproduction for the inferred mapping function. Reproduction is reasonable though there is some deviation at shorter lags. This deviation is attributed to these lags having fewer pairs. Figure 4.9 shows variogram reproduction for the 0.1 (a) and 0.9 (b) quantile indicators. The indicators show a wider band of uncertainty related to data density, particularly at shorter lag distances with fewer pairs. Figure 4.10 shows n -point connectivity reproduction for the 0.9 quantile indicator.

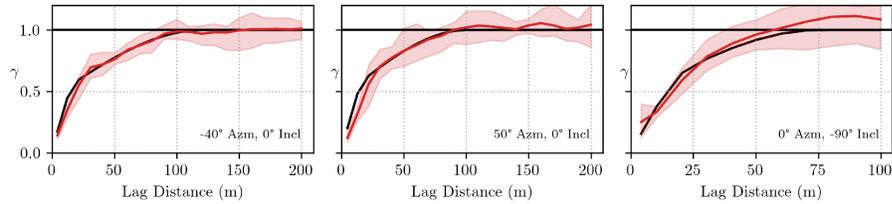


Figure 4.8: Continuous variogram reproduction for 25 realizations. The black line is the variogram model, the red line is the average variogram, and the shaded red area encloses the minimum and maximum variogram values. Left to right are the major, minor and vertical directions, respectively.

At this point in the NMR workflow, the inferred parameters reasonably approximate the mapping function between latent and observed space. The unconditional Gaussian realizations reproduce the objective function components on average when mapped through \mathcal{F}_θ . Any deviations between the optimization targets and mapped latent factors will manifest in the final NMR realizations as latent factor imputation (Chapter 5) is anchored on \mathcal{F}_θ . If present, these deviations are transferred to the imputed data realizations that condition the gridded factor realizations. This same synthetic data configuration is used to check imputed realization in Chapter 5.

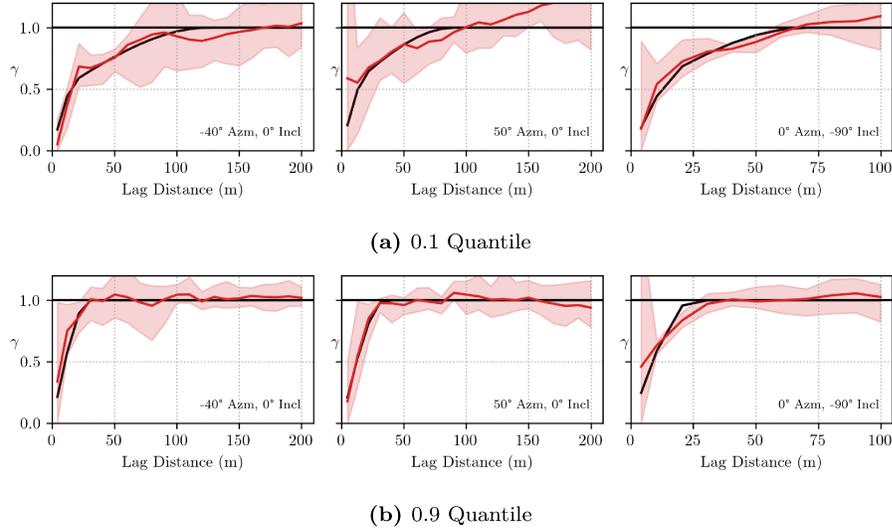


Figure 4.9: Indicator variogram reproduction for 25 realizations for the 0.1 (a) and 0.9 (b) quantiles. Left to right are the major, minor and vertical directions, respectively. The black line is the variogram model, the red line is the average variogram, and the shaded red area encloses the minimum and maximum variogram values.

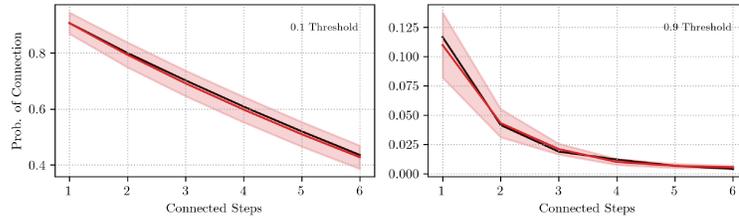


Figure 4.10: N -point connectivity function reproduction for 25 realizations for the 0.1 quantile indicator (left) and 0.9 quantile indicator (right). The black line is the target model, the red line is the average experimental value, and the shaded red area encloses the minimum and maximum experimental values.

4.4 Implementation Details

The goal of the NMR model is to infer the best possible mapping between latent and observed spaces. Given the data configuration, the design of the latent pool and the objective function components, there may be qualitative and quantitative uncertainty in the parameter vector θ . This section discusses some practical implementation details of NMR parameter inference that may contribute to or mitigate parameter uncertainty. Topics in this section include the non-uniqueness of the inferred solution, the number of data, the possibility of conflicting objectives, and overall computational considerations.

4.4.1 Non-Uniqueness

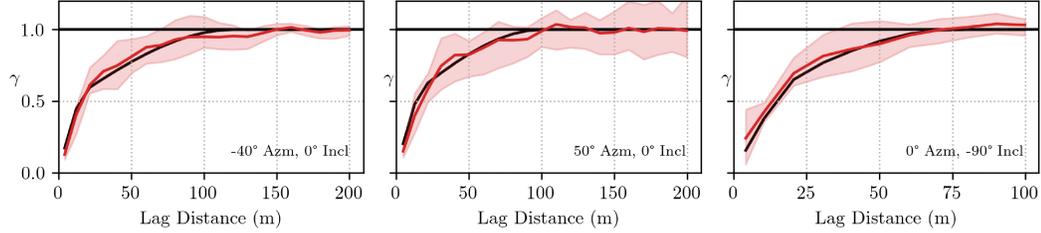
Solutions to inverse problems are often non-unique. Experimental variograms in the presence of sparse data are inherently uncertain statistics. Considering these facts, it is reasonable to assume that there are multiple (possibly infinite) network configurations and pools of Gaussian factors

that approximate the objective function targets. This non-uniqueness emphasizes the importance of considering the conceptual geological model when designing latent factors. Some latent pools are feasible from a numerical perspective but should also be geologically reasonable. The following section highlights non-uniqueness examples in NMR parameter inference. Potential non-unique scenarios are summarized as follows:

1. The NMR solution may be an arbitrary mixture of factors even if the variogram objective targets exist within the Gaussian pool.
2. A mix of short and long-range factors can reproduce a medium-range target.
3. A mix of only short-range factors cannot reproduce a medium or long-range target, and vice versa.
4. A mix of factors with NS and EW orientations can reproduce a target orientation of NNE, NNW, SSE, SSW, and so on.

Consider the same data configuration from Section 4.3.5. Synthetic data values are simulated using the NMR with a known parameter vector, θ . For clarity of the example, only the reproduction of the continuous variogram is measured. Different pools of latent factors are mixed to highlight the non-uniqueness scenarios mentioned above. The weight to each factor, a_m , is interpreted as the relative importance of each factor to the fit of θ . This weight can be compared to the known weight used to simulate the data to measure similarity. The solution is non-unique if multiple latent pools and varying parameters can reproduce the target. Another approach to understanding the dependence of each latent factor on θ is a measure of permutation feature importance (PFI) (Fisher, Rudin, et al., 2019). Feature importance can be calculated by permuting the input latent factors and calculating the increase in error in the output of the fitted model. PFI is in the same units as the objective function. Features that are pertinent to the network output will show more significant errors when they are permuted. A caveat is that shorter-range structures (more random) will always show less feature importance than longer-range structures (less random) as permutation introduces randomness. Shorter range structures may still contribute to the final fit of the model without showing high permutation feature importance.

Consider the first scenario listed above. The latent pool consists of the long-range structure of the normal score variogram model, the long-range structure of the 0.1 quantile indicator variogram model, the single structure of the 0.9 quantile indicator variogram model, and the nugget effect for $M = 3 + 1$ latent factors. Intuition suggests that if that variogram model (or its elemental components) exists within the Gaussian pool, the network will filter all irrelevant factors and weight the important components appropriately. In practice, this is only sometimes the case. Table 4.11b shows the a and ω elements of θ , plus PFI for comparison. Figure 4.11a shows the corresponding variogram reproduction. The a 's show how much weight is given to factors two and three that correspond to the nested structures of the indicator variograms. Only and portion of the weight is



(a)

	a	ω	PFI
Factor			
0	0.011	1.000	0.001
1	0.352	2.559	0.996
2	0.682	2.837	7.134
3	0.894	1.462	5.629

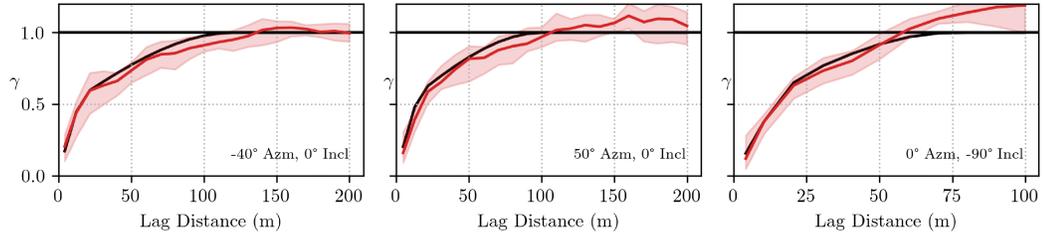
(b)

Figure 4.11: Normal score variogram reproduction (a) in the major, minor and vertical directions, respectively, and the components of θ (b) for the first scenario. PFI=permutation feature importance.

given to factor 1. The PFI backs up this weighting, showing factors two and three most important. Granted, there is a similarity between the low-grade indicator and normal score variogram models; the resulting variogram of the mixture model may be an arbitrary mixture of latent factors.

The second scenario is intuitive. A mix of short-range and long-range factors can reproduce features with intermediate-range. Again, this scenario is non-unique as one may consider any number of factors and variogram ranges are continuous and essentially unbounded. Consider a pool with two latent structures: one long-range and one short-range. The long-range structure corresponds to the second structure of the 0.1 quantile indicator variogram scaled by a factor of 1.5 in all directions. The short-range structure is the 0.9 quantile indicator model scaled by a factor of 0.7 in all directions. Table 4.12b shows the a and ω elements of θ , plus PFI for comparison. Figure 4.12a shows the corresponding variogram reproduction. The mixture of long and short-range structures can reproduce the variogram model of intermediate range with each factor receiving similar weight. This concept is also applicable to the orientation of factors in the pool. For example, the mix of factors with east-west and north-south orientations can generate features with intermediate orientation in the final model.

The third scenario is also straightforward. A mix of short-range factors cannot reproduce a longer-range target as destructive interference occurs when mixing factors. The range of structures in the final model can only be as long as the longest range structure in the latent pool. The opposite of this is also true; a mix of long-range structures will not closely reproduce a short-range target. In this scenario, the pool consists of the two nested structures of the normal score variogram model scaled by a factor of 0.3 in all directions. Table 4.12b shows the a and ω elements of θ , plus PFI for comparison. Figure 4.12a shows the corresponding variogram reproduction. As expected,

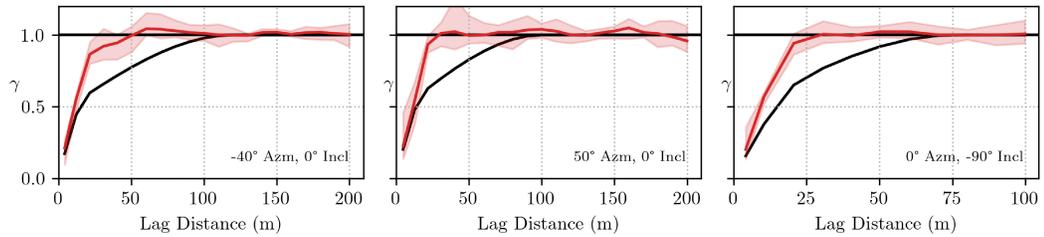


(a)

	a	ω	PFI
Factor			
0	0.162	1.000	0.118
1	0.737	1.409	12.524
2	0.662	0.811	4.639

(b)

Figure 4.12: Normal score variogram reproduction (a) in the major, minor and vertical directions, respectively, and the components of θ (b) for the second scenario. PFI=permutation feature importance.



(a)

	a	ω	PFI
Factor			
0	0.000	1.000	0.000
1	0.000	3.755	-0.000
2	0.335	1.378	5.029

(b)

Figure 4.13: Normal score variogram reproduction (a) in the major, minor and vertical directions, respectively, and the components of θ (b) for the third scenario. PFI=permutation feature importance.

the algorithm cannot converge on an acceptable solution due to poor latent feature design. The algorithm attempts to filter the first unnecessary factor by giving it a weight of zero, though the remaining factors do not have sufficient flexibility to reproduce the target.

This section presents scenarios to emphasize the non-uniqueness of the NMR solution. Numerous latent pools can reproduce objective targets. The final example presents a scenario where the design of the pool and the objective function have conflicting objectives. The scenario is designed for illustrative purposes and defies any geologic-based logic; however, it emphasizes the practitioner’s role in ensuring the latent pool and objective function are sound concerning the conceptual geology. Adding components to the objective function amplifies this issue. The requirement of “reasonable-

ness” further promotes the decomposition of all variogram targets into an initial latent Gaussian pool.

4.4.2 Number of Data

The number of data is an important consideration in the NMR workflow. Experimental statistics form the basis of the objective function components and are sensitive to the number of available data. Experimental variograms are an inherently uncertain statistic, particularly in the presence of sparse data (Ortiz & Deutsch, 2002; Pardo-Igúzquiza & Olea, 2012). The variogram value at lag vector \mathbf{h} is the mean of the squared differences between data values separated by \mathbf{h} . This mean value depends on the number of data, $n_{\mathbf{h}}$, entering the calculation, which depends on the data configuration and variogram tolerance parameters. Uncertainty in the experimental variogram points transfers to uncertainty in chosen model parameters, though this is only quantifiable with knowledge of the true variogram model. Sequence objective components are calculated downhole and are more sensitive to the total number of drillholes than the total data. However, the shape of the global distribution of runs or the global n -point connectivity function ultimately depends on the number of data.

A synthetic model is simulated to assess the sensitivity of θ to data spacing or the total number of available data. The model is simulated on a regular 56 x 56 x 56 m 3-D grid with a resolution of 1 m. The grid is sampled with regular, square data configurations ranging from 3 x 3 m to 20 x 20 m, with 1 m spacing in the vertical direction. Table 4.14a shows the resampled data configurations and the corresponding number of data. The grid is also sampled at a 1 x 1 m spacing, a reference distribution for calculating the “true” variograms, distributions of runs, and n -point connectivity functions. A parameter vector θ is inferred for each data spacing, and the unconditional latent factors are mapped to the observed space. The objective function value is the weighted sum of squared errors (Equation 4.12) between the mapped latent values and the objective components calculated from the reference distribution. Figure 4.14b shows the relationship between the square data spacings and the objective function value.

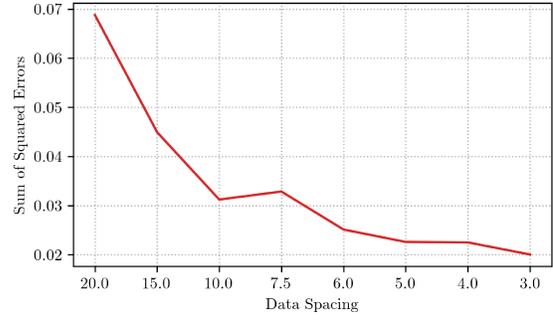
The sum of the squared errors is inversely proportional to the total number of data, which is the anticipated response. The curve decreases quickly to a spacing of 10 x 10 m or approximately 2000 data. Beyond this spacing, the curve is much flatter but does continue to decrease towards the tightest spacing. This relationship suggests that somewhere between 2000 and 5000 data are enough to provide a stable inference of θ . Beyond 5000 data, the error decreases slightly, though these improvements may be negligible when considering the increase in computation time.

4.4.3 Computational Considerations

The NMR objective function is a computationally expensive calculation. This expense is primarily due to experimental variogram calculation. As each population vector is an entirely new network,

	Number of Data
3.0m	20,216
4.0m	10,976
5.0m	8,064
6.0m	5,600
7.5m	3,584
10.0m	2,016
15.0m	896
20.0m	504

(a)



(b)

Figure 4.14: Square data spacings and corresponding numbers of data (a) and corresponding minimum sum of squared error values from NMR parameter inference(b). The total number of data increases to the right of the plot in (b).

all experimental variogram pairs must be updated on each iteration. Depending on the data configuration, this updating may account for a significant portion of the algorithm run time. Two straightforward approaches to dealing with runtime are:

1. Parallel differential evolution (PDE)
2. Constraining the maximum number of experimental variogram pairs per lag

Evaluating the objective function for each member of the population in DE is an independent task and lends itself to parallelization. PDE is a slightly different algorithm than DE. In DE, the population is updated after evaluating each trial vector (loop on line 15 in Algorithm 1), so each subsequent mutation includes information from the previous. In PDE, the entire population is mutated and evaluated before updating the next generation. This operation introduces new genetic information to the population once per generation rather than after every mutation. Though slightly different algorithms, DE and PDE produce similar, but different, final results. Figure 4.15 shows the objective function value versus iteration for two optimization runs: serial and parallel. Both scenarios use the same number of data, objective components and hyperparameters. PDE achieves a similar objective function value to serial DE, but ≈ 6.5 times quicker.

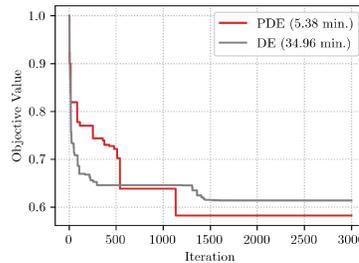


Figure 4.15: Objective function value versus iterations for runs of the DE and PDE algorithms.

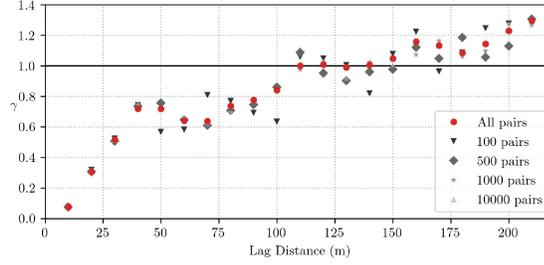


Figure 4.16: Sensitivity of experimental variogram points to the total number of pairs per lag.

The second potential speedup comes from setting a maximum number of randomly selected experimental variogram pairs to consider per lag when evaluating the objective function. Figure 4.16 shows that restricting the maximum number of pairs per lag to ≈ 10000 yields a stable variogram relative to all possible pairs. Restricting the pairs speeds up the objective function calculation, though one should check variogram stability for each data configuration.

4.5 Discussion

The NMR structure is inspired by neural networks from ML. The NMR approximates the mapping function, \mathcal{F}_θ , from latent to observed space. The parameter vector, θ , contains the weights applied to each latent factor and the corresponding MPL activation function exponents. θ permits the mixture of multivariate Gaussian spatial distributions in ways that reproduce non-Gaussian spatial features. The components of θ , a and ω , are optimized stochastically with the heuristic, genetic global optimization algorithm DE. A significant advantage of the genetic algorithm approach is the flexibility of both the activation and the loss or objective function. There is no constraint of differentiability on either the activation or objective function as DE is a gradient-free algorithm. The objective function may contain any component deemed necessary for the modeling scenario. The MPL activation function allows the influence of certain latent spatial features on certain portions of the grade range of the final mixture model. This activation function is key to achieving asymmetric spatial continuity between highs and lows. Some latent factors may only affect high and some low grades, conditional on constraints placed on ω . The practitioner has significant flexibility concerning latent factor design in both number and covariance structure. The design of the latent factors must consider the conceptual geologic model and be reasonable in the context of the modeling goals. Poorly designed factors may conflict with objective function components, preventing algorithm convergence; in some cases, reproducing the objective components may be impossible with the chosen factors. Decomposition of the nested variogram structures of the objective function components is a reasonable initial pool, provided one is confident in the variogram interpretation. One should prune redundant factors and add specific factors as required.

This chapter has introduced the first component of the NMR simulation workflow: inference

of the mapping function \mathcal{F}_θ . The following chapter introduces the second workflow component: imputation of latent factors. The ultimate modeling goal is gridded, non-Gaussian realizations for mine planning. Conditioning those realizations requires data. Data realizations are imputed for each latent factor such that they have the same spatial properties as the mixture model when mapped through \mathcal{F}_θ . The synthetic example introduced in Section 4.3.5 is carried forward into Chapter 5.

Chapter 5

Latent Factor Imputation

This chapter presents a novel algorithm for imputing latent factors called sequential Gaussian rejection imputation (SGRI). Imputation of latent factors addresses a problem where all values are missing or unknown. Conditioning gridded realizations of latent factors requires assigning valid latent values at the data locations. The goal is to impute latent factors with a unique covariance structure that, when combined with the inferred NMR mapping, returns the true observed values at the data locations within a tolerance. Traditionally, this problem has been approached through Gibbs sampling, though there are noted challenges in achieving stable convergence with spatially correlated variables (Silva, 2018). SGRI is a straightforward approach that combines SGS, rejection sampling and exact data matching. Conditional moments are calculated with SGS, ensuring the spatial relationships are correct and iterative rejection sampling ensures the collocated multivariate relationships are correct. Realizations of latent factors are generated and become conditioning data for gridded realizations. When mapped to observed space, the realizations reproduce non-Gaussian spatial features specified by the NMR and uni- and multivariate statistics. A small synthetic example continued from Chapter 4 demonstrates imputing latent factors, conditional simulation and mapping the latent space to observed space.

5.1 Imputation Concepts

Imputation is a method to “fill in” missing values (Little & Rubin, 2019). Missing values may be uni- or multivariate, and multiple mechanisms or patterns of missingness may be possible. Multivariate transformations, common in modern geostatistical workflows, such as PPMT (Barnett, Manchuk, et al., 2014), require homotopic sampling, necessitating imputation methods. Ignoring these missing values results in a loss of information or potential bias if the missingness mechanism is not random. The simplest deterministic approach to imputation is taking the global mean or median of sampled values or employing a regression model. This inference permits using all data in subsequent modeling; however, it captures no uncertainty in the imputed values.

Single imputation involves imputing a single value for each missing data value. Little and Rubin (2019) describes single imputation techniques as (1) mean imputation, where the global mean value is substituted; (2) regression imputation, where a predicted value from the regression of the missing variable on the observed variables replaces missing values; and (3) stochastic regression imputation where a predicted value plus a residual replace missing values. The single imputation paradigm indicates that the missing values are certain or constant. Though this imputation approach includes

all data in statistical analysis, this assumption results in incorrect uncertainty as the true values are unknown. Furthermore, datasets imputed with mean or regression imputation will not have the correct mean or variance (Barnett & Deutsch, 2015). For these reasons, multiple imputation techniques are preferred when characterizing imputation uncertainty, which is important.

Multiple imputation involves generating realizations of missing values, allowing assessment of imputation uncertainty. A model of the conditional distribution of the missing values given the observed values is inferred and then stochastically sampled, resulting in complete dataset realizations. Multiple imputation in a geostatistical context is often a constrained problem where the imputed values must (1) reproduce underlying multivariate relationships and (2) reproduce the spatial variability of observed values (Barnett & Deutsch, 2015). Collocated variables characterize the multivariate relationships, and the covariance structure of the observed variables characterizes the spatial variability. The conditional distribution from which the imputed values are drawn is informed by these components (Hadavand & Deutsch, 2023).

A *latent variable* is a variable that is not directly observed but is assumed related to, and can be inferred from, measured or observed variables (Everitt & Skrondal, 2010). Imputation of latent factors is a challenging imputation problem where all variables are missing or unsampled (Little & Rubin, 2019). The latent factors are not directly observed; they are a synthetic feature of the inferred mathematical model. An example of a latent model is the LMR, which is composed of multiple latent independent random factors operating at different scales (Goovaerts, 1992). The latent factors are never measured or directly observed but characterize the regionalized random variable $Z(\mathbf{u})$. Simulating geologic latent variables subject to other observations or constraints is a key component of truncated Gaussian categorical simulation techniques. The techniques in this chapter do not directly consider categorical values as constraints, though they enforce the reproduction of the true continuous values. Latent variable imputation is most commonly performed with a Gibbs Sampler (Arroyo & Emery, 2020; Emery, Arroyo, et al., 2014; Madani & Bazarbekov, 2021; Silva & Deutsch, 2017) or alternatively the sequential spectral turning bands method (Lauzon & Marcotte, 2020a, 2020b, 2023). The Gibbs sampler is a Markov chain Monte Carlo method used to sample a multivariate distribution where direct sampling is complex but sampling marginal distributions is simple. The Gibbs sampler is practical for indirectly sampling high-dimensional distributions using univariate conditional distributions, though convergence of the algorithm is a known issue with correlated variables (Silva, 2018). Many data require a restricted kriging search, which may cause the simulated Gaussian vector to deviate from the desired covariance matrix (Emery, Arroyo, et al., 2014).

5.2 Gibbs Sampler

The Gibbs sampler (Geman & Geman, 1984) is an iterative simulation algorithm designed to sample an M -dimensional multivariate distribution $f(y_1, \dots, y_M)$ of M random variables $\{Y_1, \dots, Y_M\}$ (Little & Rubin, 2019). It is particularly useful where sampling the joint distribution is difficult, but sampling the marginal conditional distributions $f(y_m | y_1, \dots, y_{m-1}, y_{m+1}, \dots, y_M)$, $m = 1, \dots, M$ is possible. The following general steps summarize the Gibbs sampler:

1. Initialize counter $t = 0$
2. Initialize a valid arbitrary vector $y^{(0)}$
3. For each dimension $m = 1, \dots, M$:
 - a) $t = t + 1$
 - b) set $y_j^{(t)} = y_j^{(t-1)} \quad \forall j \neq m$
 - c) draw $y_m^{(t)}$ from $f(y_m | y_1, \dots, y_{m-1}, y_{m+1}, \dots, y_M)$
 - d) finish if $t = t_{MAX}$
4. Return $y^{(t)}$

Over a sufficient number of iterations, the sampled vector converges on the joint distribution. In a geostatistical context, the Gibbs sampler is based on the fact that the distribution of a Gaussian vector \mathbf{Y} conditioned on other values is Gaussian; the mean and variance of this distribution are calculated by simple kriging (Emery, Arroyo, et al., 2014). The Gibbs sampler simulates both conditional and unconditional vectors. Conditional Gibbs simulation is commonly used for latent variable assignment in truncated-Gaussian techniques (Armstrong, Galli, et al., 2011; Silva & Deutsch, 2017) where latent values must respect both the mapping between categorical and continuous space and match the categorical observations when truncated. Imputation of latent factors in the context of this work does not require satisfying inequality constraints at the data locations. However, it does require satisfying the mapping condition between observed and latent space.

Silva (2018) documents Gibbs sampler convergence issues related to restricting the search neighbourhood for updating the marginal conditional distributions. Considering all n data requires the inversion of a rank $n - 1$ covariance matrix which becomes unpractical with increasing n . Restricting the search to reduce the size of the covariance results in an approximation of the conditional moments and affects algorithm convergence (Emery, Arroyo, et al., 2014; Lauzon & Marcotte, 2020b) as realization quantiles diverge to extreme highs and lows with increasing iterations. Lantuéjoul and Desassis (2012) proposed the propagative Gibbs sampler to avoid the matrix inversion requirement, though Silva (2018) shows convergence issues are still present with greater than two latent variables and complex truncation rules. These convergence challenges motivate the development of a new simulation algorithm for the imputation of latent geologic variables. The SGRI algorithm also

utilizes a restricted search neighbourhood to calculate conditional moments. However, the relaxed constraints relative to imputation for truncated Gaussian methods allow for stable convergence.

5.3 Sequential Gaussian Rejection Imputation

Generating NMR realizations requires imputation of latent factors at the data locations. These imputed factors become conditioning data for simulation on a regular grid. The problem involves assigning M unknown latent Gaussian values with the correct spatial structure at the data locations. When mapped through the NMR, these imputed values must also reproduce the observed regionalized random variable $Z(\mathbf{u})$. This solution is non-unique, and multiple combinations of latent factors could reproduce the observed values. In order to correctly transfer this latent uncertainty, a multiple imputation (Barnett & Deutsch, 2015) approach is adopted where each realization of $Z(\mathbf{u})$ is generated with a unique imputed realization, \mathbf{y} .

SGRI is an iterative simulation algorithm for imputing continuous latent Gaussian variables subject to a mapping constraint $\mathcal{F}_\theta(\mathbf{y}) = \mathbf{z} \pm \alpha$, where α is a data matching tolerance. SGRI first iteratively samples all univariate conditional distributions of the M -dimensional latent distribution until the mapped value is within a coarse tolerance with the observed value. After the simulation meets the coarse tolerance, a polishing step iteratively perturbs each initial imputed value until the mapped value is within a second, finer tolerance with the observed value. At any point during the perturbation, the sample is rejected if the new value does not decrease the error between observed and imputed. The rejection component of the algorithm is not rejection sampling in the strict statistical sense but rather a constraint to ensure the solution remains within a space of feasible solutions. Initial sampling of the conditional distributions ensures the correct covariance structure for each latent factor, and iterative polishing ensures data reproduction. Though Armstrong, Galli, et al. (2011) suggests rejection sampling is not feasible for latent imputation, the relaxed constraints relative to imputation for truncated-Gaussian techniques allow for stable algorithm convergence.

Latent factors are imputed such that when mapped through the NMR, the observed values are reproduced exactly, and each latent factor reproduces its covariance structure. Figure 5.1 shows this relationship schematically with a sketch of a drillhole with $i = 1, \dots, n$ observed data locations. The latent space consists of $m = 1, \dots, M$ factors to be imputed at each observed location. The observed values are a function of the mapped latent factors. It is straightforward to generate independent Gaussian values that reproduce observed values; however, latent spatial continuity must also be correct. $F_m(y) = 1, \dots, M$ CDFs must be standard normal, and the spatial distribution must reproduce variograms γ_m , $m = 1, \dots, M$.

The three general steps of the SGRI algorithm are as follows:

1. Calculate $m = 1, \dots, M$ conditional means and standard deviations at the imputation location \mathbf{u}_i using SK.

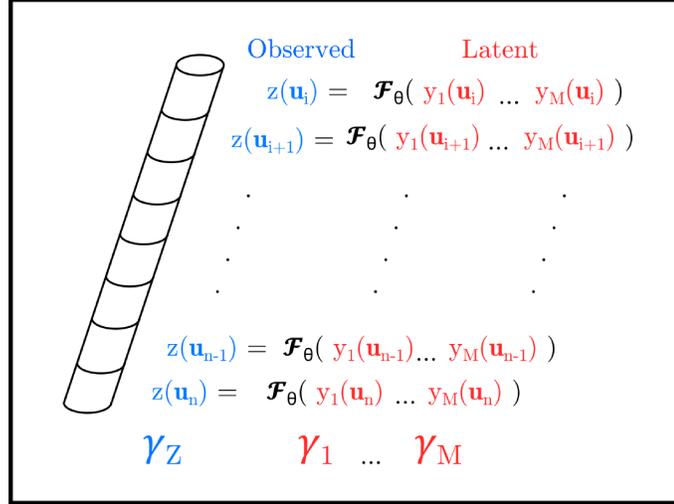


Figure 5.1: A schematic drillhole showing the multivariate relationships between observed and latent spaces. The observed data values are a function $\mathcal{F}_\theta(\mathbf{y})$ of the imputed latent factors at location \mathbf{u}_i . The imputed values must reproduce the true observed values, and each regionalized latent factor must reproduce its covariance structure γ_m .

2. Repeatedly sample the Gaussian vector $\tilde{\mathbf{y}}(\mathbf{u}_i) = \{\tilde{y}_m(\mathbf{u}_i), m = 1, \dots, M\}$ from the conditional distributions until $|\mathcal{F}_\theta(\tilde{\mathbf{y}}(\mathbf{u}_i)) - z(\mathbf{u}_i)|$ is within a specified first tolerance.
3. Iteratively refine the solution from (2) until $|\mathcal{F}_\theta(\tilde{\mathbf{y}}(\mathbf{u}_i)) - z(\mathbf{u}_i)|$ is within a specified second tolerance.

The normal score transform within the latent to observed mapping function \mathcal{F}_θ is slightly different from the context of network parameter inference in Chapter 4. The SGRI algorithm constructs a reference distribution for mapping trial latent vectors, $\tilde{\mathbf{y}}(\mathbf{u}_i)$, to a scalar value in observed space, $\tilde{z}(\mathbf{u}_i)$. The reference distribution facilitates the normal score transform of a single NMR output value. The table is constructed by mapping a $10^6 \times M$ dimensional table of standard normal independent factors through \mathcal{F}_θ , resulting in an exhaustive table of outputs of unitless, raw activation values. This output table is normal score transformed, and $\tilde{z}(\mathbf{u}_i)$ values are calculated by linearly interpolating the normal score array. The notation $\mathcal{F}_\theta(\tilde{\mathbf{y}}(\mathbf{u}_i))$ implies a scalar Gaussian deviate. In the following sections, the subscript i replaces the location vector \mathbf{u}_i to simplify notation.

5.3.1 Precedence and Constraints

Application of precedence is an optional component of the SGRI algorithm. Precedence is achieved by employing the same sigmoid weighting function optionally applied during network parameter inference in Section 4.3.3. Application of constraints is an optional component of the SGRI algorithm that may be used in conjunction with factor precedence. The use of constraints enforces the values of a certain factor to be either above or below a chosen threshold. The goal of constraint application is to ensure that some high or low values of a chosen factor correspond with the high or low values of the observed data. When enforcing factor precedence, constraints are necessary as the mapping

function \mathcal{F}_θ is inferred using unconditional realizations. There is no conditioning mechanism to ensure the low- or high-grade features of \mathbf{y}_p align with low- or high-grade features in the data. \mathcal{F}_θ is non-unique and multiple latent vectors can reproduce z_i . Therefore, it is possible, for example, to have a negative value of \mathbf{y}_p where z_i is a large positive value. This scenario is not ideal if we explicitly want the spatial features of \mathbf{y}_p where z_i is high and warrants the use of constraints on \mathbf{y}_p during imputation.

Constraints can be thought of as “seeding” some number of high or low values of \mathbf{y}_p prior to imputation by using a semi-random path. The constraint threshold and an exclusion radius parameter control the number of seeded values. A threshold is specified, and all values in the observed data, either above or below, are flagged. These data locations are the initial seed locations. The initial seed locations are sorted in descending (high values) or ascending (low values) order based on the observed grade value. The first seed location is visited, and the algorithm removes all other initial seed locations within the exclusion radius from the available possible seed locations. The next valid seed location is visited; again, the algorithm removes all other locations within the radius from the possible locations. This process continues until all initial seed locations are included or excluded. The included seed locations become the first n locations in the imputation path. Imputed values of \mathbf{y}_p at these locations are constrained to be above or below the threshold through rejection. The initial excluded seed locations and all other data locations become the remaining $n_{\text{data}} - n$ locations in the imputation path. Imputation of the remaining $M - 1$ factors uses a strictly random path through the locations.

The exclusion radius is an important parameter as the mean and variance of the imputed \mathbf{y}_p distribution are sensitive to the number of constrained locations. For example, if the 0.9 quantile is used as a threshold with no exclusion radius, approximately 10% of the imputation locations will be seeded. This strong degree of conditioning can inflate the mean and variance of the imputed values and cause a departure from the standard normal distribution. By enforcing an exclusion radius, only a subset of the locations are seeded, and the local effect of the conditioning is less pronounced. Seeding many locations may also negatively affect other factors or lead to non-convergence. Too strong of a constraint on \mathbf{y}_p may lead to a scenario where the remaining factors lack sufficient flexibility to reproduce the observed value when mapped. Practice shows that an exclusion radius of 5-10 times the composite length is sufficient to impart the desired factors of \mathbf{y}_p above or below the threshold without negatively affecting convergence.

5.3.2 Conditional Moments

The algorithm begins by determining a random (unconstrained) or semi-random (constrained) path through all data locations to be imputed. At each imputation location, the conditional moments

are solved using the normal equations:

$$y_{m,i} \sim \mathcal{N}(\mu_{m,i}, \sigma_{m,i}) \quad (5.1)$$

$$\mu_{m,i} = \sum_{j=1}^{N_{m,i}} \lambda_{j,i} \times y_{m,j} \quad (5.2)$$

$$\sigma_{m,i}^2 = 1 - \sum_{j=1}^{N_i} \lambda_{m,j,i} \times C_{m,j,i} \quad (5.3)$$

$$\sum_{j=1}^{N_{m,i}} \lambda_{m,j,i} \times C_{m,j,k} = C_{m,k,i} \quad \forall k \in N_{m,i} \quad (5.4)$$

where $N_{m,i}$ is the neighbourhood about location i using search anisotropy derived from Y_m ; this could either be all neighbours or a restricted search around location i . $C_{m,j,k}$ is the covariance between spatial locations j and k for latent variable m . The system of equations is solved M times under the assumption of zero mean and unit variance, resulting in a vector of conditional means and standard deviations. Retaining all conditional moments allows fast, repeated simulation of spatially correlated latent values in the coarse search phase.

5.3.3 Coarse Search

The coarse search begins after calculating the M conditional moments at location i . The goal of the coarse search is to find the trial vector $\tilde{\mathbf{y}}_i$ that closely reproduces z_i and has the correct spatial covariance. Values are drawn from each valid conditional distribution through MCS for a set number of iterations, j :

$$\tilde{y}_{m,i}^{(j)} = r_{\in[0,1]} \times \sigma_{m,i} + \mu_{m,i} \quad \forall M \quad (5.5)$$

$$\tilde{\mathbf{y}}_i^{(j)} = \{y_{1,i}^{(j)}, y_{2,i}^{(j)}, \dots, y_{M,i}^{(j)}\} \quad (5.6)$$

where $r_{\in[0,1]}$ is a uniform random number between 0 and 1. Simulated $\tilde{y}_{m,i}^{(j)}$ values are practically constrained $\in [-5, 5]$. The trial observed value is calculated by mapping $\mathcal{F}_\theta(\tilde{\mathbf{y}}_i)$. Coarse imputation error, e_1 , is the absolute difference between the trial value, $\tilde{z}_i^{(j)}$, and the true observed data value:

$$\tilde{z}_i^{(j)} = \mathcal{F}_\theta(\tilde{\mathbf{y}}_i^{(j)}) \quad (5.7)$$

$$e_1 = |\tilde{z}_i^{(j)} - z_i| \quad (5.8)$$

$$\tilde{\mathbf{y}}_i = \begin{cases} \tilde{\mathbf{y}}_i^{(j)}, & \text{if } e_1 < \alpha_1 \\ \tilde{\mathbf{y}}_i^{(j-1)}, & \text{otherwise} \end{cases} \quad (5.9)$$

If the error, e_1 , is less than the first rejection tolerance, α_1 , the trial vector $\tilde{\mathbf{y}}_i^{(j)}$ is retained as the initial latent vector at the i^{th} location, otherwise it is rejected. If the coarse search fails to converge after the specified number of iterations, the sample $\tilde{y}_{m,i}^{(j)}$ is flagged for resimulation. The values of both $\tilde{y}_{m,i}^{(j)}$ and a small neighbourhood of the nearest samples are reset. Resimulating the nearest

neighbours prevents situations where local conditioning leads to non-convergence. This scenario is possible with adjacent, opposite, extreme values in a drillhole, such as at a vein's or high-grade structure's boundary. Resimulation uses a default of 10 nearest neighbours. Practice shows that ≈ 0.1 is a reasonable rejection tolerance for α_1 . Exact reproduction with the observed value is not expected during this step as the multivariate relationship between M collocated factors has not been addressed. Any exact matching now would be disturbed in subsequent polishing steps.

5.3.4 Solution Polishing

Once an initial coarse solution is accepted, it can be refined to meet the collocated multivariate requirements. The imputed vector generated in the coarse search ensures Y_m reproduces $\gamma_m(\mathbf{h})$ however $\mathcal{F}_\theta(\tilde{\mathbf{y}}_i)$ must match $z_i \pm \alpha_2$, where α_2 is the polishing tolerance. Solution polishing at the i^{th} location begins by assessing the sensitivity of each component of $\tilde{\mathbf{y}}_i^{(j)}$ on the target observed value z_i . Due to the nature of the learned mapping function, it is not immediately clear how each latent factor influences the observed value. $\tilde{y}_{m,i}^{(j)}$ can either be positively or negatively correlated with \tilde{z}_i and the magnitude of the sensitivity depends on \mathcal{F}_θ . To assess the sensitivity of each factor, a step size of $\Delta_y = G^{-1}(0.525) - G^{-1}(0.500)$, or 2.5% in Gaussian probability space, is chosen. Perturbing a component, m , of $\tilde{\mathbf{y}}_i^{(j)}$ by $\pm\Delta_y$ results in two trial vectors $\tilde{\mathbf{y}}_{m+,i}^{(j)}$, and $\tilde{\mathbf{y}}_{m-,i}^{(j)}$. The trial vectors permit calculation of sensitivity as $\Delta\tilde{z}_{m+,i} = \tilde{z}_i - \mathcal{F}_\theta(\tilde{\mathbf{y}}_{m+,i}^{(j)})$ and $\Delta\tilde{z}_{m-,i} = \tilde{z}_i - \mathcal{F}_\theta(\tilde{\mathbf{y}}_{m-,i}^{(j)})$. Iterating over $m = 1, \dots, M$ components establishes the sensitivity and direction of change of \tilde{z}_i with respect to $\tilde{y}_{m,i}^{(j)}$. Figure 5.2 shows this sensitivity graphically as a “tornado” chart.

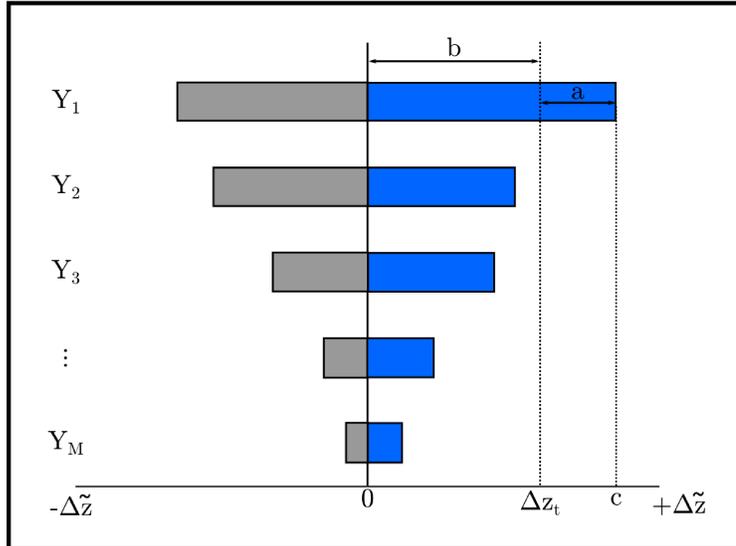


Figure 5.2: Schematic representation of latent factor sensitivity as a tornado chart. The x-axis shows $\Delta\tilde{z}$, or how much \tilde{z} changes for a given perturbation of Y_m . In this scenario, the target delta is a positive value. The y-axis shows $m = 1, \dots, M$ latent factors sorted by descending sensitivity. $a = \Delta z_t - c$; $b = \Delta z_t$; $c =$ the largest possible change to \tilde{z} .

Factors are sorted from most to least sensitive. Δz_t is the target delta, or the difference between

z_i and \tilde{z}_i . Δz_t tells us which direction to move, and the tornado bars tell us which factor(s) are sensitive enough to achieve the target delta. a represents the delta between the most sensitive factor and the target delta, b is equal to the target delta, and c is equal to the largest possible change in \tilde{z}_i given the current vector $\tilde{\mathbf{y}}_i^{(j)}$. Suppose the target delta lies within one or more of the sensitivity bars, as shown in Figure 5.2. In that case, the algorithm adjusts the least sensitive factor to solve $\mathcal{F}_\theta(\tilde{\mathbf{y}}_i^{(j)}) = z_i \pm \alpha_2$ by performing a binary search (Nowak, 2008) on the interval $\tilde{y}_{m,i}^{(j)} \pm \Delta y$. The least sensitive factor is selected to have the smallest possible impact on the initial vector from the coarse search. Suppose the target delta lies outside the maximum sensitivity of any factor. In that case, the most sensitive factor is perturbed to the maximum amount, and the algorithm reassesses the sensitivity of all latent factors. This perturbing and sensitivity assessment loop repeats until a binary search can solve the problem. The benefit of this approach is the latent vector $\tilde{\mathbf{y}}_i^{(j)}$ is always being perturbed in the correct direction, and Δz_t is always approaching zero. Given a sufficient number of iterations, solution polishing must converge. That being said, there may be situations where the required amount of polishing negatively influences the covariance structure imposed in the coarse search. If at any point $|\tilde{z}_i^{(j)} - z_i| < \alpha_2$ the loop breaks as $\mathcal{F}_\theta(\tilde{\mathbf{y}}_i^{(j)}) \approx z_i$. Practice shows that ≈ 0.01 is a reasonable tolerance for t_2 .

Assessing latent sensitivity permits the calculation of a measure of “difficulty” in the polishing phase. Consider the equality:

$$b = \frac{\Delta z_t - c}{\Delta z_t} \quad (5.10)$$

If the target delta falls within the sensitivity of one or more factors, $b < 0$, and the correct vector can be solved with binary search. In this scenario, solution polishing converges quickly. If the target delta falls outside the sensitivity of all factors, $b > 0$, factor sensitivity must be repeatedly assessed, and convergence takes additional iterations. This ratio provides insight into locations or observed samples that are more difficult to impute. However, it is possible there are locations that are difficult to impute but require little polishing. Consider a location where resimulation is necessary for the coarse search, but the algorithm eventually generates a latent vector where $e_1 \approx \alpha_2$. This scenario is still difficult to impute, though b does not reflect this. When evaluating problematic locations, the number of resimulations at each data location should also be considered. Data locations persistently challenging to impute across realization are subject to further investigation. Algorithm 2 summarizes the complete pseudocode for the SGRI algorithm.

5.3.5 Imputation Checks

There are any number of latent vectors $\tilde{\mathbf{y}}_i$ that can reproduce the observed value z_i . As the solution is highly non-unique, multiple imputed realizations of the latent factors should be considered. Considering a multiple imputation framework transfers the uncertainty in the latent variables to subsequent model realizations (Silva & Deutsch, 2017). SGRI imputes realizations of latent vari-

Algorithm 2 SGRI pseudo code.

```

1: Initialize search parameters based on covariance
2: for  $\ell = 1, \dots, L$  do                                     ▷ Main loop over realizations
3:   establish random or semi-random path through data
4:    $nresim = 0$ 
5:   for  $i = 1, \dots, ndata$  do                               ▷ Loop over data locations
6:     for  $m = 1, \dots, M$  do                                 ▷ Loop over factors at  $i^{th}$  location
7:       establish search neighbourhood  $N_{m,i}$ 
8:       calculate conditional moments  $\mu_{m,i}, \sigma_{m,i}^2$ 
9:     end for                                               ▷ End loop over factors
10:     $j = 0$ 
11:    while  $e_1 < \alpha_1$  do                                 ▷ Start coarse search
12:       $j = j + 1$ 
13:      for  $m = 1, \dots, M$  do
14:         $y_{m,i} = r_{\in[0,1]} \times \sigma_{m,i} + \mu_{m,i}$        ▷ Monte Carlo simulation
15:      end for
16:       $\tilde{z}_i = \mathcal{F}_\theta(\tilde{\mathbf{y}}_i^{(j)})$ 
17:       $e_1 = |\tilde{z}_i^{(j)} - z_i|$                                ▷ Compare to observed value
18:      if  $j > iter1$  then
19:        break
20:      end if
21:    end while                                             ▷ End coarse search
22:    if  $e_1 > \alpha_1$  then
23:       $nresim = nresim + 1$ 
24:      cycle data loop
25:    end if
26:     $j = 0$ 
27:     $e_2 = e_1$ 
28:    while  $e_2 < \alpha_2$  do                               ▷ Start solution polishing
29:       $j = j + 1$ 
30:      assess latent sensitivity
31:      if  $\frac{a}{b} < 0$  then
32:        solve  $\tilde{\mathbf{y}}_i^{(j)}$  with binary search
33:      else
34:        set most sensitive factor to its bound
35:        reassess latent sensitivity
36:      end if
37:       $\tilde{z}_i = \mathcal{F}_\theta(\tilde{\mathbf{y}}_i^{(j)})$ 
38:       $e_2 = |\tilde{z}_i^{(j)} - z_i|$                                ▷ Retain new absolute difference
39:      if  $j > iter2$  then
40:        break
41:      end if
42:    end while                                             ▷ End solution polishing
43:  end for                                                 ▷ End loop over data locations
44:  if  $nresim > 0$  then
45:    resimulate at  $nresim$  locations
46:  end if
47: end for                                                 ▷ End loop over realizations

```

ables with the correct spatial structure, are standard normal, and are independent. The minimum acceptance criteria of the imputed values in practice are:

1. Mapping the latent vector $\tilde{\mathbf{y}}_i$ (red in Figure 5.1) to observed space via \mathcal{F}_θ reproduces the observed value $z_i \pm \alpha_2$ (blue in Figure 5.1).
2. Latent factors are standard normal: $E\{Y_m\} = 0$, $E\{Y_m^2\} = 1$
3. Latent factors are independent: $E\{Y_{m,i}Y_{n,j}\} = 0$, $\forall m \neq n, \forall i \neq j$
4. Latent factors reproduce their respective variogram model $\gamma_m(\mathbf{h})$, $\forall \mathbf{h}$

Consider the same small synthetic example from Chapter 4; the problem involves imputing three latent factors at 746 data locations. Factor three is the nugget effect. The SGRI algorithm is run using a search neighbourhood of 40 nodes, a maximum of 20,000 and 10,000 iterations for the coarse search and polishing steps, respectively, and rejection tolerances of 0.1 and 0.01 for the coarse search and polishing steps, respectively. Figure 5.3 shows a scatter plot matrix between all imputed factors, the mapped values in observed space, \tilde{z} , and the true data values, z for a single realization. As expected, the correlation between z and \tilde{z} is 1.0. The top row of the matrix in Figure 5.3 is somewhat redundant due to this perfect correlation; however, it highlights the exact data matching and validates item (1). The marginal distributions (histograms in Figure 5.3) are all standard normal validating item (2). Deviation from the standard normal distribution is possible, and one must consider the variogram range relative to the domain size. Long-range variogram structures may generate low-variance imputed distributions. Imputed latent factors are uncorrelated with roughly concentric density contours and $|\rho| < 0.10$ validating item (3). The relationships between factors 1 and 2 with the mapped values highlight the influence of the ω parameter. Factor 1 is correlated with the mapped value when it is low, and factor 2 is correlated when it is high. Figure 5.4 shows variogram reproduction for each imputed factor. There is no nugget effect variogram model, so the reproduction is not shown. The red-shaded area highlights uncertainty in the variogram across all imputed realizations. The expected imputed variograms reproduce the single structure variogram models reasonably well for 746 data.

Beyond the minimum acceptance criteria, one should investigate data locations requiring multiple resimulations. Samples at transitions between extreme grade ranges or local neighbourhood outliers may require multiple simulation passes for $\tilde{\mathbf{y}}_i$ to converge. Table 5.1 shows six adjacent samples from a single drillhole, where z is the observed value, \tilde{z} is the mapped standard normal value, Factors 1-3 are the imputed latent Gaussian values, b is the ratio in Equation 5.10, and nresim is the number of resimulations at the given data location. The third sample represents a transition from 0.563 to -1.306 or approximately the 71st to the 9th quantile of the Gaussian distribution. This short-scale, extreme change is a challenge for the SGRI algorithm, though it can converge with a sufficient number of rejection iterations. It is noteworthy that $b < 0$, or the coarse search was able to produce a latent vector $\tilde{\mathbf{y}}_i^{(j)}$ where minimal solution polishing is required.

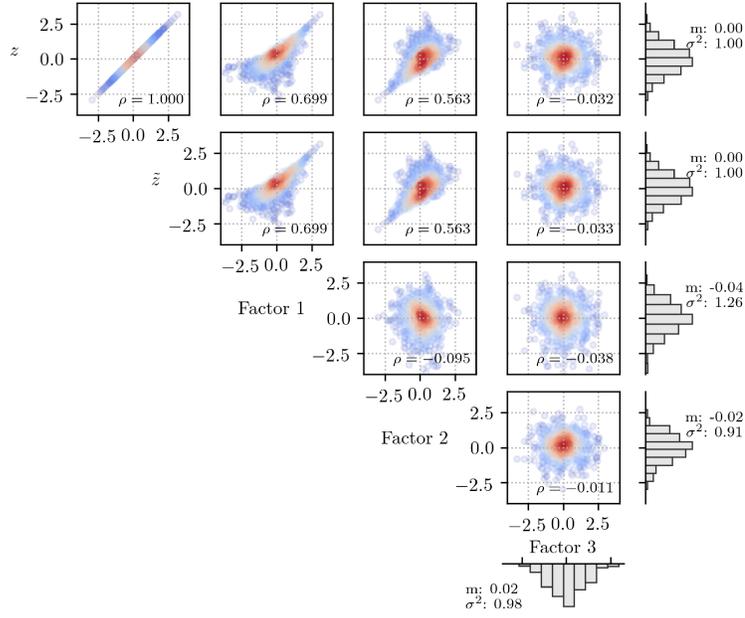


Figure 5.3: Scatter plot matrix for a single imputed realization highlighting the perfect correlation between z and \tilde{z} , the uncorrelated nature of all latent factors, and the standard normal nature of all marginal distributions. These features are the first three minimum acceptance criteria for SGRI imputed realizations.

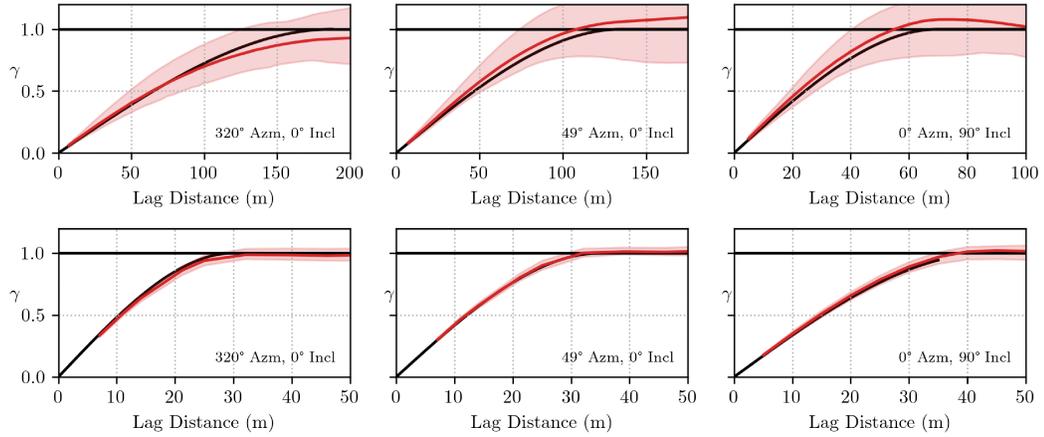


Figure 5.4: Latent factor variogram reproduction for the first two factors in the major (left), minor (center) and vertical (right) directions, respectively. The expected variograms of the imputed realizations reproduce the models well, given the data density.

Table 5.1: Six adjacent samples from a single drillhole, where z is the observed value, \tilde{z} is the mapped standard normal value, Factors 1-3 are the imputed latent Gaussian values, b is the ratio in Equation 5.10 and nresim is the number of resimulations at the given data location. The polishing tolerance is 0.01.

z	\tilde{z}	Factor 1	Factor 2	Factor 3	$\frac{a}{b}$	nresim
0.784	0.791	1.223	0.501	-1.616	-0.411	0
0.563	0.564	1.134	0.206	-2.068	-0.834	0
-1.306	-1.310	-0.820	-1.937	-1.185	-1.638	2
-0.307	-0.305	0.165	-1.153	0.739	-0.020	0
0.988	0.982	0.255	0.974	0.202	-1.481	0

If solution polishing does not converge after a reasonable number of iterations, the algorithm can draw a valid Gaussian from the internal SGRI reference distribution. $\tilde{\mathbf{y}}_i^{(jmax)}$ is selected from the “lookup table” such that $|\tilde{z}_i^{(jmax)} - z_i|$ is minimized. A reference distribution with 10^6 entries is sufficiently large to contain a latent vector that satisfies the polishing tolerance when mapped to observed space. Frequent use of a lookup table may affect multivariate and spatial properties of the latent factors, so SGRI issues a warning if more than 1% of the imputed samples are drawn from the reference distribution.

5.4 Latent Factor Simulation

NMR realizations are generated by conditionally simulating latent factors on a grid and then mapping the gridded factors to observed space. Latent factors imputed at the data locations become the conditioning data for any conditional simulation algorithm. Standard conditional simulation algorithms include SGS (Gómez-Hernández & Journel, 1993), turning-bands (Journel, 1974) and LU simulation (Davis, 1987). SGS is likely the most commonly implemented algorithm due to its simplicity and availability in commercial software (Rossi & Deutsch, 2013), though any conditional algorithm is valid. Figure 5.5 shows plan view sections through the first conditional realization of the three gridded factors.

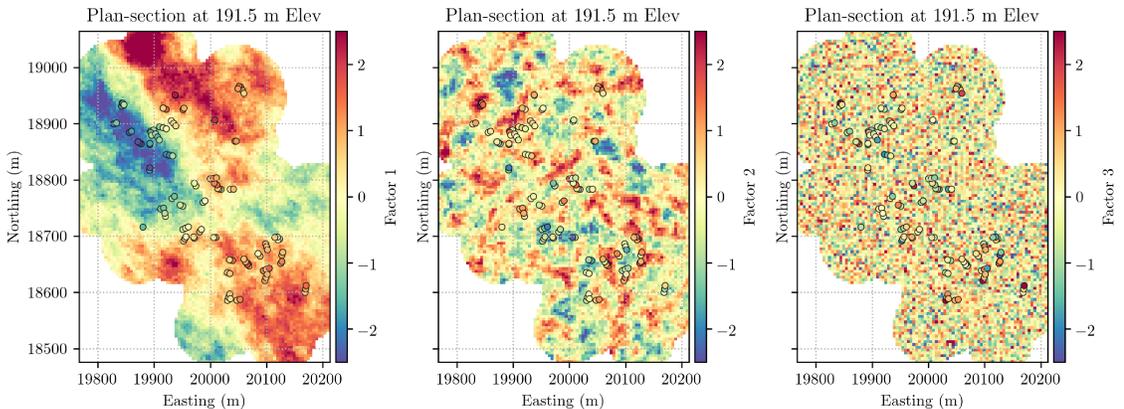


Figure 5.5: Plan view sections through the first conditional realization of the three gridded factors with imputed conditioning data.

Once the latent factors are defined at every grid node, the realizations are mapped to the observed space through \mathcal{F}_θ . Figure 5.6 shows sections through the first NMR realization mapped to Gaussian units. The non-Gaussian characteristics of the realizations are evident in these sections. The localized high-grade values clearly overprint the low-grade, longer-range background values defined by the Gaussian pool and ω constraints. These features are also evident in the back-transformed realizations in Figure 5.7. The realizations show localized but internally connected high-grade regions that correspond to the covariance structure of the 0.9 quantile indicator variogram.

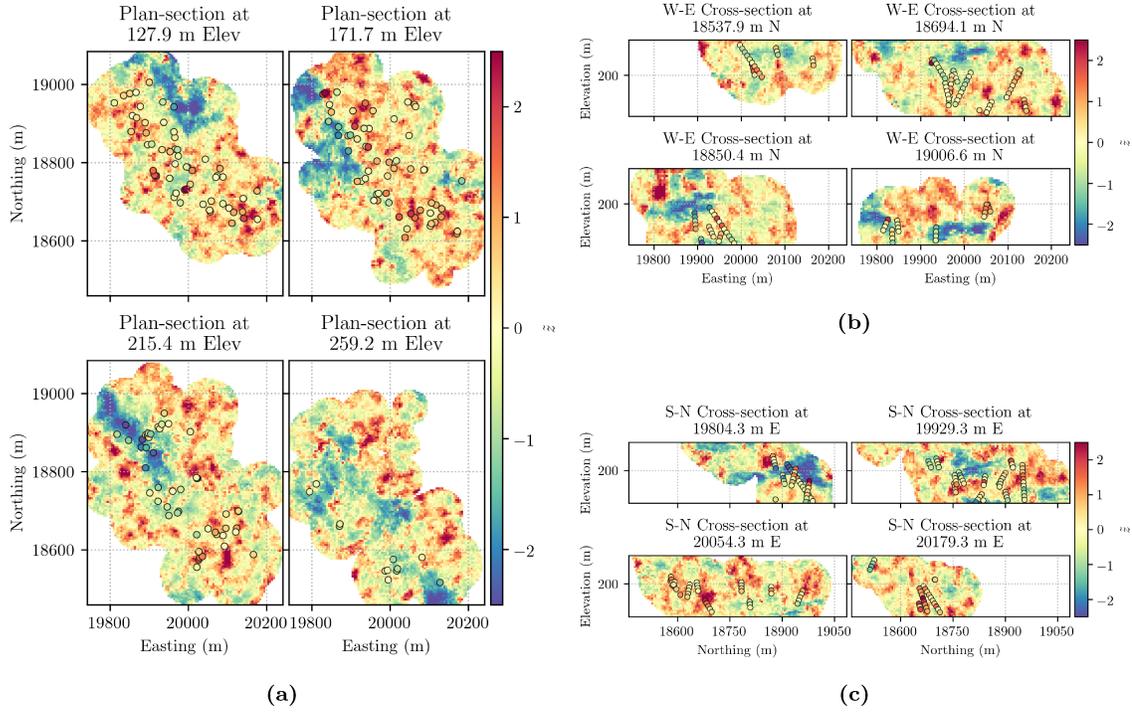


Figure 5.6: Plan (a), east-west (b) and north-south (c) sections through the first NMR realization mapped to Gaussian space with observed data values in Gaussian units.

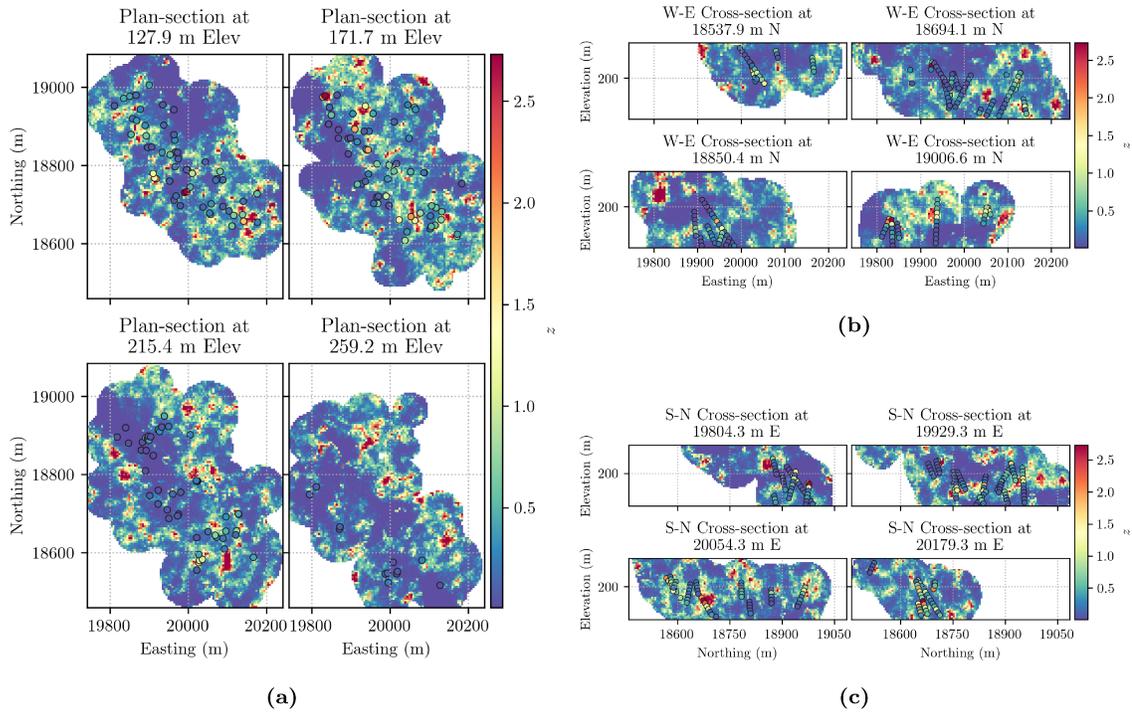


Figure 5.7: Plan (a), east-west (b) and north-south (c) sections through the first NMR realization back-transformed to original units.

At this point, the gridded NMR realizations should reproduce the continuous variogram model, specified indicator variogram models, cumulative run frequencies and n -point connectivity functions.

As the gridded models reproduce the data at the data locations (particularly implementations of SGS that assign data to grid nodes), an additional step is required to check downhole statistics. For checking runs and n -point connectivity reproduction, the entire drillhole configuration is translated within the domain. This translation retains the drillhole configuration but allows sampling simulated values from the grid. The resampled values are then used to check multi-point statistical reproduction.

5.4.1 Simulation Checks

The minimum acceptance criteria for checking continuous realizations (Leuangthong, McLennan, et al., 2004) must be applied, similar to any other geostatistical model. The gridded realizations must reproduce (1) the data values at data locations, (2) the declustered input CDF and summary statistics for the latent Gaussian pool and the original variable, and (3) the input covariance model, both for the latent Gaussian pool and the original variable.

Figure 5.8 shows CDF reproduction (left) and data reproduction (center) for the gridded NMR realizations in Gaussian units. The Gaussian realizations are, on average, standard normal, and the data values are reproduced exactly at collocated grid nodes. In some locations, more than one data value occupies a single grid node, and there is a deviation from the 1:1 line in the scatter plot. In this scenario, the data value closest to the grid node centroid is retained for comparison. Figure 5.8 (right) shows CDF reproduction of the gridded realizations back-transformed to original units. The realizations, on average, reproduce the declustered CDF. The final component of the traditional model checks is checking continuous variogram reproduction. Figure 5.9 shows gridded continuous variogram reproduction for the NMR realizations in Gaussian units. As the realizations are generated with a mapping function that considers multiple variogram components, there are some deviations from the input continuous model. In this example, the 0.1 and 0.9 quantile indicator variograms influence the continuous variogram reproduction in the major and minor directions. The 0.9 indicator variogram introduces a slight increase in variance in the short-range lags and a slight decrease in the longer-range lags from the 0.1 indicator variogram. Continuous variogram reproduction remains reasonable, though the influence of other objective components is evident in the covariance structure of the final gridded models.

Beyond the traditional simulation model checks, the NMR realizations must be checked to reproduce all the objective function components. These checks include indicator variograms and sequences. Figure 5.10 shows gridded reproduction of the strongly asymmetric 0.1 and 0.9 quantile indicator variograms. The gridded models closely reproduce the non-Gaussian indicator structure that is apparent in Figures 5.6 and 5.7, though there is some deviation at shorter lags. This deviation is attributed to shorter range lags having few pairs given the data configuration. In a mining context, one is predominantly concerned with high values, and in this example, cumulative run frequencies

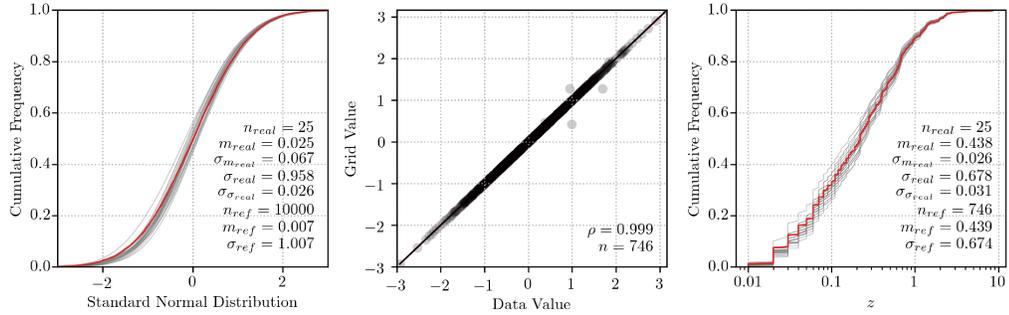


Figure 5.8: CDF and data reproduction for the NMR realizations in Gaussian units (left and center) and CDF reproduction in original units (right). The original units reference distribution shown in red considers declustering weights.

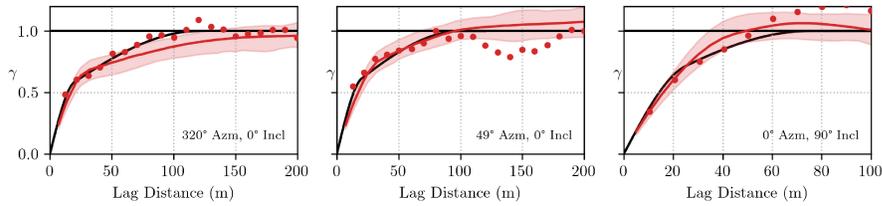


Figure 5.9: NMR continuous variogram reproduction. The black line is the variogram model; the red dots are the experimental variogram of the imputed data; the red line is the average variogram of the gridded NMR realizations, the shaded red area encloses the minimum and maximum gridded NMR variogram values. Left to right are the major, minor and vertical directions, respectively.

are considered above the 0.9 quantile threshold. When considering sequences above a threshold, the n -point connectivity function and cumulative run-length frequencies are analogous; one can be calculated from the other. Only cumulative run-length frequencies are considered for this reason. The chosen simulation algorithm may assign the data values to the grid, resulting in a distribution of runs with zero uncertainty. To overcome this, the entire drillhole configuration is translated within the domain and values are sampled from the grid at the “new” data locations. Four translations are performed (25 meters to the NE, SE, SW and NW), and the final result is the expected value across translations. Figure 5.11 shows gridded cumulative run-length frequencies for the 0.9 quantile indicator. The solid red line is the expected value across translations. The black line is the target value calculated from the drillholes. The expected value deviates slightly from the target, though the uncertainty bandwidth captures the target.

Suppose the NMR realizations meet the minimum acceptance criteria. In that case, the realizations are (1) univariate Gaussian, (2) reproduce the observed data at the data locations, (3) reproduce the continuous variogram model, (4) reproduce the indicator variogram models for chosen thresholds and (5) reproduce chosen multi-point measures of connectivity. Points (4) and (5) ensure the realizations are not multivariate Gaussian.

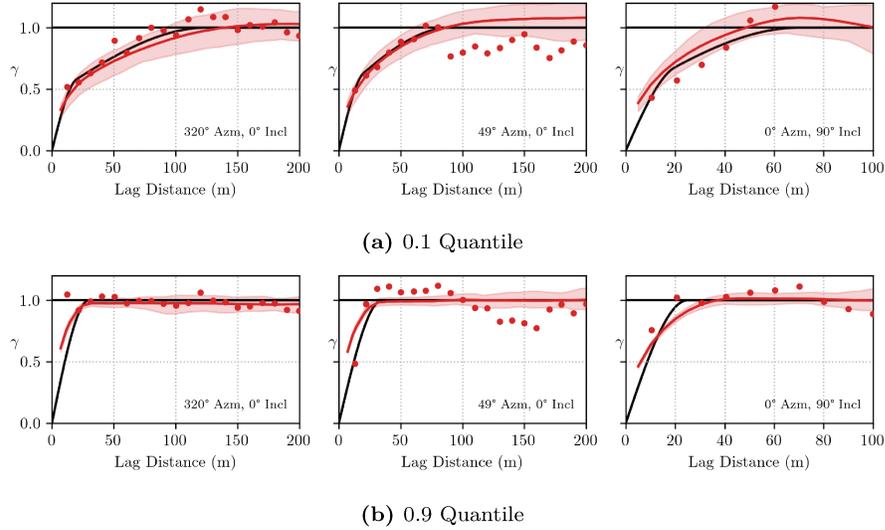


Figure 5.10: NMR indicator variogram reproduction for the 0.1 (a), 0.5 (b) and 0.9 (c) quantiles. The black line is the variogram model; the red dots are the experimental variogram of the imputed data; the red line is the average variogram of the gridded NMR realizations, the shaded red area encloses the minimum and maximum gridded NMR variogram values. Left to right are the major, minor and vertical directions, respectively.

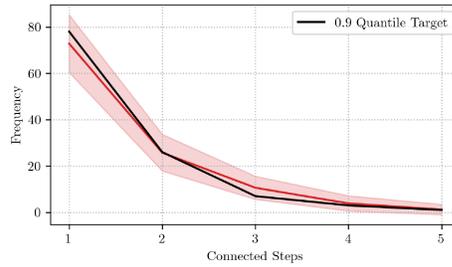


Figure 5.11: NMR 0.9 quantile cumulative run-length frequency reproduction. The black line is the target calculated from the drillholes, the red line is the average gridded value, and the shaded red area encloses the minimum and maximum gridded values.

5.5 Discussion

SGRI is a novel algorithm for imputing latent Gaussian factors for use in the NMR simulation framework. SGRI is an alternative to the Gibbs sampler approach and permits the imputation of any number of independent, standard normal Gaussian vectors with the correct spatial structure. When mapped from latent to observed space, the imputed factors reproduce the observed data value. The mapping function \mathcal{F}_θ is highly flexible and incorporates components of the conceptual geological model and features embedded in the design of the Gaussian pool. The algorithm is straightforward and incorporates elements of SGS and rejection sampling. Simple kriging calculates the moments of local conditional CDFs, and values are drawn with MCS subject to rejection. The rejection step ensures that the imputed vector remains within a feasible solution space. The algorithm shows stable convergence in both constrained and unconstrained forms, though heavy constraints may

lead to departure from the standard normal distribution. SGRI implements optional constraints by seeding some number of imputation locations with values above or below a threshold, ensuring the imputed values are extreme where observed values are extreme. The constraints are enforced through rejection where y_p values must be above or below the threshold.

SGRI is the second component the NMR framework for simulation of non-Gaussian spatial features. Chapters 4 and 5 present a synthetic example of the NMR workflow highlighting all aspects of latent factor design, parameter inference, factor imputation and simulation, factor mapping and checking NMR realizations. The following chapter takes these components to a real 3-D dataset where the connectivity of extreme values is a defining characteristic of the data.

Chapter 6

Case Study

This chapter presents an application of the complete NMR workflow to a real 3-D dataset. The NMR workflow is designed to capture the spatial features of non-Gaussian distributions in a way that the multi-Gaussian assumption and a LMR cannot. The case study aims to construct an NMR spatial model and compare it to a traditional SGS model. The dataset comes from a mineral deposit where the project operator observes non-Gaussian characteristics within drillholes that are not well reproduced by SGS. Within local stope volumes SGS cannot reproduce the connectivity of high-grades present in drillhole data.

The workflow consists of (1) inferring the parameters, θ , of the latent-observed mapping function, \mathcal{F}_θ , using the methodology from Chapter 4, (2) imputing the latent factors at the data locations using the methodology from Chapter 5, and (3) conditional simulation of the imputed factors on the grid. The learned mapping is non-linear. The idea is that the NMR realizations can capture non-Gaussian spatial features such as asymmetric indicator variograms and connectivity of extreme values. Twenty-five conditional realizations are generated with the NMR, and pertinent model statistics are checked and compared to SGS. Defining characteristics of the deposit and mineralization type are omitted for confidentiality purposes.

6.1 Exploratory Data Analysis

The complete dataset consists of 27,983 values nominally composited to 5 meters. All coordinates are transformed such that the spatial centroid of the data is ($x = y = z = 0$). For cross-validation purposes, approximately 30% of the data is withheld as a “test set”, and the remaining 70 % is the “training set”. Figure 6.2 shows plan-view, east-west and north-south sections through the training dataset coloured by the variable of interest. The test data is not exposed to any subsequent modeling step. Figure 6.1 shows CDFs and basic summary statistics for the training and test datasets. There is a notable imbalance between the training and testing sets, with the latter being 14% higher grade. The testing set is not a random data partition but rather a collection of data drilled after a certain date. This choice is a reasonable approach to partitioning as it ensures complete drillholes in both datasets, preventing optimistic metrics with leave-one-sample-out style cross-validation. The downside of this approach is that the holes drilled after the cutoff are for stope definition purposes, leading to a higher mean grade. This imbalance must be considered when assessing the reproduction of the test set. All modeling considers the training set only, and Section 6.6 considers the test set.

Experimental variograms are first calculated using an omni-directional search that informs the

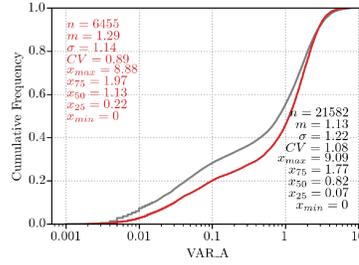


Figure 6.1: CDFs for training (grey) and testing (red) datasets.

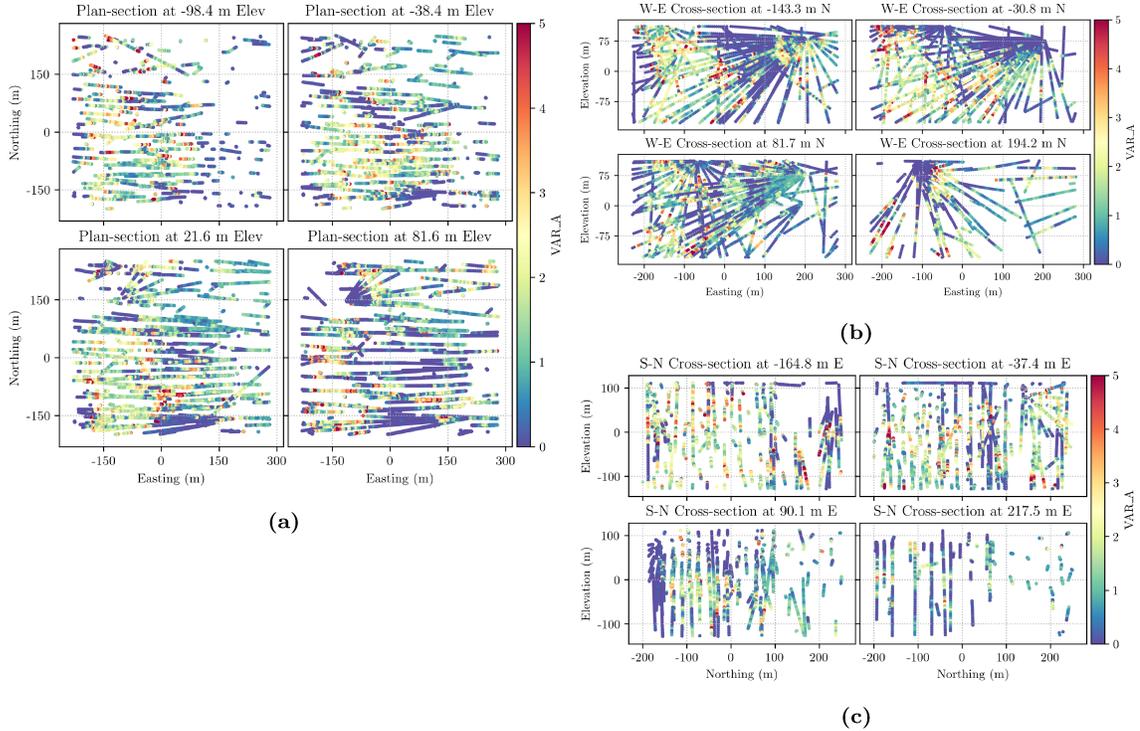


Figure 6.2: Plan (a), east-west (b) and north-south (c) sections through the training dataset.

nugget effect. This search uses short lags to highlight the shape and range of the short-scale continuity. The omni-directional variogram suggests a nugget effect of near zero, and that an exponential variogram shape is appropriate for the modeled variable. Directional experimental variograms are calculated first in original units and fitted with an exponential model. Figure 6.3 shows the fitted experimental points, and Table 6.1 shows the model parameters. The original unit variogram defines the search orientation and anisotropy for estimator-based declustering. Indicator variograms are calculated for the grade distribution's 0.1, 0.5 and 0.9 quantiles. Figures 6.4 show the fitted indicator variogram models, and Table 6.2 summarizes the model parameters. Interestingly, the 0.1 and 0.9 indicator variograms are not highly asymmetric with this dataset.

Transformation to Gaussian space requires a representative CDF. Inverse distance squared (ID2) estimation produces declustering weights. A grid of 5 x 5 x 5 meter blocks is used considering a 25-meter buffer in the x, y, and z directions around all available data (the SMU grid). ID2 estimation

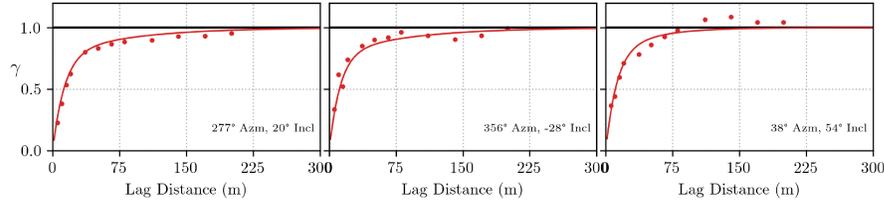


Figure 6.3: Fitted original unit variograms in the major, minor and tertiary directions (left to right).

Table 6.1: Original unit variogram model parameters.

	Nugget	Structure 1	Structure 2
Contribution	0.000	0.655	0.345
Model Shape		exponential	exponential
Angle 1		277.0	277.0
Angle 2		20.0	20.0
Angle 3		30.0	30.0
Range 1		36.3	250.0
Range 2		33.2	125.0
Range 3		33.2	100.0

Table 6.2: Indicator variogram model parameters. All models have zero nugget.

	0.1 Quantile		0.5 Quantile		0.9 Quantile	
	Structure 1	Structure 2	Structure 1	Structure 2	Structure 1	Structure 2
Contribution	0.530	0.470	0.489	0.511	0.359	0.641
Model Shape	exponential	exponential	exponential	exponential	exponential	exponential
Angle 1	277.0	277.0	277.0	277.0	277.0	277.0
Angle 2	20.0	20.0	20.0	20.0	20.0	20.0
Angle 3	30.0	30.0	30.0	30.0	30.0	30.0
Range 1	13.2	200.0	18.0	150.0	8.5	61.8
Range 2	11.0	50.3	17.4	121.0	5.1	56.9
Range 3	19.3	91.3	32.5	84.5	5.1	53.3

weights for each data are accumulated during an estimation run and used as declustering weights. The idea is that data in dense regions receive less total weight as many data are present in the search neighbourhood. In sparse areas, data receive more weight as fewer data are in the search neighbourhood. Estimator declustering is more robust than traditional cell declustering for irregular data configurations, as the search ellipse can capture the principal orientation of continuity. The limits of the estimation grid must be well-defined to prevent unintended edge effects (Wilde, 2007). Figure 6.5 shows the clustered (gray) and declustered CDFs (red).

After the normal score transformation, the experimental variogram of the normal scores is calculated and fitted with a model. Figure 6.6 shows the fitted experimental points, and Table 6.3 shows the model parameters. A normal score variogram permits the calculation of accuracy metrics, particularly the fraction of true values that fall within specified probability intervals. Figure 6.7 shows a leave-one-drillhole out cross-validation and accuracy plot for the fitted normal score variogram model. The scatter plot shows a reasonable slope of regression (SOR) and reproduction of the true

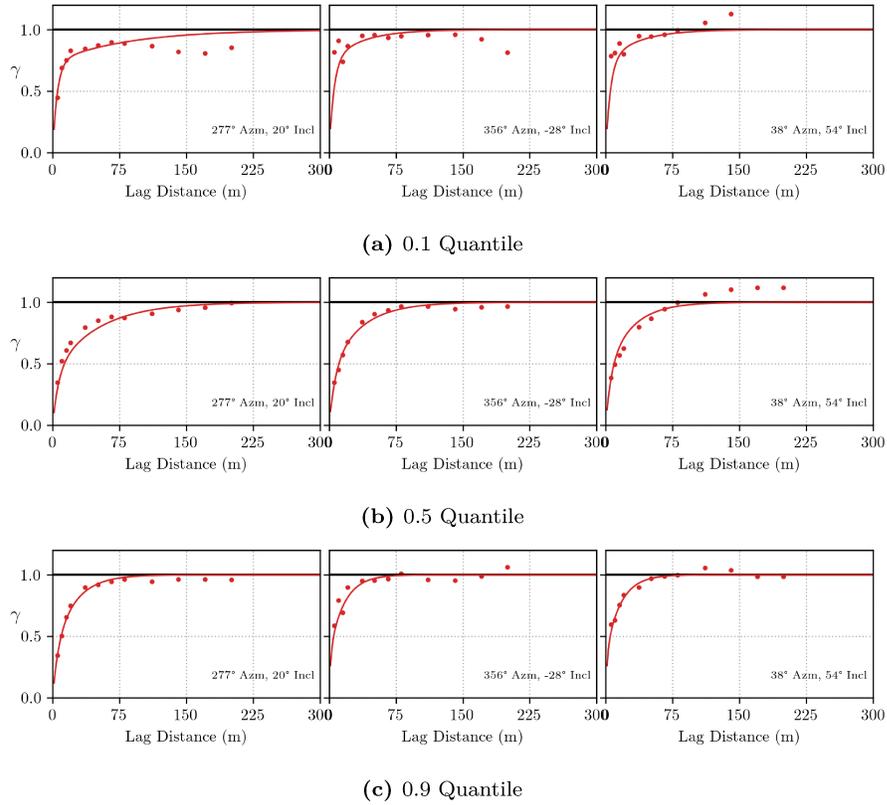


Figure 6.4: Fitted experimental indicator variograms for the 0.1 (a), 0.5 (b) and 0.9 (c) quantiles in the major, minor and tertiary directions (left to right).

mean. The accuracy plot suggests that the variogram model characterizes a reasonable uncertainty distribution.

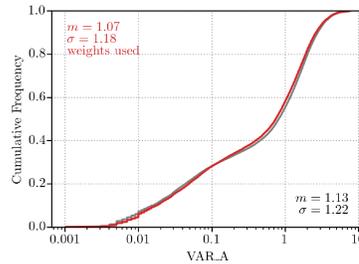


Figure 6.5: Clustered (gray) and declustered CDFs (red) considering ID2 weights.

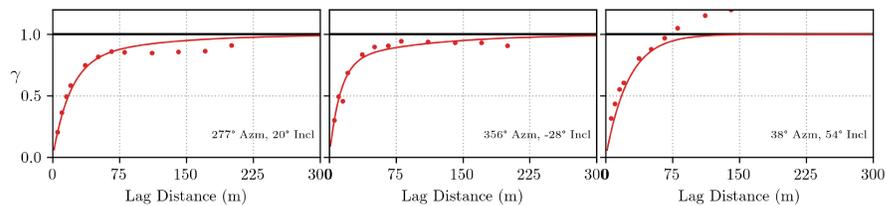
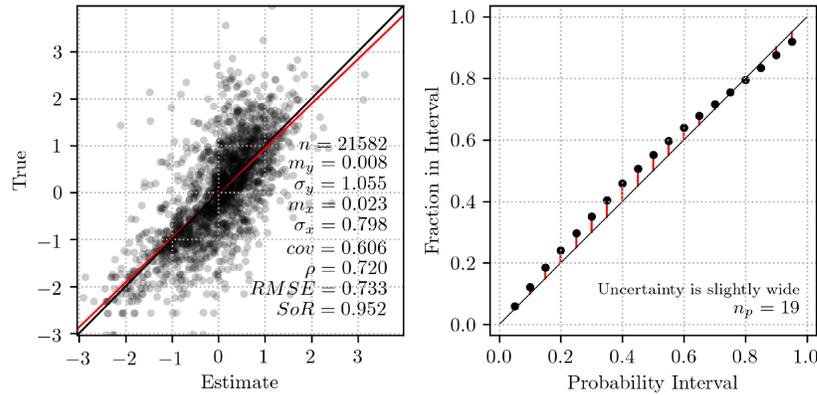


Figure 6.6: Fitted normal score variograms in the major, minor and tertiary directions (left to right).

Table 6.3: Normal score variogram model parameters.

	Nugget	Structure 1	Structure 2
Contribution	0.000	0.668	0.332
Model Shape		exponential	exponential
Angle 1		277.0	277.0
Angle 2		20.0	20.0
Angle 3		30.0	30.0
Range 1		47.6	300.0
Range 2		45.5	124.9
Range 3		45.5	100.0

**Figure 6.7:** Normal score variogram leave-one-out cross-validation (LOOCV) (left) and accuracy plot (right).

The final step of exploratory data analysis (EDA) is to generate a reference model for validation purposes. An exhaustive model is generated with ordinary kriging (OK) considering a minimum of 6 and a maximum of 40 composites using a search ellipse with ranges and angles aligned with the fitted original units variogram model (Table 6.1).

6.2 Network Parameter Inference

Parameterization of the NMR requires the definition of (1) optimization targets, (2) a pool of latent Gaussian factors, (3) network architecture and (4) optimization hyperparameters. The continuous normal score and indicator variograms from Section 6.1 are optimization targets for parameterizing the NMR. Target run distributions are calculated directly from the drillhole data. The n -point connectivity function is not considered, as the case study focuses on indicators above a high-grade threshold. When considering indicators above a threshold, cumulative run-length frequencies and the n -point connectivity function are analogous, and one can be calculated from the other. Thus, considering both in the objective function is redundant and would increase computation time.

The initial step in defining the Gaussian pool is checking how well a traditional simulation algorithm (e.g. SGS) reproduces the optimization target only considering the normal score variogram.

This “base-case” model provides insight into the required features to achieve the desired targets. SGS performs well reproducing the normal score and all indicator variograms, likely due to the data density of the training dataset. Though variograms are well reproduced, there is a discrepancy between downhole sequences from the data and the base-case model. Figure 6.8 shows cumulative run length frequencies above the 0.9 quantile calculated directly from the data (red) and the expected value of gridded SGS realizations (black). This discrepancy suggests additional high-grade continuity is required up to approximately 30m (six connected five-meter composites).

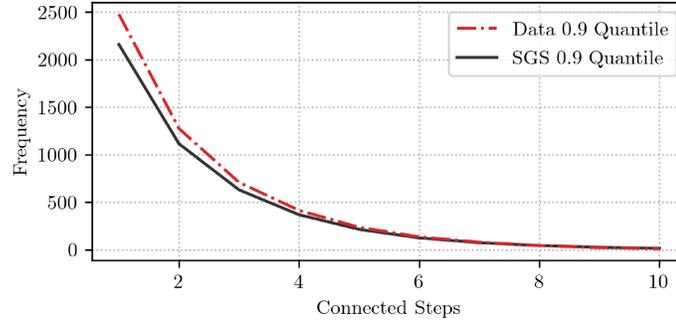


Figure 6.8: Cumulative run length frequencies above the 0.9 quantile from the drillhole data and the base-case SGS model.

The decomposition of the continuous normal score variogram structures into individual components forms the initial latent Gaussian pool. Each nested variogram structure defines the covariance of a latent factor with zero nugget and sill of 1.0. As the continuous variogram also reproduces the indicator variograms, the initial pool consists of the two nested structures in Table 6.3 for two initial factors. Additional structures are added to the pool that correspond with high-grade features of the conceptual geological model and the overall orientation of drilling. Factor 3 is strongly anisotropic in the vertical direction to capture the continuity of sub-vertical, high-grade breccia pipe structures. Ranges of the vertical factor are exaggerated as they will be reduced through mixing in the final model. The fourth and final factor is the nugget effect. As a pure nugget factor is added to the pool, any structure with a major range of less than 30 meters is pruned to prevent adding too much noise to the network inputs. Though the network can filter this noise, experiments show that some level of pruning improves the final parameter inference. Table 6.4 shows the covariance structure of each component in the final pool, excluding the nugget factor. Factors 1 and 2 are from the normal score variogram model and factor 3 is the additional high-grade structure.

The choice of activation function exponent, ω , controls each factor’s influence on either high or low values. $\omega < 1$ emphasizes low values (<0 in Gaussian units) and $\omega > 1$ emphasizes high values (>0 in Gaussian units). Table 6.5 shows the minimum and maximum bound for each trainable ω parameter. Factors 1 and 2 are essentially unconstrained and free to influence either high or low values. Factor 3 is constrained to influence high values only. The nugget effect factor is constrained

Table 6.4: Covariance structures of the Gaussian pool (excluding the nugget).

	Factor 1	Factor 2	Factor 3
Contribution	1.000	1.000	1.000
Model Shape	exponential	exponential	spherical
Angle 1	277.0	277.0	0.0
Angle 2	20.0	20.0	-90.0
Angle 3	30.0	30.0	0.0
Range 1	47.6	300.0	200.0
Range 2	45.5	124.9	100.0
Range 3	45.5	100.0	100.0

to $\omega = 1$. The sigmoid weight parameter controls the weight to remaining factors when the factor with precedence is above or below a particular threshold. A negative weight emphasizes high values, and a positive weight emphasizes lows. As the weight magnitude increases, the threshold approaches zero and trends towards a binary step function. The weight to other factors is ≈ 0 when $y_3 > 3.0$ with a sigmoid weight of -1.5.

Table 6.5: ω bounds by factor.

	ω_{min}	ω_{max}	precedence	σ wt
Factor				
1	0.25	4.00	3	-
2	0.25	4.00	2	-
3	1.00	4.00	1	-1.50
4	0.25	4.00	4	-

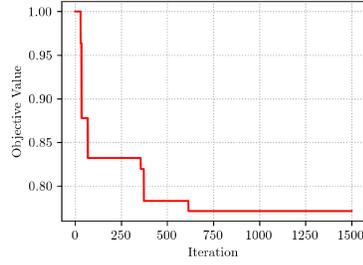
The number of factors in the Gaussian pool dictates the network architecture. The pool consists of three structures, plus one for the nugget, for four latent Gaussian factors, resulting in an input layer of four nodes. Each input node has a single connection weight to a single hidden layer of the same dimension as the input. The network has $2 \cdot (M + 1) = 8$ trainable parameters: four connection weights, and four ω values. Table 6.6 shows DE parameters. The indicator thresholds used for optimization correspond to the indicator variograms in Section 6.1, that is, $G^{-1}([0.1, 0.5, 0.9])$ where G^{-1} is the inverse of the Gaussian CDF. Runs consider a maximum of 10 steps ($\approx 50m$).

Table 6.6: Differential Evolution parameters.

	Population	Lower bound	Upper bound	F	CR lo	CR hi	Iterations
value	30	0.25	1.00	0.80	0.10	0.80	1500

Network optimization runs for 1500 iterations, resulting in a reasonably stable objective function value. The objective function value (Figure 6.9a) does not reach zero as it considers the expected value across all realizations. Though the objective function components are weighted, any deviation in the number of runs can lead to larger objective function values. Table 6.9b shows the final optimized connection weight and ω parameters. Figure 6.10 shows plan-view, east-west and north-south sections through the first unconditional realization at the data locations generated through parameter inference. Objective components are calculated using these realizations.

6. Case Study



(a) DE objective value versus iteration.

	Weight	ω
Factor		
1	0.77	3.98
2	0.43	3.72
3	0.34	1.00
4	0.15	1.00

(b) Optimal connection weight and ω parameters.

Figure 6.9: DE objective function value (a) and final optimized network parameters (b).

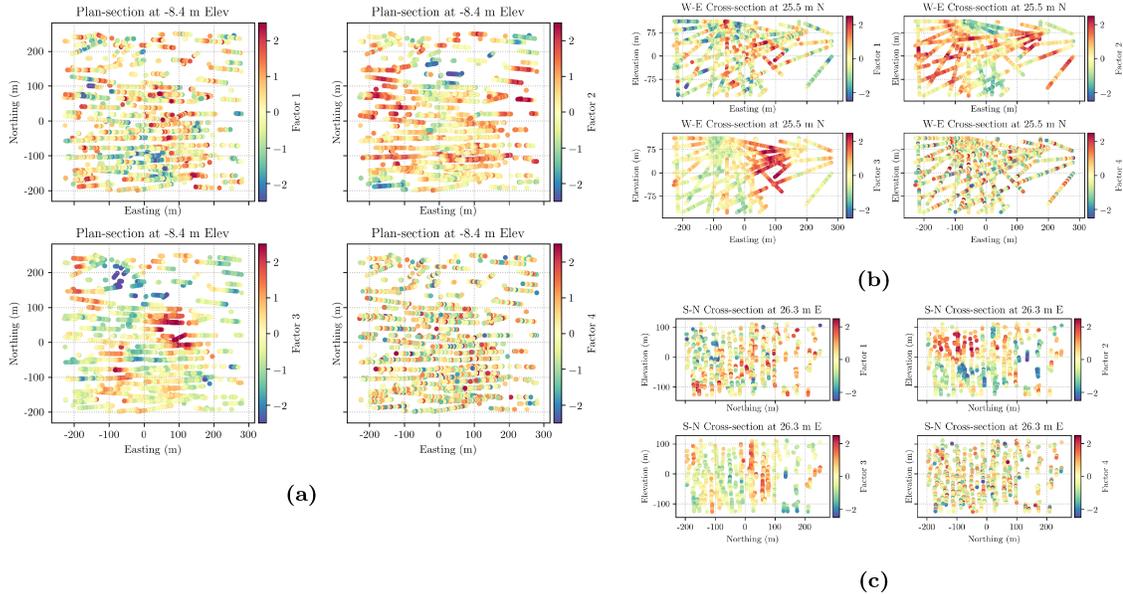


Figure 6.10: Plan (a), east-west (b) and north-south (c) sections through the first unconditional realization of Gaussian values generated through parameter inference. Objective function components are evaluated using these realizations.

A critical component of the workflow is checking the reproduction of the optimization targets. Deviations from these targets will likely manifest in the imputed conditioning data and final realizations. Figures 6.11 to 6.13 show variogram, indicator variogram and cumulative run-length frequency reproduction with the optimal NMR parameters. Variogram and indicator variogram reproduction are reasonable, though there is some deviation from the target models, particularly at shorter ranges in the principal directions. Cumulative run-length frequencies show good reproduction for all quantiles.

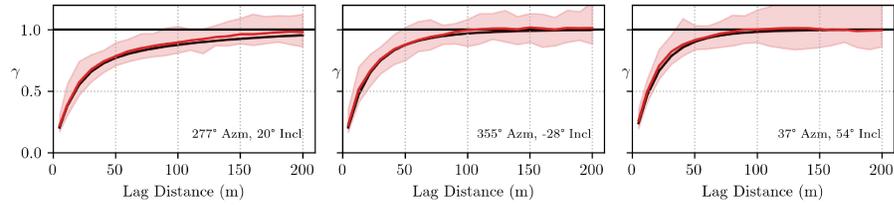


Figure 6.11: Continuous variogram reproduction for 25 realizations. The black line is the variogram model, the red line is the average variogram, and the shaded red area encloses the minimum and maximum variogram values. Left to right are the major, minor and vertical directions, respectively.

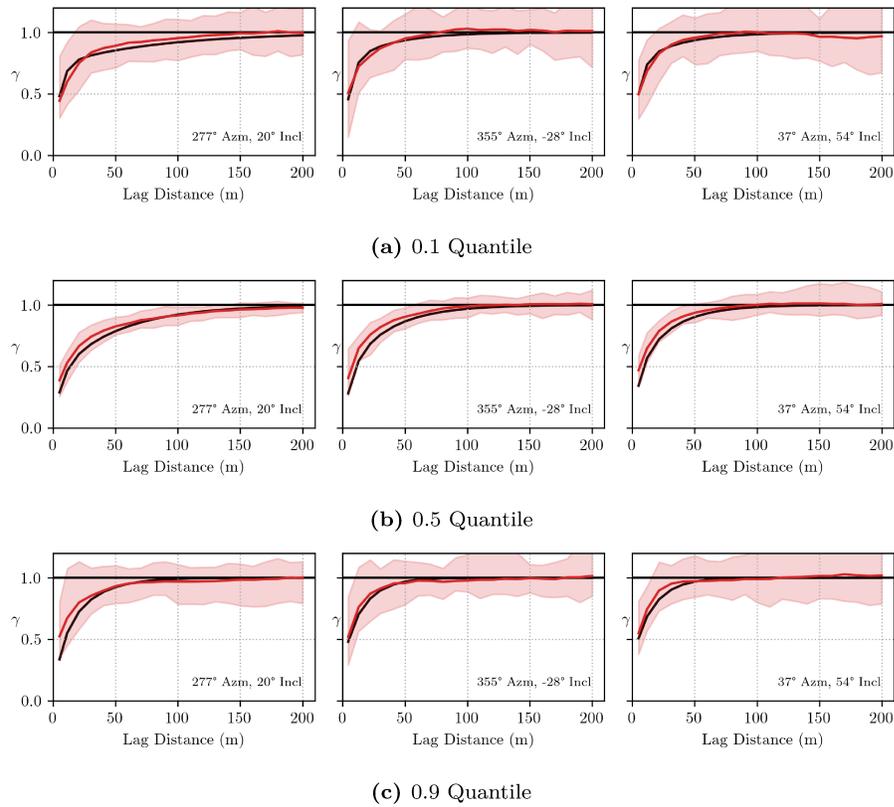


Figure 6.12: Indicator variogram reproduction for 25 realizations for the 0.1 (a), 0.5 (b) and 0.9 (c) quantiles. Left to right are the major, minor and vertical directions, respectively. The black line is the variogram model, the red line is the average variogram, and the shaded red area encloses the minimum and maximum variogram values.

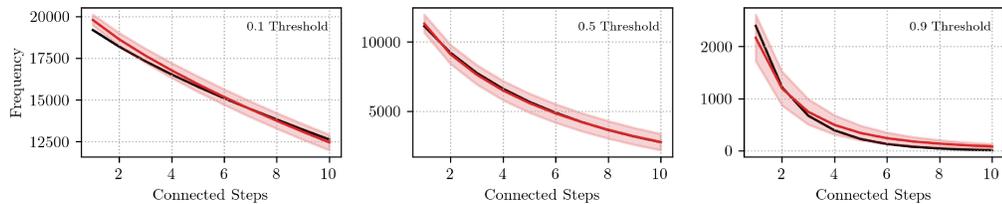


Figure 6.13: Cumulative run-length frequency reproduction for 25 realizations. Thresholds 1, 2, and 3 correspond to the 0.1, 0.5 and 0.9 quantiles. The black line is the target run-length frequency model, the red line is the average experimental value, and the shaded red area encloses the minimum and maximum experimental values.

6.3 Latent Factor Imputation

The optimized NMR parameters from Section 6.2 are the basis for latent factor imputation. The optimized mapping function, \mathcal{F}_θ , and Gaussian pool are imputation inputs, simulating latent factor realizations via SGRI. A high-grade constraint threshold at the 0.9 quantile (in Gaussian units) and an exclusion radius of 40 m are enforced for imputation. Enforcing the high-grade constraint ensures that some high values of the factor with precedence are seeded where the observed data values are high. The exclusion radius ensures that not all of these locations are seeded, as seed locations must be a minimum of 40 m apart. At the final seed locations, the value of the factor with precedence must be above the threshold. Seeding too many imputation locations may cause a strong departure from the standard normal distribution. Using the 0.9 quantile and a 40 m radius leads to $\approx 1.2\%$ of the data locations being seeded. 25 realizations are imputed for each factor specified in the Gaussian pool, plus the nugget effect. All imputed realizations are checked to ensure they (1) reproduce the data within the specified polishing tolerance (Table 6.7); (2) imputed realizations are on average univariate Gaussian; (3) imputed factors are on average uncorrelated, and (4) each imputed factor on average reproduces its single structure variogram model.

Table 6.7: Imputation parameters.

Parameter	Value
Max nodes	40
Max iterations: coarse	25000
Max iterations: polish	15000
Reject tolerance: coarse	0.10
Reject tolerance: polish	0.01
Constraint quantile	0.90
Exclusion radius	40

Figure 6.14 shows a scatter plot matrix of the original data values, plus the first realization of the mapped imputed values and all latent factors. Scatters are coloured by KDE where warmer colours correspond to a higher density of points. The top row of the matrix is somewhat redundant; however, it highlights the perfect correlation between the observed data and mapped imputed values. The marginal histograms show that each latent distribution is standard normal, with reasonable statistical fluctuation. The scatter plots between each latent factor show roughly concentric bivariate Gaussian density contours and $|\rho| < 0.10$. The average correlation coefficient across all realizations is within this tolerance, suggesting the latent factors are effectively uncorrelated. Figure 6.15 shows variogram reproduction for each imputed factor. There is some deviation from the specified target models, though overall, variogram reproduction is reasonable. Deviation from the specified model is largely present in factor 2, which shows decreased short-range continuity. This deviation is attributed to the effect of strong conditioning and does not appear to impart any negative consequences in the final NMR model.

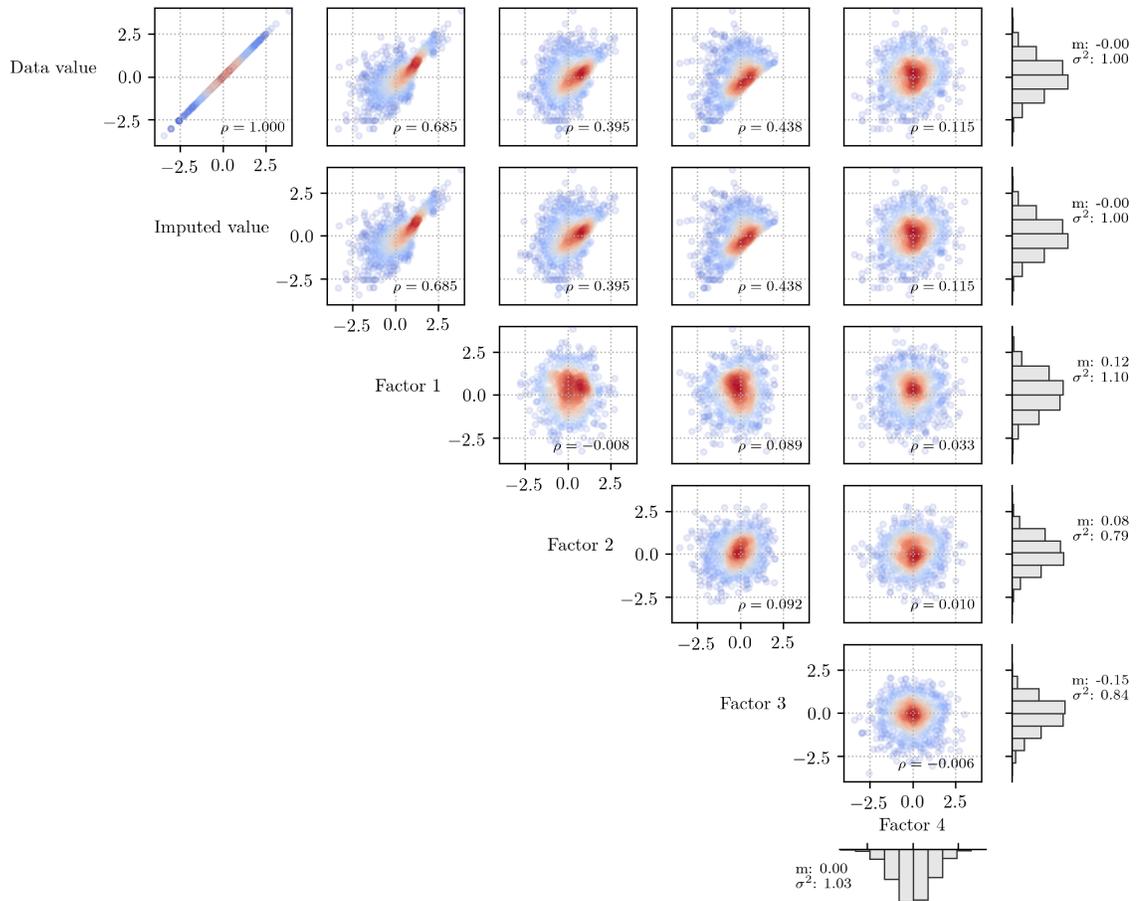


Figure 6.14: Scatter plot matrix coloured by KDE illustrating the relationships between the observed data values, mapped imputed values and latent factors in Gaussian units. The histograms show the marginal distributions of each variable by row.

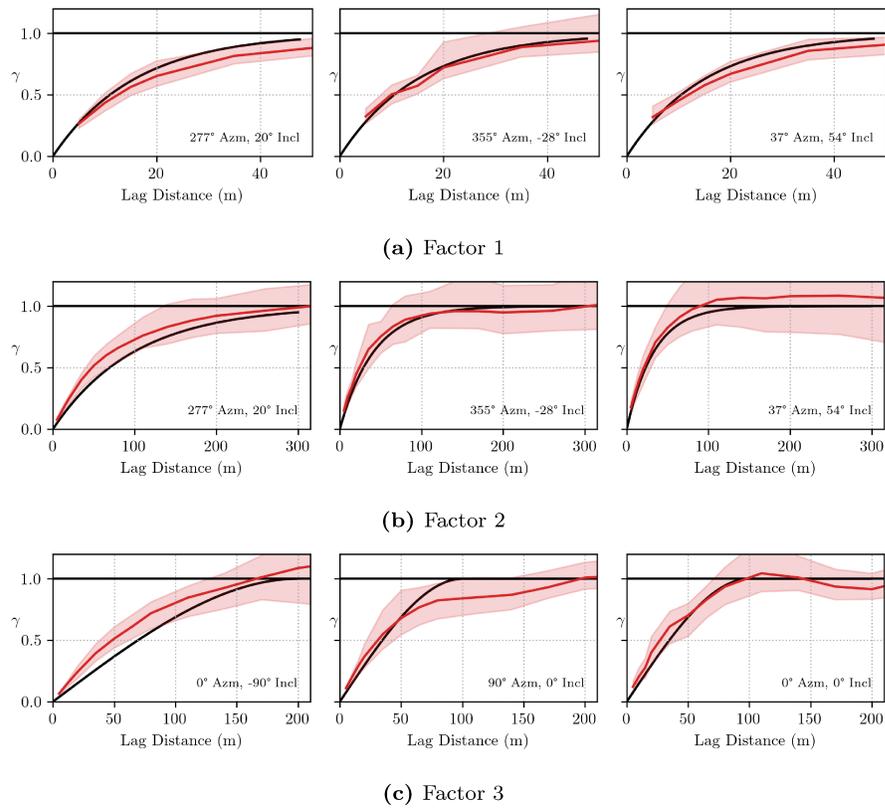


Figure 6.15: Variogram reproduction for 25 imputed realizations of each factor. The black line is the variogram model, the red line is the average variogram, and the shaded red area encloses the minimum and maximum variogram values. Table 6.4 specifies the Gaussian pool.

6.4 Latent Factor Simulation

Imputed factor realizations from Section 6.3 become conditioning data for gridded factor realizations. Each univariate Gaussian factor realization conditions a gridded realization in a one-for-one fashion. That is, each simulated realization uses a unique data realization. Gridded realizations in this case study are simulated using SGS though any conditional simulation algorithm is valid. The point scale simulation grid utilizes a cell size of $2.5 \times 2.5 \times 5\text{m}$ for 2,236,948 nodes; Table 6.8 summarizes the parameters of the grid. 25 realizations of each factor are simulated and checked similarly to any other Gaussian simulation workflow (Leuangthong, McLennan, et al., 2004). The realizations are, on average, univariate Gaussian and reproduce the variogram. Figure 6.16 shows an east-west section through the first gridded realization of the latent factors to highlight the effect of seeding a particular factor. The fourth panel shows the 0.9 quantile indicator transform of the latent factors mapped to observed space with the 0.9 quantile indicator transform of the Gaussian variable. As factor 3 is seeded and given high-grade preference during imputation, the latent high-grade features correspond with high-grade regions in the observed data.

Table 6.8: Simulation grid parameters.

	Easting	Northing	Elevation
minimum (m)	-256.08	-227.04	-130.86
maximum (m)	308.92	277.96	114.14
size (m)	2.50	2.50	5.00
number	226	202	49

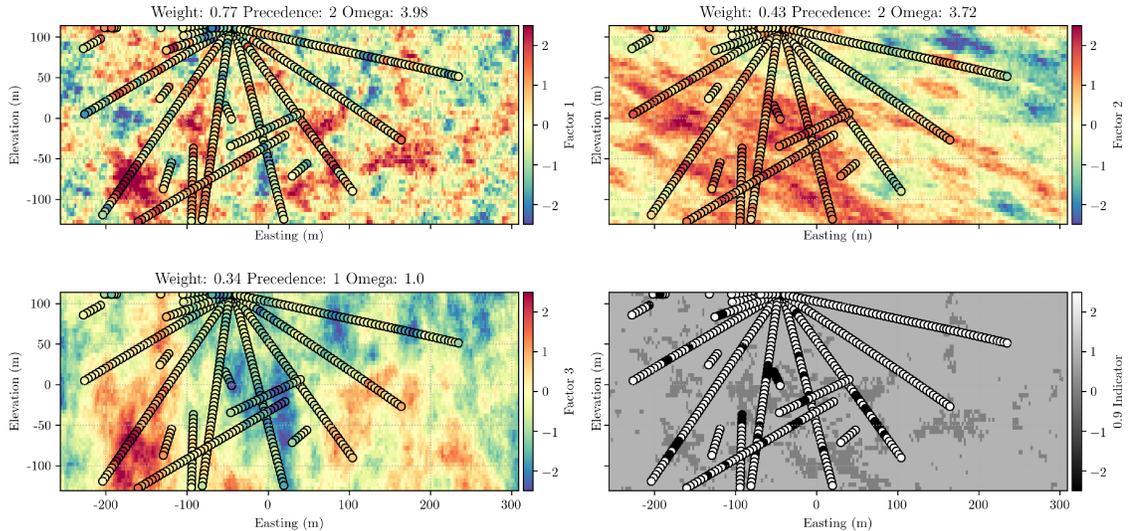


Figure 6.16: East-west section through the first realization of the gridded factors, plus the 0.9 quantile indicator transform of the latent factors mapped to observed space.

Once factors are defined at all grid nodes, they are mapped from latent to observed space by applying $\mathcal{F}_\theta(\mathbf{y})$ where \mathbf{y} is the vector of gridded Gaussian realizations. Mapping consists of inter-

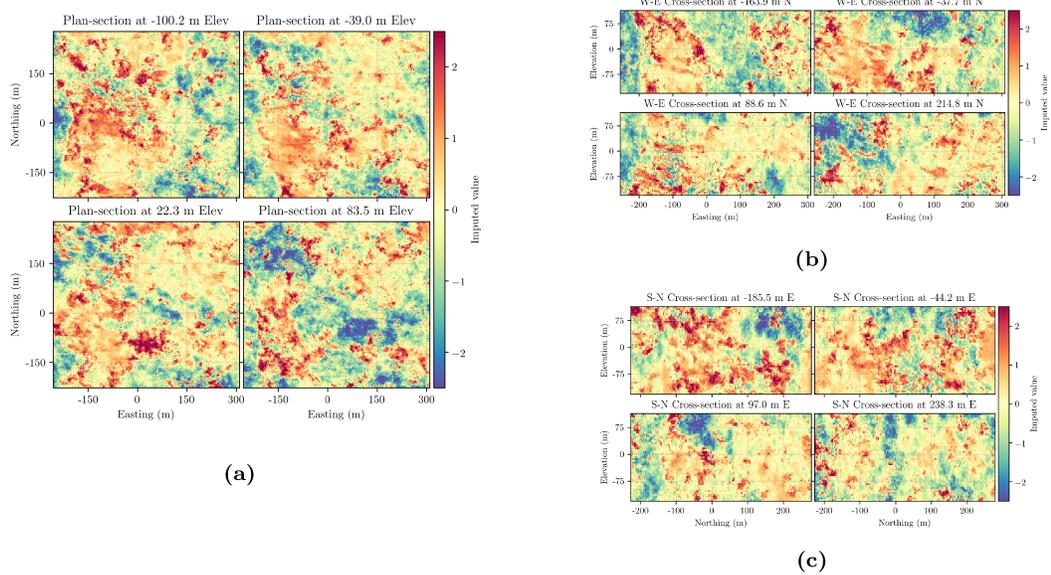


Figure 6.17: Plan (a), east-west (b) and north-south (c) sections through the first NMR realization mapped to Gaussian units.

polating the raw activation values (the NMR output) with the normal score reference distribution generated from imputation. Using a single reference distribution ensures consistency between imputation and gridded simulation. The NMR realizations are checked to ensure they reproduce the continuous variogram model, indicator variogram models, and cumulative run-length frequencies, and the realizations are, on average, univariate Gaussian. Figure 6.17 shows plan-view, east-west and north-south sections through the first NMR realization mapped to Gaussian units. Figure 6.18 shows CDF reproduction of the NMR realizations in Gaussian units (left panel). The realizations show reasonable variance given the level of conditioning and are, on average, univariate Gaussian. Figure 6.19 shows plan-view, east-west and north-south sections through the first NMR realization back-transformed to original units. These figures show the vertical and downhole high-grade continuity imparted by factors 3 and 4 (Table 6.4). The overall background continuity imparted by the continuous variogram model dips to the east, and is clear in the east-west sections (Figure 6.19 (b)). Figure 6.18 shows CDF reproduction of the NMR realizations in back-transformed to original units (right panel). The realizations reproduce the declustered histogram of the training data well.

Figure 6.20 shows gridded continuous variogram reproduction for the NMR realizations in Gaussian units. The experimental variogram of the first imputed data realization and the average of SGS realizations are shown for reference. There is reasonable agreement between the target model (black line), the imputed data (red dots) and the gridded realization average (red line). There is some deviation; however, these are the same deviations present from the inference of the network parameters (Figure 6.11). Indicator variogram reproduction in Figure 6.21 shows similar deviations observed in the optimization process, though the gridded variograms reproduce the target model and imputed

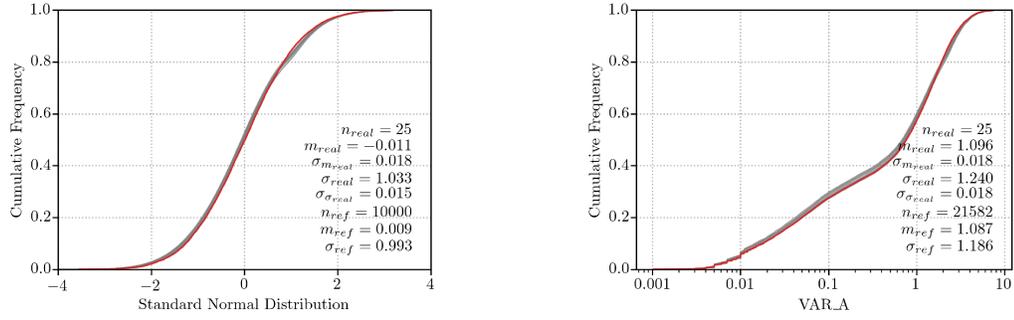


Figure 6.18: CDF reproduction for 25 NMR realizations mapped to Gaussian units (left) and original units (right).

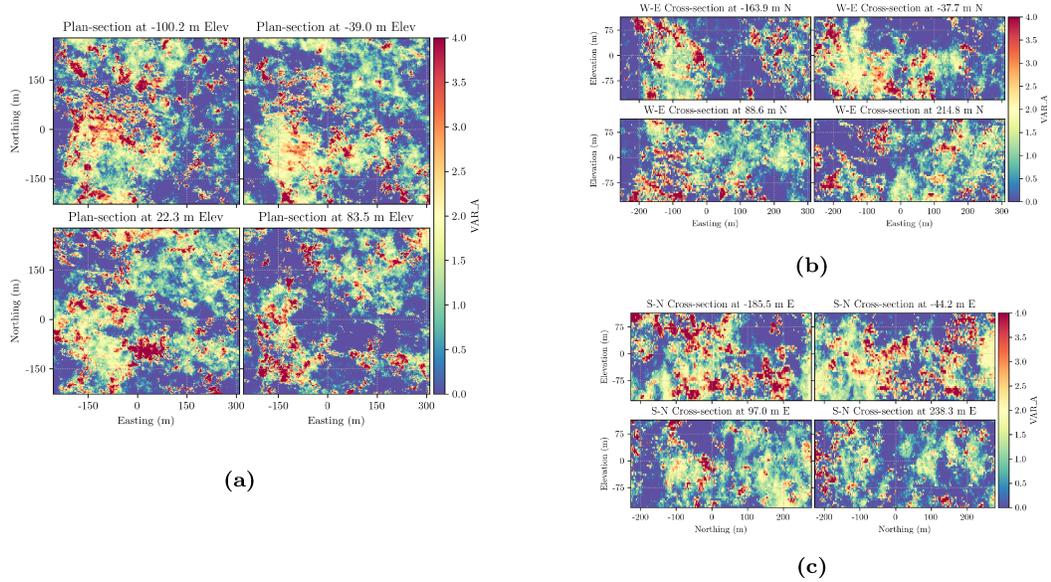


Figure 6.19: Plan (a), east-west (b) and north-south (c) sections through the first NMR realization back-transformed to original units.

data reasonably well. Figure 6.22 shows gridded cumulative run-length frequency reproduction for both the NMR and SGS realizations, respectively. The expected value of the SGS realizations is shown as the solid black line. As expected, the NMR realizations show increased continuity over SGS for all connected steps up to a run length of 6 composites. This increase in downhole continuity is consistent with observations made when designing the Gaussian pool (Figure 6.8).

6.5 Post Processing

Point scale realizations are block-averaged to the SMU volume for comparison at a relevant scale. The SMU scale grid extents are consistent with the simulation grid, though subsequent analyses are focused on the west-southwest portion of the grid that encompasses most of the test set. Table 6.9 shows the parameters of the SMU grid. As a goal of the NMR is to improve local high-grade

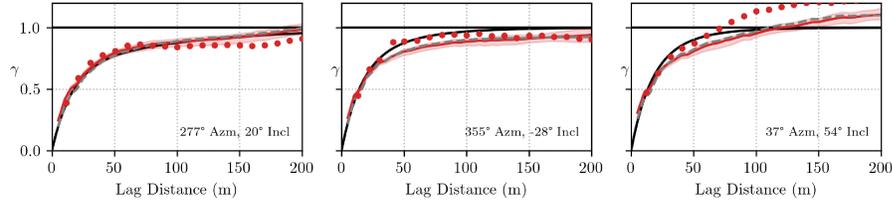
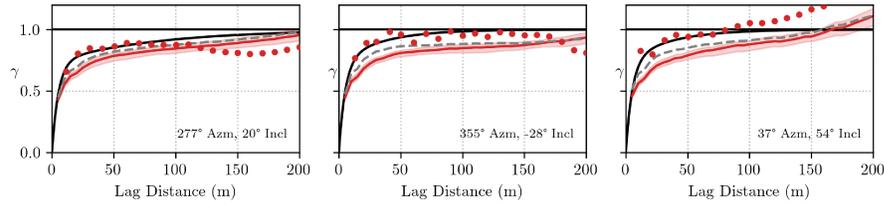
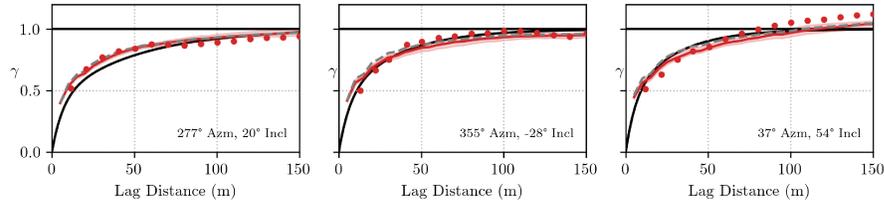


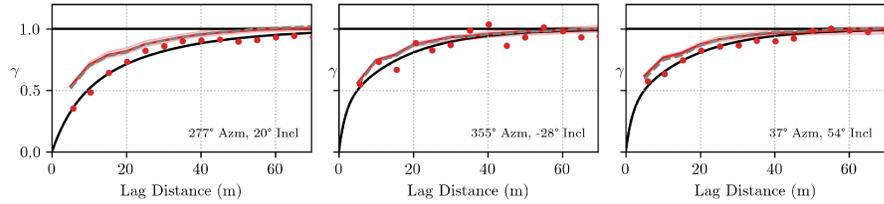
Figure 6.20: NMR normal score variogram reproduction for 25 realizations. The black line is the variogram model, the red dots are the experimental variogram of the imputed data, the red line is the average variogram of the gridded NMR realizations, the shaded red area encloses the minimum and maximum gridded NMR variogram values, and the dashed grey line is the average variogram of the gridded SGS realizations. Left to right are the major, minor and vertical directions, respectively.



(a) 0.1 Quantile



(b) 0.5 Quantile



(c) 0.9 Quantile

Figure 6.21: NMR indicator variogram reproduction for 25 realizations for the 0.1 (a), 0.5 (b) and 0.9 (c) quantiles. The black line is the variogram model, the red dots are the experimental variogram of the imputed data, the red line is the average variogram of the gridded NMR realizations, the shaded red area encloses the minimum and maximum gridded NMR variogram values, and the dashed grey line is the average variogram of the gridded SGS realizations. Left to right are the major, minor and vertical directions, respectively.

resources conditional to certain events, a simplified stope optimization process is implemented to mimic the selection of local high-grade resources at the SMU scale. The reference OK model from Section 6.1 is used for stope optimization. Stopes with a fixed footprint of 30 x 30 m and a minimum height of 40 m are selected considering the OK reference model and a cutoff grade (COG) of 2.0 units. An initial grid of stope footprint centroids is generated across the reference model. All valid footprints about the centroid are evaluated, and the highest grade stope meeting the minimum

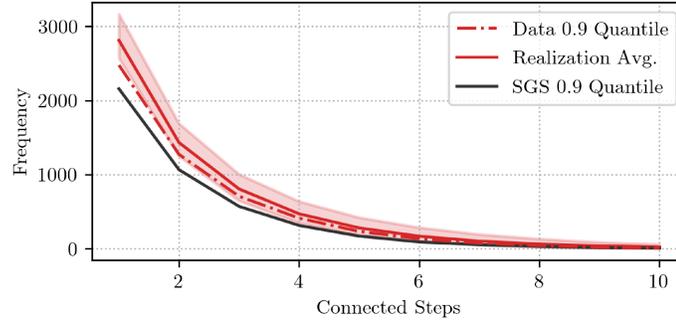


Figure 6.22: NMR 0.9 quantile cumulative run-length frequency reproduction for 25 realizations. 0.9 quantile from SGS is shown for reference.

volume requirement is selected. The initial stope's z-dimension is then expanded one grid level at a time until the volume is no longer above COG or a maximum height of 80 m is reached. This process is simplified and intended to be a straightforward proxy for mining selectivity. The resulting volumes are not true optimums. Figure 6.23 shows a plan view section through the reference OK model showing the stope shapes and the test data locations.

Table 6.9: SMU grid parameters.

	Easting	Northing	Elevation
minimum (m)	-256.08	-227.04	-130.49
maximum (m)	128.92	127.96	-0.49
size (m)	5.00	5.00	5.00
number	77	71	26

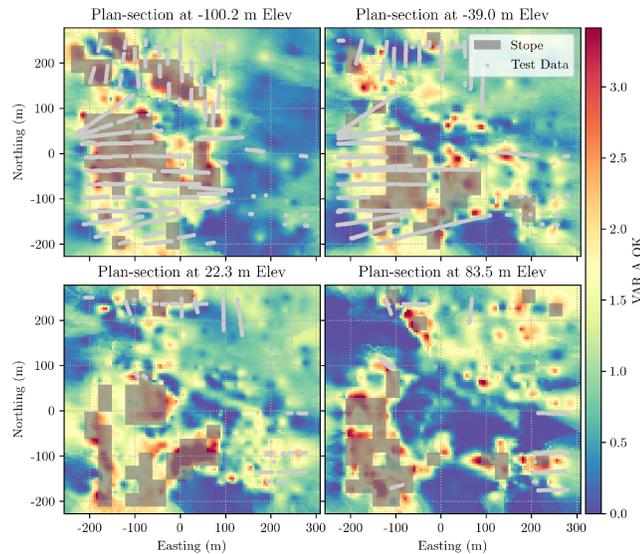


Figure 6.23: Plan view section through the reference OK model showing stope shapes and the test dataset.

Stopes containing composites from the test dataset are used to evaluate the performance of the NMR realizations relative to SGS in local, high-grade regions of the model. The test composites are

coded by stope ID and compared directly to simulated realizations within the same volume. The following section summarizes model performance.

6.6 Model Comparison

The following section compares the NMR and SGS realizations at a global scale and conditional to high-grade stope volumes. These tests show how well each set of realizations reproduces data the model has not seen. The main focus of the NMR implementation is getting the high-grade connectivity correct at the local stope volume. Modeling results are discussed globally across all test set data locations and locally within high-grade stope volumes. Recall from Figure 6.1 that the test set is systematically higher grade than the training set. Though the conceptual geologic model is embedded in the NMR, it is unrealistic to expect a data-driven model to reproduce unseen data that is significantly higher or lower grade. For this reason, stopes where the deviation between the test and training set's mean grade is greater than 20% are excluded from the comparison. Spatially partitioning complete drillholes into training and test sets with balanced features may not always be possible with geologic data.

6.6.1 Global Results

The most straightforward check of model performance is the prediction error at test set locations. Each value from the test set is compared to the simulated value at the data location for both NMR and SGS models. The prediction error is the difference between the true and simulated data values. Figure 6.24 shows the distributions of prediction errors across all realizations for both the NMR and SGS models. Both histograms are symmetric and centered about zero. The NMR model shows a slight reduction in error with a mean of -0.13 versus -0.14 for the SGS model. The negative means indicate that both models underestimate the true values, which is expected due to the increased mean grade of the test set relative to the training set. Figure 6.25 shows cross-validation scatter plots between NMR and SGS realizations and true data values. Again, the NMR model shows similar root mean squared error (RMSE) and the correlation between the estimate and the truth.

Probabilistic accuracy can be assessed by considering the local conditional CDFs defined by the set of simulated realizations at the test set locations. Probabilistic accuracy amounts to checking the number of true values that fall within a given probability interval. For example, 50% of the true values should fall within the 0.5 probability interval (between the 25th and 75th quantiles). The true value comes from the test set, and the realizations define the probability interval. Accuracy plots are generated by plotting the actual fraction of true values within the interval against the probability interval. An accurate and precise model will plot with points close to the 45-degree line. Figure 6.26 shows accuracy plots for the NMR and SGS models. Both models plot close to the 45-degree line, suggesting reasonable modeling techniques. Uncertainty is slightly wide for the NMR

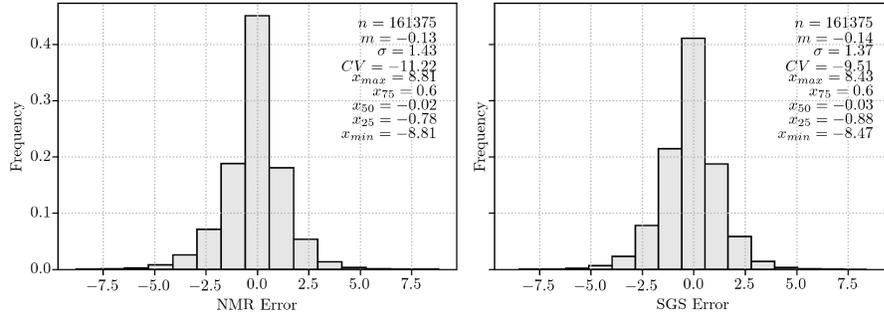


Figure 6.24: Distributions of prediction error for NMR (left) and SGS models (right) at all test set locations.

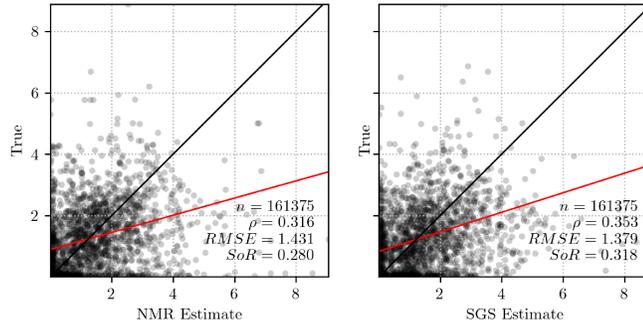


Figure 6.25: Cross validation scatter plots of NMR (left) and SGS (right) models and true data values at all test set locations. The black line is 1:1, and the red line is the regression of the truth on the estimate. RMSE=root mean squared error; SoR=slope of regression.

model, meaning the realizations are slightly higher variance compared to the test set; the same is true for the SGS model. The acceptable width tolerance is set to 5%; for both models, the average error (probability interval - fraction in interval) is less than this threshold. Accuracy and precision metrics are reported for a quantitative measure. However, little deviation from the 1:1 line is the most important feature (Deutsch, 2010) .

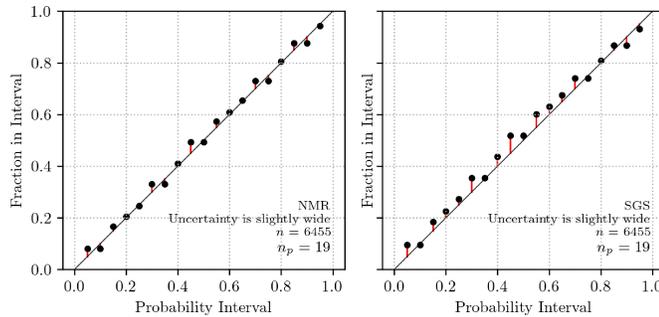


Figure 6.26: Accuracy plots of NMR (left) and SGS (right) models at all test set locations. n_p is the number of probability intervals.

At a global scale, the NMR model shows a modest improvement in prediction error, both in magnitude and RMSE with the truth. Both NMR and SGS models show good accuracy at the

global scale. The following section explores validation metrics at the local scale; the goal is that the NMR outperforms SGS conditional to high-grade stope volumes.

6.6.2 Local Results

Local results consider test set locations that fall within stope volumes. Prediction error, RMSE and accuracy plots are checked within stope volumes. Grade-tonnage (GT) curves are checked for all stope volumes and on a stope-by-stope basis. Stopes where the NMR outperforms SGS are investigated in greater detail. The NMR shows modest improvements compared to SGS overall. However, some stopes exist where SGS is superior. These stopes are also investigated in greater detail. Figure 6.27 shows distributions of prediction error for the NMR and SGS models. Again, the NMR shows a slight improvement with a mean error closer to zero than the SGS error distribution. Figure 6.28 shows cross-validation scatter plots for true and predicted values across all realizations for both models. The NMR model shows similar results in terms of RMSE and correlation coefficient relative to the SGS model.

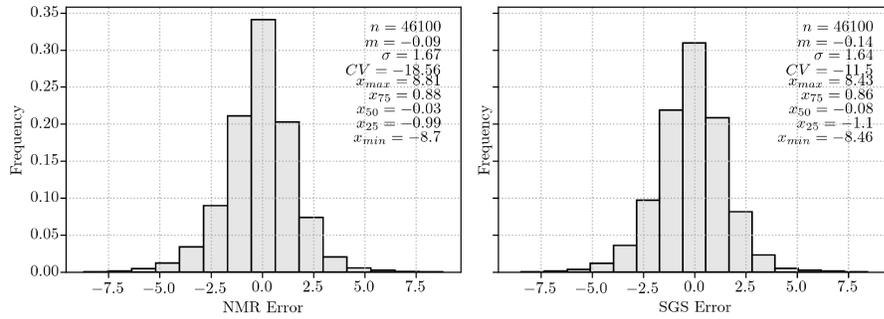


Figure 6.27: Distributions of prediction error for NMR (left) and SGS models (right) at test set locations flagged within stope volumes.

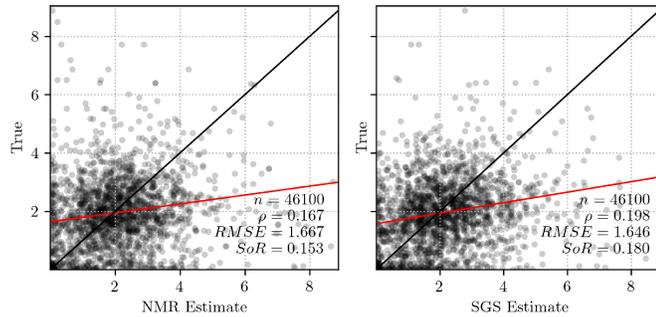


Figure 6.28: Cross validation scatter plots of NMR (left) and SGS (right) models and true data values at test set locations flagged within stope volumes. The black line is 1:1, and the red line is the regression of the truth on the estimate. RMSE=root mean squared error; SoR=slope of regression.

Figure 6.29 shows accuracy plots for test set data values flagged within stope volumes. Again, both models plot close to the 45-degree line though the NMR model slightly under-predicts local

uncertainty compared to all test set locations in Figure 6.26. This reduction of local uncertainty is likely due to the effect of strong conditioning and constraints enforced during the imputation of factor 3 in high-grade locations. The SGS model shows similar accuracy within the stope volumes.

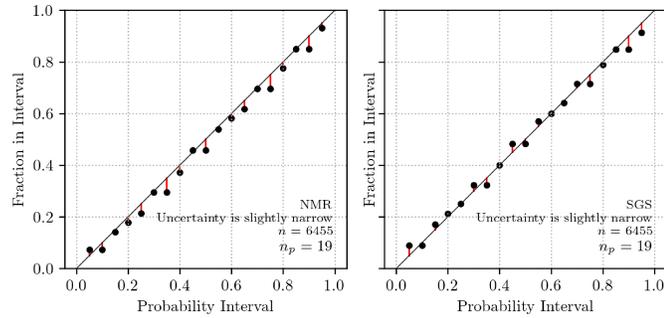


Figure 6.29: Accuracy plots of NMR (left) and SGS (right) models at all test set locations. n_p is the number of probability intervals.

Figure 6.30 shows GT curves for all stope volumes for the NMR (left) and SGS (right) SMU scale realizations. Test data tonnes and grades come from weighted composites in the test set flagged within the stope volumes. Across all stopes, the SGS model is similar in grade and higher in tonnage than the test composites. NMR realizations are overall lower in grade and tonnage relative to the SGS realizations and more closely reproduce the test data. This reproduction is likely due to the anisotropies of certain factors constraining high-grade features rather than being controlled by a single continuous variogram (that considers the full range of grade values) in the SGS context.

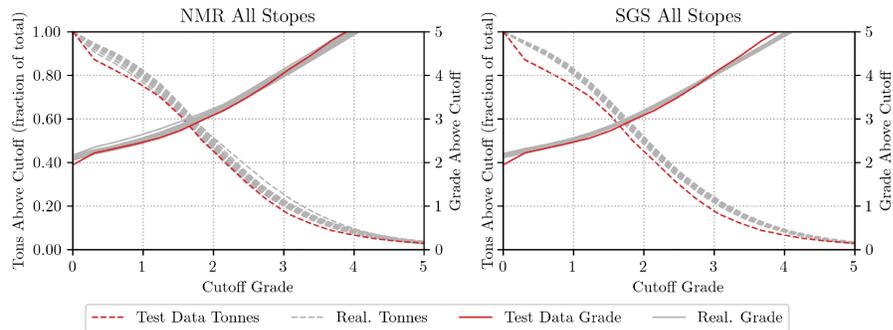


Figure 6.30: GT curves for all stope volumes for the NMR (left) and SGS (right) models.

Table 6.10 shows more detailed grade and test set deviation on a stope-by-stope basis. The table shows the number of test data in each stope, the mean grade of the composites, and the percent deviation from the test data mean for the OK, SGS and NMR expected values. Training and test set mean values are the declustered mean grade of the respective data set, flagged within the stope volume. Gridded mean values are the equal-weighted mean of all blocks flagged with that stope ID, as no density data is available. Values are sorted in ascending order by the training set mean grade. In some situations, all models perform poorly, both over- and under-estimating the mean of the test

data. These large deviations occur in stopes where the training and testing data are unbalanced. That is, there is a significant deviation between the mean of the conditioning data and the mean of the hold-out data. Figure 6.31 shows the percent deviation from the test data mean for the training data and NMR and SGS expected values within each stope. Subsequent comparisons consider stopes where this deviation is less than 20%. Where the training and testing data are balanced concerning grade, the NMR model performs well relative to SGS, particularly in higher grade stopes. For stopes 39, 158 and 40, the NMR shows improvement over SGS, with stope 39 showing a 7% improvement relative to the test set mean. Notably, the NMR almost always under-predicts the mean relative to SGS. All other stopes show minor improvements except for stope 641, 47 and 1138, where the test set mean is greater than the training set and the NMR under-predicts.

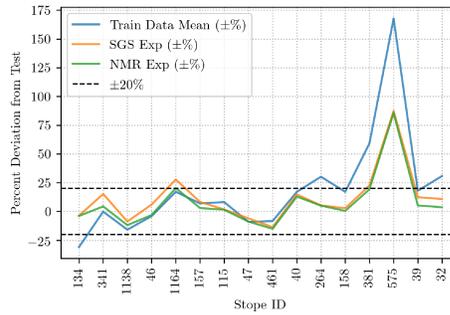


Figure 6.31: Percent deviation from the test data set by stope ID. Stopes where train/test deviation is less than 20% are retained for comparison.

Table 6.10: % difference in expected value between OK mean, SGS expected value and NMR expected value relative to test set mean by stope (≥ 2.0 unit COG). The table shows stopes with greater than 30 test data and test set deviation of less than 20%. Values are sorted in ascending order by training data mean value.

Stope ID	Num Test Data	Train Data Mean	Test Data Mean	OK Mean ($\pm\%$)	SGS Exp ($\pm\%$)	NMR Exp ($\pm\%$)
341	40	1.85	1.86	7.7	15.2	4.4
1138	39	1.95	2.32	-13.2	-8.8	-12.0
46	40	1.99	2.08	2.6	5.9	-3.1
1164	30	2.06	1.76	15.3	27.8	20.3
157	52	2.08	1.95	8.6	8.3	2.9
115	73	2.15	1.98	1.6	1.7	1.4
47	31	2.20	2.42	-6.9	-6.0	-8.9
461	31	2.29	2.49	-15.8	-13.7	-15.1
40	33	2.34	2.00	10.5	14.8	12.9
158	33	2.56	2.19	3.9	2.8	0.4
39	38	2.72	2.31	8.1	12.4	5.2

Stope 39 is a high-grade stope where the NMR performs well relative to SGS. Figure 6.33 shows the NMR vs SGS GT curve for stope 39. These GT curves show similar characteristics to Figure 6.30 where the SGS model is higher grade through virtually all cutoff grades. The NMR model shows slightly better reproduction of the test composites grades and the fraction of total tonnes

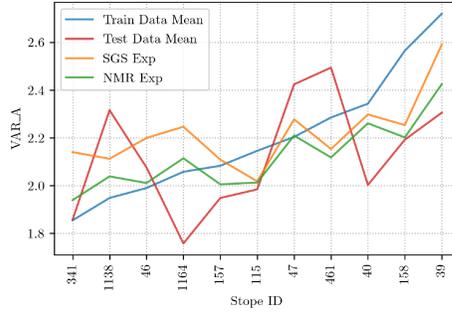


Figure 6.32: Mean grades by stope ID where test set deviation is less than 20%, and at least 30 data are present in the stope volume.

above COG. Figure 6.34 shows east-west sections through the spatial centroid of stope 39 with the reference OK model, the SGS e-type and the NMR e-type models with the training and test data. Visually, the NMR realization does a good job of reproducing the test data, and vertical high-grade continuity imparted by factor 3 is apparent. Stope 47 is a high-grade stope where SGS performs well relative to the NMR. Figure 6.35 shows the NMR vs SGS GT curve for stope 47. The NMR model appears to reproduce the grade of the test composites well; however, the tonnage above cutoff is less for almost all cutoffs. This reduction in tonnage above cutoff leads to a lower expected value relative to the test set and SGS. Factor 3 imparts vertical continuity visible to the east of the stope volume in the bottom panel of Figure 6.36. Rather than east-dipping continuity characterized by the continuous variogram, this vertical continuity constrains the high-grade mineralization between the training set drillholes. This constraint underestimates the test set grades given the deterministic model with anisotropy defined by the continuous variogram.

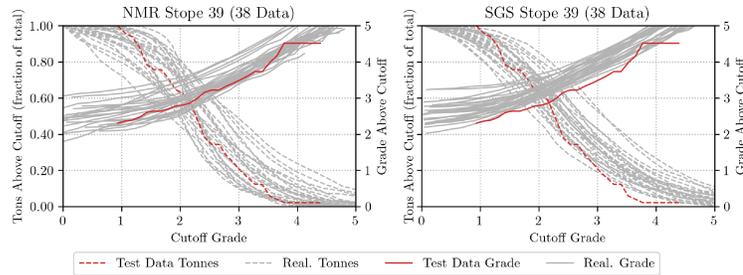


Figure 6.33: GT curves for stope 39 for NMR (left) and SGS (right) models. Stope 39 is a high-grade stope where the NMR performs well relative to SGS.

Throughout all cutoff grades the NMR realizations are lower grade than the SGS realizations. This reduction in grade leads to better reproduction of the test data across all stope volumes. However, discrepancies exist stope-by-stope. The NMR outperforms SGS when SGS overestimates the true grade. If SGS underestimates the true grade, typically the NMR further underestimates. This phenomenon is likely due to strongly anisotropic latent factors imparting high-grade continuity (factor 3) that form a “hard” boundary with background mineralization. This feature is evident in

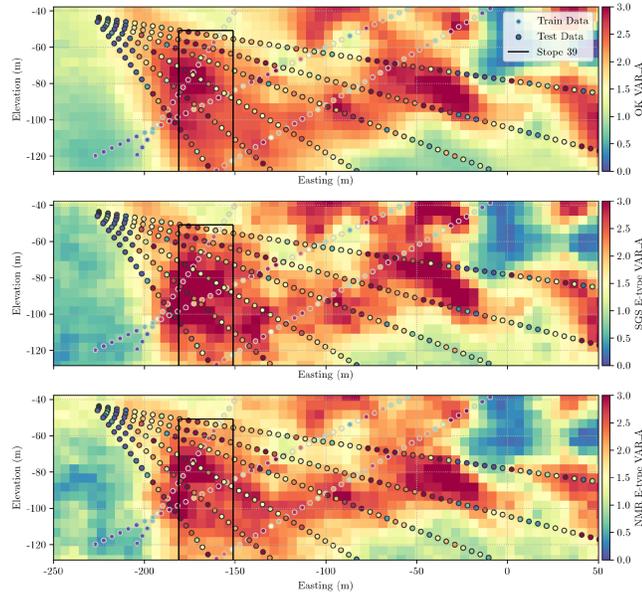


Figure 6.34: Sections through stope 39 showing the test data, OK model, and the first block averaged realization of the SGS and NMR models. Stope 39 is a high-grade stope where the NMR performs well relative to SGS. Training data is outlined in gray circles, while test data is in black.

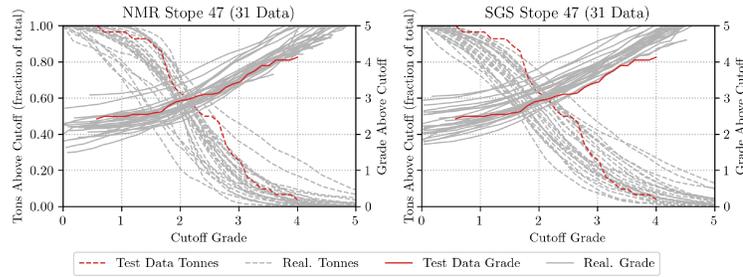


Figure 6.35: GT curves for stope 47 for NMR (left) and SGS (right) models. Stope 47 is a high-grade stope where SGS performs well relative to the NMR.

the bottom panel of Figure 6.34 where sharper vertical boundaries are present between high-grade structures (>3.0 units) and lower-grade background mineralization. Stope placement, considering a smooth deterministic model, may not be able to capture these features adequately in probabilistic realizations.

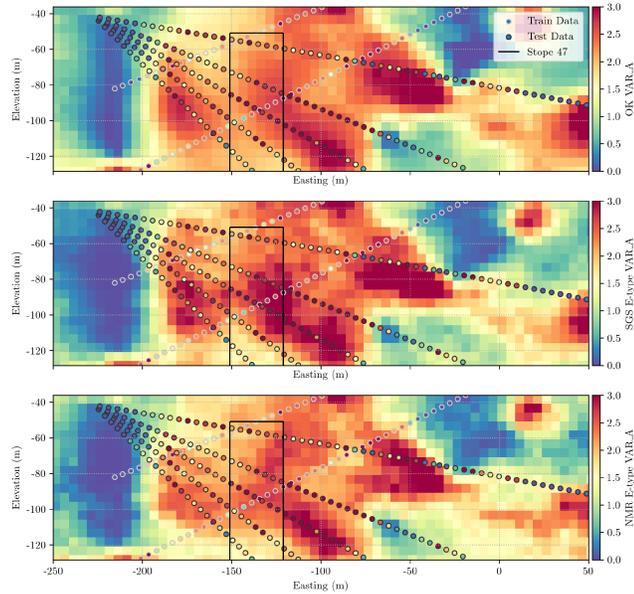


Figure 6.36: Sections through stope 47 showing the test data, OK model, and the first block averaged realization of the SGS and NMR models. Stope 47 is a high-grade stope where SGS performs well relative to the NMR. Training data is outlined in gray circles, while test data is in black.

6.7 Discussion

The NMR can reproduce high-grade continuity features that cannot be captured with a continuous variogram alone. The specification of ω bounds allows the practitioner to impart specific spatial features onto specific portions of the continuous grade range. The resulting NMR realizations are univariate Gaussian; however, they capture non-Gaussian connectivity features. In this case study, the connectivity of extreme values is particularly important. A base-case SGS model shows that realizations generated under a multivariate-Gaussian assumption cannot alone capture the high-grade continuity observed in the drillholes. The NMR parameter inference workflow determines the optimal weight of each factor and the exponent ω for the network activation function. Parameters are optimized to reproduce continuous and indicator variograms and cumulative run-length frequencies observed in the drillhole data in the final gridded realizations. The indicator variograms do not show significant asymmetry in this case study, though the NMR can reproduce these non-Gaussian features. In addition, high-grade continuity, above what is observed in the drillholes, is embedded in the NMR through highly anisotropic factors where $\omega \geq 1$. This increased continuity is required when the conditional simulation of the Gaussian factors leads to the destructuring of extreme values after passing through the NMR.

Across all stope volumes, the NMR realizations more accurately reproduce the grade and tonnage of the test dataset than the SGS realizations. On a stope-by-stope basis, the performance of the NMR model is variable, though major deviations in grade between the train and test composites within any given stope should be considered. For critical stopes such as stopes 39, 40, and 158 the

NMR shows up to a 7% improvement relative to a conventional SGS workflow. This improvement is modest, but deemed significant in a local, high-grade context. Comparison of composite averages to stope volume expected values is assumed to be an adequate proxy to production reconciliation data; this “test” dataset scenario is likely as close as we can get without having true as-mined shapes and production grade information. It is also important to note that the stope optimization workflow is not generating a true optimal shape. The stope shapes are a geotechnically feasible, heuristic proxy to mimic the mining selectivity.

Chapter 7

Conclusions and Future Work

The connectivity of extreme values is of great practical importance. In most real-world scenarios, transfer functions are sensitive to extreme values, and correctly characterizing their two- and multi-point spatial connectivity can have significant downstream impacts. A small proportion of extreme gold grades may be the economic foundation of a mine, while extremely low permeability facies may preclude the production of a hydrocarbon reservoir. The geologic processes responsible for generating these features are not completely disorganized; strings of connected extreme values exist in nature are likely the most consequential spatial features. Traditional estimation and simulation algorithms are maximum entropy models based on multivariate Gaussian assumptions. Though practical from a dimensionality perspective, using Gaussian RFs smooths high grades and tends toward disconnected extreme values. These reasons motivate the development of the NMR framework. A mixture of Gaussian RF components may result in a highly non-Gaussian RV. The NMR exploits this idea while considering high-order connectivity measures from strings data and the potential spatial asymmetry between extreme highs and lows.

7.1 Summary of Contributions

The main contribution of this thesis is the development of the NMR framework for simulating continuous non-Gaussian spatial fields. The framework includes important contributions to quantifying non-Gaussianity from spatial data and multiple imputation for latent Gaussian factors.

7.1.1 Extreme Value Connectivity

The connectivity of extreme values is an important metric for understanding the potential shortcomings of multivariate Gaussian models. Extracting multi-point connectivity measures from drill strings is not a new idea (Boisvert, Pyrcz, et al., 2007; Ortiz, 2003); however, using the n -point connectivity function and distribution of runs to calculate a proxy for non-Gaussianity is a contribution of this thesis. This non-Gaussianity metric can provide valuable insight into spatial regions poorly modeled by a Gaussian RF. These regions could be sub-domained or simply warrant further investigation.

Extreme value connectivity is a defining feature of the NMR framework that sets it apart from traditional multivariate Gaussian simulation algorithms. Introducing high-order statistics to the mixing model overcomes the maximum entropy characteristics of a Gaussian RF. It is shown that generating NMR realizations with explicit consideration of extreme value connectivity results in

spatial connectivity features that a traditional algorithm cannot reproduce. The n -point connectivity function and distribution of runs provide straightforward access to high-order statistics without using a training image.

7.1.2 Network Model of Regionalization

The NMR is a material contribution to this thesis. It expands the LMR concept for univariate spatial modeling. Both models construct a random function as a combination of independent, standard normal factors, each with its basic covariance model operating at different spatial scales. The key difference with the NMR is that the combination of factors is not limited to be linear. Activation of the latent factors with a novel MPL function prior to the linear combination permits the spatial structure of certain features to be imparted on the distribution’s tails. Traditionally, the LMR is limited to four or five factors, while there is no practical limit with the NMR. This flexibility allows the creative mixing of latent covariance structures, achieving complex modeling goals. Structures with various anisotropies and orientations can be combined to reproduce non-stationary features. A key concept of the NMR is that a mixture of Gaussian factors can be highly non-Gaussian.

The NMR is parameterized by $2 \cdot (M + 1)$ parameters: $M + 1$ weights and $M + 1$ power law exponents. Inference of these parameters is an inverse problem. The spatial features of the observed data are known. However, the parameters that map the latent space to the observed space are unknown—framing this as an optimization problem permits using a multi-component objective function. Through this objective function, the NMR gains the ability to reproduce statistics beyond the second order. Two- and multi-point statistics are introduced directly through the optimization process. The network parameters are optimized such that the NMR output has the correct continuous and indicator variogram models, n -point connectivity function, and distributions of runs. The NMR approximates the non-linear mapping function between the latent factor space and the observed data space. The approach is highly flexible, with no limitations on the number of latent factors or objective function components. The NMR is an important contribution to geostatistics, particularly for continuous simulation of high-order connectivity features.

7.1.3 Sequential Gaussian Rejection Imputation

SGRI is a novel algorithm for multiple imputation of latent factors within the NMR framework and is a practical alternative for the Gibbs sampler. The algorithm is designed to overcome Gibbs sampler convergence issues with spatially correlated data (Silva, 2018). SGRI combines elements of SGS and rejection sampling to impute spatially correlated latent variables that reproduce data observations when mixed through the NMR. The algorithm directly samples all univariate conditional distributions of the M -dimensional latent distribution until the mapped value is within a specified tolerance. Sampling the conditional distributions ensures variogram reproduction. Next,

each sampled value is iteratively perturbed until the mapped value matches the observed value. The algorithm rejects samples anytime they do not improve the solution. Iterative refinement of the solution ensures the collocated multivariate relationships are correct while still honouring the spatial covariance structure of each factor.

When constrained and unconstrained, the algorithm shows stable convergence, correctly reproducing factor variogram models and collocated multivariate relationships. SGRI is a notable contribution concerning latent imputation. One could adapt it (with an appropriate mapping function) to truncated Gaussian categorical modeling techniques such as hierarchical truncated pluri-Gaussian simulation.

7.1.4 Simulation of Continuous High-Order Features

The complete NMR framework for the continuous simulation of high-order spatial features consists of the NMR forward model and the SGRI algorithm. This framework is the primary contribution of this thesis. The NMR methodology is developed to overcome the maximum entropy characteristic of multivariate Gaussian RFs. The NMR permits the simulation of continuous variables with connectivity features beyond the second order. These multi-point spatial connectivity features are critical when characterizing the continuity of extreme values. A key differentiator from existing high-order simulation methodologies (Mustapha & Dimitrakopoulos, 2011) is that the NMR framework does not require a training image. All high-order statistics are extracted from strings of drillhole data and embedded in the parameterization of the latent-observed mapping function. The order of these statistics is practically limited by drillhole length; however, conceptually, the n -point connectivity function can consider any number of steps n . Another key component is that the NMR framework leverages the simplicity of the Gaussian RF model for generating non-Gaussian spatial structures. This spatial Gaussian mixture model concept simplifies the implementation of the framework as both unconditional and conditional Gaussian realization can be generated with any algorithm.

Imputation of latent factors proceeds after defining all parameters of the NMR. Considering the mapping function during imputation ensures that the latent factors used for conditioning have the correct spatial features. Finally, mapping the gridded latent realizations to observed space with the NMR ensures the models reproduce all target two- and multi-point statistics. Practical considerations for inference of NMR parameters are discussed in detail, including (1) latent factor design, (2) MPL activation function parameterization, (3) objective function design, (4) potential non-uniqueness of the solution, and (5) checking of forward model outputs. A detailed discussion of the practical aspects of latent imputation is also provided, including the minimum acceptance criteria for imputed realizations. Examples show the NMR framework can effectively characterize multi-point high-grade connectivity features in a range of two- and three-dimensional scenarios, with the ability to adapt to highly non-stationary domains. The framework is demonstrated on a

dataset from a producing underground mine, where the non-Gaussian NMR realizations show up to a 7% improvement in contained metal over a conventional SGS workflow in high-grade stopes. Though a modest improvement, improving forecasts of local high-grade resources is of great practical importance.

7.1.5 Spatial Outlier Detection

An algorithm is presented to address the spatial component of outlier detection. Traditionally, the mining industry has employed graphical outlier detection techniques that neglect the spatial neighbourhood of potential outliers. The spatial correlation of geoscience data is pertinent information for describing a sample's degree of outlierness. The algorithm considers the relationship of each sample within a local neighbourhood plus its probability density from a fitted GMM. The idea is that true outliers should be either sufficiently different from their spatial neighbours, be from a low density region of the GMM, or a combination of both. The algorithm also considers the data configuration through an area of influence. Extremely high- or low-grade samples in sparse data regions pose a risk of overestimation, and the outlier score considers the area of influence. A drawback of the approach, and others from Section 2.2.1, is that a subjective threshold must be selected to delineate inliers and outliers. Selecting this threshold may be challenging in some instances, though practice shows that clustering of outliers is generally apparent. The algorithm is shown to effectively identify both outliers and extreme values in one- and two-dimensional examples. The algorithm is not intended to replace existing methodologies but to be an additional tool in the practitioners' toolbox. Best practice dictates the identification of outliers with an ensemble of methodologies.

7.2 Limitations

Despite the developments made in this thesis, limitations exist with the proposed methodologies. The following sections discuss the limitations of the NMR framework and latent variable imputation.

7.2.1 Network Model of Regionalization

A limitation of the NMR approach is that the final models reproduce linear features. The model cannot adapt to non-linear features as two-point and linear sequences of MPS characterize the objective function. Limiting MPS to 1-D sequences is a practical trade-off for the use of a TI. However, this precludes the generation of curvilinear features. One could conceivably generate curvilinear features with a sufficient number of latent factors at appropriate orientations, however the objective function components remain linear. This linear nature is a limitation compared to other high-order simulation algorithms. Another challenge is that even though a mixture of Gaussian RFs can be non-Gaussian, the mixture may experience destructured extreme values depending on the

conditioning data. Structuring the extreme values can be controlled through the objective function components. However, this may not be possible with sparse data.

Another limitation of the NMR methodology is the necessity of numerical inversion. The outputs of the forward model are known; however, the mapping parameters that yield the correct forward outputs are unknown. Inverse problems are generally ill-posed or lack a unique solution. This non-uniqueness means the solution is sensitive to initial parameter choices like the latent Gaussian pool, ω , and \mathbf{a} parameters. The potential exists for the algorithm to converge on a numerically reasonable but geologically unreasonable solution. The practitioner must make initial parameter choices to ensure the optimization process explores an appropriate solution space. The potential exists for conflicting objective function components or components with confounding relationships. There may be uncertainty in the objective components if they are derived from experimental statistics. The practitioner must ensure a carefully crafted objective function, suitable conversion criteria, and appropriate regularization through prior geologic knowledge and parameter bounds.

Numerical inversion is an iterative process and is typically computationally expensive. The speed of NMR convergence negatively correlates with the number of available data. As each inversion iteration generates a newly parameterized network, experimental variograms must be recalculated each time. Though data paring is only performed once, updating each variogram lag is computationally expensive. This expense is somewhat counteracted by parallelization of DE, yet algorithm run times may be significant with greater than 10000 data.

7.2.2 Latent Imputation

Though the SGRI algorithm shows stable convergence in many scenarios, a limitation of the methodology is the somewhat “brute-force” approach to imputation. Sequentially sampling the conditional distributions can lead to scenarios where imputed values at one data location yield local conditional CDFs at another location that does not permit convergence. This non-convergence is practically overcome by resetting all imputed values in a neighbourhood about the location and resimulating. However, the algorithm relies on sufficient iterations to overcome this issue. For these reasons, the rejection sampling approach may necessitate many iterations for convergence at each data location. Like the NMR workflow, conversion may be slow with many data.

SGRI is a practical alternative to the Gibbs sampler paradigm for imputation. However, the implementation is not general and requires a fully parameterized NMR to impute latent variables. Overall, this is not a limitation of the methodology but a consideration for adaptation to other workflows featuring latent imputation.

7.3 Future Work

Future work should address the known limitations of the proposed methodology. The NMR framework poses challenges as the solution space is non-unique and necessitates a numerical approach. Another algorithm for parameter inference may perform better than DE. DE is chosen due to its ease of implementation and widespread use in engineering problems, though another heuristic algorithm may perform well. Regardless of the chosen optimization algorithm, future work could improve the computational efficiency of both the NMR and SGRI algorithms. An efficient approximation of the experimental variogram could drastically speed up NMR parameter inference. As the NMR objective function may contain any number of components, one could explore additional components such as directional asymmetry or third-order spatial moments from $[\mathbf{h}_1, \mathbf{h}_2]$ triplets (Lauzon & Marcotte, 2020b).

Another area of future research is exploring more complex network structures. The current network architecture is simple, and there is no interaction of latent factors before the output layer. A more complex or classical multi-layer perceptron architecture could capture even more complex, non-Gaussian, or non-linear spatial features.

References

- Aggarwal, C. C. (2016). *Outlier Analysis*. Springer International Publishing.
- Ahmad, M. F., Isa, N. A. M., Lim, W. H., & Ang, K. M. (2022, May). Differential evolution: A recent review based on state-of-the-art works. *Alexandria Engineering Journal*, *61*(5), 3831–3872. doi: 10.1016/j.aej.2021.09.013
- Armstrong, M., Galli, A., Beucher, H., Loc'h, G., Renard, D., Doligez, B., ... Geffroy, F. (2011). *Plurigaussian Simulations in Geosciences*. Berlin, Heidelberg: Springer. doi: 10.1007/978-3-642-19607-2
- Arroyo, D., & Emery, X. (2020, October). Iterative algorithms for non-conditional and conditional simulation of Gaussian random vectors. *Stochastic Environmental Research and Risk Assessment*, *34*(10), 1523–1541. doi: 10.1007/s00477-020-01875-0
- Artemis Gold Inc. (2020). *Blackwater Gold Project British Columbia - NI 43-101 Technical Report on Pre-Feasibility Study* (Tech. Rep.).
- Athens, N., & Caers, J. (2022, February). Stochastic Inversion of Gravity Data Accounting for Structural Uncertainty. *Mathematical Geosciences*, *54*(2), 413–436. doi: 10.1007/s11004-021-09978-2
- Babakhani, M. (2014). *Geostatistical Modeling in Presence of Extreme Values* (Doctoral dissertation). doi: 10.7939/R3VQ2SJ85
- Balkaya, Ç., Ekinçi, Y. L., Göktürkler, G., & Turan, S. (2017, January). 3D non-linear inversion of magnetic anomalies caused by prismatic bodies using differential evolution algorithm. *Journal of Applied Geophysics*, *136*, 372–386. doi: 10.1016/j.jappgeo.2016.10.040
- Bárdossy, A., & Hörning, S. (2016, September). Random Mixing: An Approach to Inverse Modeling for Groundwater Flow and Transport Problems. *Transport in Porous Media*, *114*(2), 241–259. doi: 10.1007/s11242-015-0608-4
- Barnett, R. M., & Deutsch, C. V. (2015, October). Multivariate Imputation of Unequally Sampled Geological Variables. *Mathematical Geosciences*, *47*(7), 791–817. doi: 10.1007/s11004-014-9580-8
- Barnett, R. M., Manchuk, J. G., & Deutsch, C. V. (2014, April). Projection Pursuit Multivariate Transform. *Mathematical Geosciences*, *46*(3), 337–359. doi: 10.1007/s11004-013-9497-7
- Barnett, V., & Lewis, T. (1984). *Outliers in statistical data*.
- Bellman, R. E. (1961). *Adaptive control processes: A guided tour*. Princeton: Princeton University Press. doi: doi:10.1515/9781400874668
- Bilal, Pant, M., Zaheer, H., Garcia-Hernandez, L., & Abraham, A. (2020, April). Differential Evolution: A review of more than two decades of research. *Engineering Applications of Artificial Intelligence*, *90*, 103479. doi: 10.1016/j.engappai.2020.103479

- Boisvert, J. B., Pyrcz, M. J., & Deutsch, C. V. (2007, December). Multiple-Point Statistics for Training Image Selection. *Natural Resources Research*, 16(4), 313–321. doi: 10.1007/s11053-008-9058-9
- Boukerche, A., Zheng, L., & Alfandi, O. (2021, May). Outlier Detection: Methods, Models, and Classification. *ACM Computing Surveys*, 53(3), 1–37. doi: 10.1145/3381028
- Brown, B. M., & Resnick, S. I. (1977). Extreme values of independent stochastic processes. *Journal of Applied Probability*, 732–739.
- Caers, J., Beirlant, J., & Maes, M. A. (1999a). Statistics for modeling heavy tailed distributions in geology: Part I. Methodology. *Mathematical geology*, 31(4), 391–410.
- Caers, J., Beirlant, J., & Maes, M. A. (1999b). Statistics for modeling heavy tailed distributions in geology: Part II. Applications. *Mathematical geology*, 31(4), 411–434.
- Carbas, S., Toktas, A., & Ustun, D. (2021). *Nature-inspired metaheuristic algorithms for engineering optimization applications*. Springer.
- Cardinal Resources. (2019). *Namdini Gold Project Feasibility Study NI 43-101 Technical Report, Ghana, West Africa* (Tech. Rep.).
- Carvalho, D., & Deutsch, C. V. (2017, January). *An Overview of Multiple Indicator Kriging*. <https://geostatisticslessons.com/lessons/mikoverview>.
- Chen, D., Lu, C.-T., Kou, Y., & Chen, F. (2008). On detecting spatial outliers. *Geoinformatica*, 12(4), 455–475.
- Chilès, J.-P., & Delfiner, P. (2012). *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons.
- CIM Mineral Resource & Mineral Reserve Committee. (2019, November). *CIM Estimation of Mineral Resources & Mineral Reserves Best Practice Guidelines* (Tech. Rep.). Canadian Institute of Mining, Metallurgy and Petroleum.
- Coles, S., Bawa, J., Trenner, L., & Dorazio, P. (2001). *An introduction to statistical modeling of extreme values* (Vol. 208). Springer.
- Conn, A. R., Scheinberg, K., & Vicente, L. N. (2009). *Introduction to derivative-free optimization*. SIAM.
- Costa, J. F. (2003). Reducing the impact of outliers in ore reserves estimation. *Mathematical geology*, 35(3), 323–345.
- Cui, E. H., Zhang, Z., Chen, C. J., & Wong, W. K. (2024, April). Applications of nature-inspired metaheuristic algorithms for tackling optimization problems across disciplines. *Scientific Reports*, 14(1), 9403. doi: 10.1038/s41598-024-56670-6
- da Silva, C. Z., & Costa, J. F. C. L. (2019, January). The treatment of missing ‘not at random’ geological data for ore grade modelling. *Applied Earth Science*, 128(1), 15–26. doi: 10.1080/25726838.2018.1547508
- Dávila Rodríguez, I. A., Palafox González, A., Guerrero Arroyo, E. A., Becerra López, F. I., &

- Fregoso Becerra, E. (2024, April). Three-Dimensional Inversion of Magnetic Anomalies Using a Low-Level Representation and an Evolution Strategy for Archaeological Studies. *Mathematical Geosciences*, *56*(3), 511–539. doi: 10.1007/s11004-023-10090-w
- Davis, M. W. (1987, February). Production of conditional simulations via the LU triangular decomposition of the covariance matrix. *Mathematical Geology*, *19*(2), 91–98. doi: 10.1007/BF00898189
- Davison, A. C., & Huser, R. (2015). Statistics of extremes. *Annual Review of Statistics and its Application*, *2*, 203–235.
- Davison, A. C., Huser, R., & Thibaud, E. (2013). Geostatistics of dependent and asymptotically independent extremes. *Mathematical Geosciences*, *45*(5), 511–529.
- de Carvalho, J. P., Dimitrakopoulos, R., & Minniakhmetov, I. (2019, August). High-Order Block Support Spatial Simulation Method and Its Application at a Gold Deposit. *Mathematical Geosciences*, *51*(6), 793–810. doi: 10.1007/s11004-019-09784-x
- De Haan, L., & Ferreira, A. (2007). *Extreme value theory: An introduction*. Springer Science & Business Media.
- Deligne, N. I., Coles, S. G., & Sparks, R. S. J. (2010). Recurrence rates of large explosive volcanic eruptions. *Journal of Geophysical Research: Solid Earth*, *115*(B6). doi: 10.1029/2009JB006554
- Deutsch, C. V. (1992). *Annealing techniques applied to reservoir modeling and the integration of geological and engineering (well test) data* (Unpublished doctoral dissertation). stanford university.
- Deutsch, C. V. (2007). A Recall of Factorial Kriging with Examples and a Modified Version of kt3d, 8.
- Deutsch, C. V. (2010). Display of cross validation/jackknife results. *Centre for Computational Geostatistics Annual Report*, *12*(406), 1–4.
- Deutsch, C. V., & Journel, A. G. (1992). Geostatistical software library and user’s guide. *New York*, *119*(147).
- Dimitrakopoulos, R., Mustapha, H., & Gloaguen, E. (2009, December). High-order Statistics of Spatial Random Fields: Exploring Spatial Cumulants for Modeling Complex Non-Gaussian and Non-linear Phenomena. *Mathematical Geosciences*, *42*(1), 65. doi: 10.1007/s11004-009-9258-9
- Drumond, D. A., Rolo, R. M., & Costa, J. F. C. L. (2019, January). Using Mahalanobis Distance to Detect and Remove Outliers in Experimental Covariograms. *Natural Resources Research*, *28*(1), 145–152. doi: 10.1007/s11053-018-9399-y
- Dubey, S. R., Singh, S. K., & Chaudhuri, B. B. (2022, September). Activation functions in deep learning: A comprehensive survey and benchmark. *Neurocomputing*, *503*, 92–108. doi: 10.1016/j.neucom.2022.06.111

- Dutaut, R. V., & Marcotte, D. (2021, May). A new grade-capping approach based on coarse duplicate data correlation. *Journal of the Southern African Institute of Mining and Metallurgy*, 121(5), 193–200. doi: 10.17159/2411-9717/1379/2021
- Eidsvik, J., Mukerji, T., & Bhattacharjya, D. (2015). *Value of information in the earth sciences: Integrating spatial modeling and decision analysis*. Cambridge University Press.
- Eldorado Gold Corporation. (2020). *Technical Report Kışladağ Gold Mine Turkey* (Tech. Rep.).
- Emery, X., Arroyo, D., & Peláez, M. (2014, April). Simulating Large Gaussian Random Vectors Subject to Inequality Constraints by Gibbs Sampling. *Mathematical Geosciences*, 46(3), 265–283. doi: 10.1007/s11004-013-9495-9
- Ernst, M., & Haesbroeck, G. (2017). Comparison of local outlier detection techniques in spatial multivariate data. *Data mining and knowledge discovery*, 31(2), 371–399.
- Everitt, B., & Skrondal, A. (2010). *The cambridge dictionary of statistics*. Cambridge University Press.
- Filzmoser, P., & Gregorich, M. (2020, November). Multivariate Outlier Detection in Applied Data Analysis: Global, Local, Compositional and Cellwise Outliers. *Mathematical Geosciences*, 52(8), 1049–1066. doi: 10.1007/s11004-020-09861-6
- Filzmoser, P., Ruiz-Gazen, A., & Thomas-Agnan, C. (2014, February). Identification of local multivariate outliers. *Statistical Papers*, 55(1), 29–47. doi: 10.1007/s00362-013-0524-z
- Fisher, A., Rudin, C., & Dominici, F. (2019, December). *All Models are Wrong, but Many are Useful: Learning a Variable’s Importance by Studying an Entire Class of Prediction Models Simultaneously* (No. arXiv:1801.01489). arXiv. doi: 10.48550/arXiv.1801.01489
- Fisher, R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical proceedings of the Cambridge philosophical society* (Vol. 24, pp. 180–190). Cambridge University Press.
- Fourie, A., Morgan, C., & Minnitt, R. (2019). Limiting the influence of extreme grades in ordinary kriged estimates. *Journal of the Southern African Institute of Mining and Metallurgy*, 119(4). doi: 10.17159/2411-9717/18/090/2019
- Fréchet, M. (1927). Sur la loi de probabilité de l’écart maximum. *Ann. Soc. Math. Polon.*, 6, 93–116.
- Fu, J. C., & Lou, W. Y. W. (2003). *Distribution Theory of Runs and Patterns and Its Applications: A Finite Markov Chain Imbedding Approach*. WORLD SCIENTIFIC. doi: 10.1142/4669
- Gaume, J., Eckert, N., Chambon, G., Naaim, M., & Bel, L. (2013). Mapping extreme snowfalls in the French Alps using max-stable processes. *Water Resources Research*, 49(2), 1079–1098.
- Geman, S., & Geman, D. (1984, November). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6), 721–741. doi: 10.1109/TPAMI.1984.4767596
- Georgioudakis, M., & Plevris, V. (2020). A Comparative Study of Differential Evolution Variants

- in Constrained Structural Optimization. *Frontiers in Built Environment*, 6.
- Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems.* ” O’Reilly Media, Inc.”.
- Ghorbanidehno, H., Kokkinaki, A., Lee, J., & Darve, E. (2020, December). Recent developments in fast and scalable inverse modeling and data assimilation methods in hydrology. *Journal of Hydrology*, 591, 125266. doi: 10.1016/j.jhydrol.2020.125266
- Gilli, M., & Kellezi, E. (2006). An application of extreme value theory for measuring financial risk. *Computational Economics*, 27(2), 207–228.
- Giraud, J., Lindsay, M., Ogarko, V., Jessell, M., Martin, R., & Pakyuz-Charrier, E. (2019, January). Integration of geoscientific uncertainty into geophysical inversion by means of local gradient regularization. *Solid Earth*, 10(1), 193–210. doi: 10.5194/se-10-193-2019
- Gnedenko, B. (1943). Sur La Distribution Limite Du Terme Maximum D’Une Serie Aleatoire. *Annals of Mathematics*, 44(3), 423–453. doi: 10.2307/1968974
- Gómez-Hernández, J. J., & Journel, A. G. (1993). Joint Sequential Simulation of MultiGaussian Fields. In A. Soares (Ed.), *Geostatistics Tróia ’92: Volume 1* (pp. 85–94). Dordrecht: Springer Netherlands. doi: 10.1007/978-94-011-1739-5_8
- Gómez-Hernández, J. J., & Srivastava, R. M. (2021, February). One Step at a Time: The Origins of Sequential Simulation and Beyond. *Mathematical Geosciences*, 53(2), 193–209. doi: 10.1007/s11004-021-09926-0
- Gómez-Hernández, J. J., & Wen, X.-H. (1998, February). To be or not to be multi-Gaussian? A reflection on stochastic hydrogeology. *Advances in Water Resources*, 21(1), 47–61. doi: 10.1016/S0309-1708(96)00031-0
- Goovaerts, P. (1992). Factorial kriging analysis: A useful tool for exploring the structure of multivariate spatial soil information. *Journal of Soil Science*, 43(4), 597–619. doi: 10.1111/j.1365-2389.1992.tb00163.x
- Goovaerts, P. (1997). *Geostatistics for natural resources evaluation.* Oxford University Press on Demand.
- Grana, D., Azevedo, L., de Figueiredo, L., Connolly, P., & Mukerji, T. (2022, July). Probabilistic inversion of seismic data for reservoir petrophysical characterization: Review and examples. *Geophysics*, 87(5), M199-M216. doi: 10.1190/geo2021-0776.1
- Guardiano, F. B., & Srivastava, R. M. (1993). Multivariate Geostatistics: Beyond Bivariate Moments. In F. M. Gradstein & A. Soares (Eds.), *Geostatistics Tróia ’92* (Vol. 5, pp. 133–144). Dordrecht: Springer Netherlands. doi: 10.1007/978-94-011-1739-5_12
- Guthke, P. (2013). *Non-multi-Gaussian spatial structures: Process-driven natural genesis, manifestation, modeling approaches, and influences on dependent processes* (Unpublished doctoral dissertation).
- Guthke, P., & Bárdossy, A. (2017, May). On the link between natural emergence and manifestation

- of a fundamental non-Gaussian geostatistical property: Asymmetry. *Spatial Statistics*, 20, 1–29. doi: 10.1016/j.spasta.2017.01.003
- Hadavand, M., & Deutsch, C. V. (2023, August). Spatial multivariate data imputation using deep learning and lambda distribution. *Computers & Geosciences*, 177, 105376. doi: 10.1016/j.cageo.2023.105376
- Harding, B., Lagos, R., Pfeiffer, N., & Deutsch, C. V. (2023, October). Probabilistic Modeling of the Round Mountain Gold Deposit: A Case Study. *Mining, Metallurgy & Exploration*, 40(5), 1987–2006. doi: 10.1007/s42461-023-00787-1
- Harris, P., Brunson, C., Charlton, M., Juggins, S., & Clarke, A. (2014). Multivariate spatial outlier detection using robust geographically weighted methods. *Mathematical Geosciences*, 46(1), 1–31.
- Hawkins, D. M., & Cressie, N. (1984, January). Robust kriging—A proposal. *Journal of the International Association for Mathematical Geology*, 16(1), 3–18. doi: 10.1007/BF01036237
- Hodge, V., & Austin, J. (2004). A survey of outlier detection methodologies. *Artificial intelligence review*, 22(2), 85–126.
- Holland, J. H. (1992). Genetic Algorithms. *Scientific American*, 267(1), 66–73.
- Hong, S., & Deutsch, C. V. (2007). Improved Factorial Kriging for Feature Identification and Extraction. , 16.
- Isaaks, E. H. (1990). *The application of Monte Carlo methods to the analysis of spatially correlated data* (Unpublished doctoral dissertation). Stanford University, United States – California.
- Journal, A. G. (1974, August). Geostatistics for Conditional Simulation of Ore Bodies. *Economic Geology*, 69(5), 673–687. doi: 10.2113/gsecongeo.69.5.673
- Journal, A. G. (1983, June). Nonparametric estimation of spatial distributions. *Journal of the International Association for Mathematical Geology*, 15(3), 445–468. doi: 10.1007/BF01031292
- Journal, A. G. (2005). Beyond Covariance: The Advent of Multiple-Point Geostatistics. In O. Leuangthong & C. V. Deutsch (Eds.), *Geostatistics Banff 2004* (pp. 225–233). Dordrecht: Springer Netherlands. doi: 10.1007/978-1-4020-3610-1_23
- Journal, A. G., & Alabert, F. (1989). Non-Gaussian data expansion in the Earth Sciences. *Terra Nova*, 1(2), 123–134. doi: 10.1111/j.1365-3121.1989.tb00344.x
- Journal, A. G., & Deutsch, C. V. (1993, April). Entropy and spatial disorder. *Mathematical Geology*, 25(3), 329–355. doi: 10.1007/BF00901422
- Kabluchko, Z., Schlather, M., De Haan, L., et al. (2009). Stationary max-stable fields associated to negative definite functions. *The Annals of Probability*, 37(5), 2042–2065.
- Kerrou, J., Renard, P., Hendricks Franssen, H.-J., & Lunati, I. (2008, January). Issues in characterizing heterogeneity and connectivity in non-multiGaussian media. *Advances in Water Resources*, 31(1), 147–159. doi: 10.1016/j.advwatres.2007.07.002
- Lantuéjoul, C., & Desassis, N. (2012). Simulation of a Gaussian random vector: A propagative

- version of the Gibbs sampler. In *The 9th international geostatistics congress* (pp. 174–181).
- Lauzon, D., & Marcotte, D. (2020a, February). Calibration of random fields by a sequential spectral turning bands method. *Computers & Geosciences*, *135*, 104390. doi: 10.1016/j.cageo.2019.104390
- Lauzon, D., & Marcotte, D. (2020b, November). The sequential spectral turning band simulator as an alternative to Gibbs sampler in large truncated- or pluri- Gaussian simulations. *Stochastic Environmental Research and Risk Assessment*, *34*(11), 1939–1951. doi: 10.1007/s00477-020-01850-9
- Lauzon, D., & Marcotte, D. (2023, September). Joint hydrofacies-hydraulic conductivity modeling based on a constructive spectral algorithm constrained by transient head data. *Hydrogeology Journal*, *31*(6), 1647–1664. doi: 10.1007/s10040-023-02638-1
- Lee, J.-H., Kim, H., Park, H.-J., & Heo, J.-H. (2021, January). Temporal prediction modeling for rainfall-induced shallow landslide hazards using extreme value distribution. *Landslides*, *18*(1), 321–338. doi: 10.1007/s10346-020-01502-7
- Leuangthong, O., Khan, K. D., & Deutsch, C. V. (2008). *Solved problems in geostatistics*. John Wiley & Sons.
- Leuangthong, O., McLennan, J. A., & Deutsch, C. V. (2004, September). Minimum Acceptance Criteria for Geostatistical Realizations. *Natural Resources Research*, *13*(3), 131–141. doi: 10.1023/B:NARR.0000046916.91703.bb
- Leuangthong, O., & Nowak, M. (2015). Dealing with high-grade data in resource estimation. *Journal of the Southern African Institute of Mining and Metallurgy*, *115*(1), 27–36.
- Leung, R., Balamurali, M., & Melkumyan, A. (2021, January). Sample Truncation Strategies for Outlier Removal in Geochemical Data: The MCD Robust Distance Approach Versus t-SNE Ensemble Clustering. *Mathematical Geosciences*, *53*(1), 105–130. doi: 10.1007/s11004-019-09839-z
- Li, Z., Zhao, Y., Hu, X., Botta, N., Ionescu, C., & Chen, G. H. (2022, March). ECOD: Unsupervised Outlier Detection Using Empirical Cumulative Distribution Functions. *arXiv:2201.00382 [cs, stat]*.
- Linde, N., Renard, P., Mukerji, T., & Caers, J. (2015, December). Geological realism in hydrogeological and geophysical inverse modeling: A review. *Advances in Water Resources*, *86*, 86–101. doi: 10.1016/j.advwatres.2015.09.019
- Little, R. J. A., & Rubin, D. B. (2019). *Statistical analysis with missing data* (3rd ed.). Wiley.
- Madani, N., & Bazarbekov, T. (2021, March). Enhanced conditional Co-Gibbs sampling algorithm for data imputation. *Computers & Geosciences*, *148*, 104655. doi: 10.1016/j.cageo.2020.104655
- Mahalanobis, P. C. (2018). On the Generalized Distance in Statistics. *Sankhyā: The Indian Journal of Statistics, Series A (2008-)*, *80*, S1-S7.

- Maleki, M., Madani, N., & Emery, X. (2014). Capping and kriging grades with long-tailed distributions. *Journal of the Southern African Institute of Mining and Metallurgy*, 114(3), 255–263.
- Mariethoz, G., & Caers, J. (2014). *Multiple-Point Geostatistics* (1st ed.). John Wiley & Sons, Ltd. doi: 10.1002/9781118662953
- Matheron, G. (1963). Principles of geostatistics. *Economic geology*, 58(8), 1246–1266.
- Matheron, G. (1982). *Pour une analyse krigeante des données régionalisées* (Tech. Rep. No. N-732). Ecole des Mines de Paris.
- Matheron, G. (2019). *Matheron's Theory of Regionalised Variables* (V. Pawlowsky-Glahn & J. Serra, Eds.). Oxford: Oxford University Press. doi: 10.1093/oso/9780198835660.001.0001
- Matheron, G., Beucher, H., de Fouquet, C., Galli, A., Guérillot, D., & Ravanne, C. (1987). Conditional simulation of the geometry of fluvio-deltaic reservoirs. In *SPE annual technical conference and exhibition?* (pp. SPE-16753). Spe.
- McLachlan, G. J., Lee, S. X., & Rathnayake, S. I. (2019, March). Finite Mixture Models. *Annual Review of Statistics and Its Application*, 6(Volume 6, 2019), 355–378. doi: 10.1146/annurev-statistics-031017-100325
- Medgold Resources Corp. (2021). *PRELIMINARY ECONOMIC ASSESSMENT AND NI 43-101 TECHNICAL REPORT FOR THE MEDGOLD TLAMINO PROJECT LICENCES, SERBIA* (Tech. Rep.).
- Meng, Z., Chen, Y., & Li, X. (2020). Enhancing Differential Evolution With Novel Parameter Control. *IEEE Access*, 8, 51145–51167. doi: 10.1109/ACCESS.2020.2979738
- Miniussi, A., Marani, M., & Villarini, G. (2020, February). Metastatistical Extreme Value Distribution applied to floods across the continental United States. *Advances in Water Resources*, 136, 103498. doi: 10.1016/j.advwatres.2019.103498
- Minniakhmetov, I., & Dimitrakopoulos, R. (2017, January). Joint High-Order Simulation of Spatially Correlated Variables Using High-Order Spatial Statistics. *Mathematical Geosciences*, 49(1), 39–66. doi: 10.1007/s11004-016-9662-x
- Minniakhmetov, I., & Dimitrakopoulos, R. (2022, January). High-Order Data-Driven Spatial Simulation of Categorical Variables. *Mathematical Geosciences*, 54(1), 23–45. doi: 10.1007/s11004-021-09943-z
- Minniakhmetov, I., Dimitrakopoulos, R., & Godoy, M. (2018, October). High-Order Spatial Simulation Using Legendre-Like Orthogonal Splines. *Mathematical Geosciences*, 50(7), 753–780. doi: 10.1007/s11004-018-9741-2
- Mirjalili, S. (2019). *Evolutionary Algorithms and Neural Networks* (Vol. 780). Cham: Springer International Publishing. doi: 10.1007/978-3-319-93025-1
- Mohamed, A. W. (2014, November). RDEL: Restart Differential Evolution algorithm with Local Search Mutation for global numerical optimization. *Egyptian Informatics Journal*, 15(3), 175–188. doi: 10.1016/j.eij.2014.07.001

- Mustapha, H., & Dimitrakopoulos, R. (2010, July). High-order Stochastic Simulation of Complex Spatially Distributed Natural Phenomena. *Mathematical Geosciences*, *42*(5), 457–485. doi: 10.1007/s11004-010-9291-8
- Mustapha, H., & Dimitrakopoulos, R. (2011, September). HOSIM: A high-order stochastic simulation algorithm for generating three-dimensional complex geological patterns. *Computers & Geosciences*, *37*(9), 1242–1253. doi: 10.1016/j.cageo.2010.09.007
- Nava-Flores, M., Ortiz-Alemán, C., & Urrutia-Fucugauchi, J. (2023, March). High Resolution Model of the Vinton Salt-Dome Cap Rock by Joint Inversion of the Full Tensor Gravity Gradient Data with the Simulated Annealing Global Optimization Method. *Pure and Applied Geophysics*, *180*(3), 983–1014. doi: 10.1007/s00024-023-03227-9
- Nevada Gold Mines LLC. (2020). *TECHNICAL REPORT ON THE CARLIN COMPLEX, EU-REKA AND ELKO COUNTIES, STATE OF NEVADA, USA* (Tech. Rep.).
- Neves, M. M. (2015). Geostatistical Analysis in Extremes: An Overview. *Mathematics of Energy and Climate Change*, 229–245.
- Nguyen, M., Veraart, A. E. D., Taisne, B., Tan, C. T., & Lallemand, D. (2023, October). A Dynamic Extreme Value Model with Application to Volcanic Eruption Forecasting. *Mathematical Geosciences*. doi: 10.1007/s11004-023-10109-2
- Nowak, M., & Leuangthong, O. (2019). Optimal drill hole spacing for resource classification. In *Mining Goes Digital*. CRC Press.
- Nowak, M., Leuangthong, O., & Srivastava, R. M. (2013). Suggestions for good capping practices from historical literature. In *Proceedings of the 23rd World Mining Congress 2013*. Canadian Institute of Mining, Metallurgy and Petroleum Montreal.
- Nowak, R. (2008, September). Generalized binary search. In *2008 46th Annual Allerton Conference on Communication, Control, and Computing* (pp. 568–574). doi: 10.1109/ALLERTON.2008.4797609
- Opitz, T. (2013). Extremal t processes: Elliptical domain of attraction and a spectral representation. *Journal of Multivariate Analysis*, *122*, 409–413.
- Ortiz, J. (2003). *Characterization of high order correlation for enhanced indicator simulation*. (Unpublished doctoral dissertation). University of Alberta.
- Ortiz, J., & Deutsch, C. V. (2002, February). Calculation of Uncertainty in the Variogram. *Mathematical Geology*, *34*(2), 169–183. doi: 10.1023/A:1014412218427
- Osisko Gold Royalties Ltd. (2020). *NI 43-101 Technical Report and Mineral Resource Estimate for the Cariboo Gold Project, British Columbia, Canada* (Tech. Rep.).
- Pang, G., Shen, C., Cao, L., & van den Hengel, A. (2022, March). Deep Learning for Anomaly Detection: A Review. *ACM Computing Surveys*, *54*(2), 1–38. doi: 10.1145/3439950
- Pardo-Igúzquiza, E., & Olea, R. A. (2012, April). VARBOOT: A spatial bootstrap program for semivariogram uncertainty assessment. *Computers & Geosciences*, *41*, 188–198. doi: 10.1016/

j.cageo.2011.09.002

- Parker, HM. (1991). Statistical treatment of outlier data in epithermal gold deposit reserve estimation. *Mathematical Geology*, 23(2), 175–199.
- Parker, HM. (2006). *Technical Report of the Rock Creek Property, Nome, Alaska, USA* (Tech. Rep.).
- Parrish, I. (1997). Geologist’s gordian knot: To cut or not to cut. *Mining Engineering*, 49, 45–49.
- Pasofino Gold Ltd. (2020). *DUGBE GOLD PROJECT, LIBERIA NI 43-101 TECHNICAL REPORT* (Tech. Rep.).
- Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of statistics*, 3(1), 119–131.
- Pimentel, M. A., Clifton, D. A., Clifton, L., & Tarassenko, L. (2014). A review of novelty detection. *Signal Processing*, 99, 215–249.
- Pinto, F. C. (2020). *Independent Factor Simulation for Improved Multivariate Geostatistics* (Unpublished doctoral dissertation). University of Alberta.
- Piotrowski, A. P. (2017, February). Review of Differential Evolution population size. *Swarm and Evolutionary Computation*, 32, 1–24. doi: 10.1016/j.swevo.2016.05.003
- Pretium Resources Inc. (2020). *Technical Report on the Brucejack Gold Mine, Northwest British Columbia* (Tech. Rep.).
- Price, K. V. (2013). Differential Evolution. In I. Zelinka, V. Snášel, & A. Abraham (Eds.), *Handbook of Optimization: From Classical to Modern Approach* (pp. 187–214). Berlin, Heidelberg: Springer. doi: 10.1007/978-3-642-30504-7_8
- Pyrzcz, M. J., & Deutsch, C. V. (2014). *Geostatistical reservoir modeling*. Oxford university press.
- Qu, J., Du, Q., Li, Y., Tian, L., & Xia, H. (2021, November). Anomaly Detection in Hyperspectral Imagery Based on Gaussian Mixture Model. *IEEE Transactions on Geoscience and Remote Sensing*, 59(11), 9504–9517. doi: 10.1109/TGRS.2020.3038722
- Renard, P., Straubhaar, J., Caers, J., & Mariethoz, G. (2011, November). Conditioning Facies Simulations with Connectivity Data. *Mathematical Geosciences*, 43(8), 879–903. doi: 10.1007/s11004-011-9363-4
- Rios, L. M., & Sahinidis, N. V. (2013, July). Derivative-free optimization: A review of algorithms and comparison of software implementations. *Journal of Global Optimization*, 56(3), 1247–1293. doi: 10.1007/s10898-012-9951-y
- Rivoirard, J., Demange, C., Freulon, X., Lécureuil, A., & Bellot, N. (2013). A top-cut model for deposits with heavy-tailed grade distribution. *Mathematical geosciences*, 45(8), 967–982.
- Rojas, R. (1996). The Backpropagation Algorithm. In R. Rojas (Ed.), *Neural Networks: A Systematic Introduction* (pp. 149–182). Berlin, Heidelberg: Springer. doi: 10.1007/978-3-642-61068-4_7
- Roscoe, W. (1996, April). Cutting curves for grade estimation and grade control in gold mines. In *98th annual general meeting*. Canadian Institute of Mining, Metallurgy and Petroleum.

- Rossi, M. E., & Deutsch, C. V. (2013). *Mineral resource estimation*. Springer Science & Business Media.
- Rousseeuw, P. J., & Driessen, K. V. (1999, August). A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics*, *41*(3), 212–223. doi: 10.1080/00401706.1999.10485670
- Rukhin, A., Soto, J., Nechvatal, J., Barker, E., Leigh, S., Levenson, M., ... Dray, J. (2010). A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications. , 131.
- Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes*, *5*(1), 33–44.
- Schlather, M., & Tawn, J. A. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika*, *90*(1), 139–156.
- Sen, M. K., & Stoffa, P. L. (2013). *Global Optimization Methods in Geophysical Inversion*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511997570
- Sharma, S., Sharma, S., & Athaiya, A. (2020, May). ACTIVATION FUNCTIONS IN NEURAL NETWORKS. *International Journal of Engineering Applied Sciences and Technology*, *04*(12), 310–316. doi: 10.33564/IJEAST.2020.v04i12.054
- Silva, D. A. (2014). *Guide to MPS Simulation with SNESIM* (Vol. 18).
- Silva, D. S. F. (2018). *Enhanced Geologic Modeling of Multiple Categorical Variables* (Doctoral dissertation, University of Alberta). doi: 10.7939/R30G3HD9R
- Silva, D. S. F., & Deutsch, C. V. (2017, November). Multiple imputation framework for data assignment in truncated pluri-Gaussian simulation. *Stochastic Environmental Research and Risk Assessment*, *31*(9), 2251–2263. doi: 10.1007/s00477-016-1309-4
- Silva, D. S. F., & Deutsch, C. V. (2018, October). Multivariate data imputation using Gaussian mixture models. *Spatial Statistics*, *27*, 74–90. doi: 10.1016/j.spasta.2016.11.002
- Silva, V. M. (2021, July). On the classification and treatment of outliers in a spatial context: A Bayesian Updating approach. *REM - International Engineering Journal*, *74*, 379–389. doi: 10.1590/0370-44672021740003
- Smith, R. L. (1990). Max-stable processes and spatial extremes. *Unpublished manuscript*, *205*, 1–32.
- Solow, A. R. (1985, October). Bootstrapping correlated data. *Journal of the International Association for Mathematical Geology*, *17*(7), 769–775. doi: 10.1007/BF01031616
- Storn, R., & Price, K. (1997, December). Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. *Journal of Global Optimization*, *11*(4), 341–359. doi: 10.1023/A:1008202821328
- Strebelle, S. (2002, January). Conditional Simulation of Complex Geological Structures Using Multiple-Point Statistics. *Mathematical Geology*, *34*(1), 1–21. doi: 10.1023/A:1014009426274
- Tahmasebi, P. (2018). Multiple Point Statistics: A Review. In B. Daya Sagar, Q. Cheng, &

- F. Agterberg (Eds.), *Handbook of Mathematical Geosciences: Fifty Years of IAMG* (pp. 613–643). Cham: Springer International Publishing. doi: 10.1007/978-3-319-78999-6_30
- Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics. doi: 10.1137/1.9780898717921
- TriStar Gold Inc. (2021). *Mineral Resource Update for the Castelo de Sonhos Gold Project, Pará State, Brazil* (Tech. Rep.).
- Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2). Reading, MA.
- Ünal, H. T., & Başçiftçi, F. (2022, March). Evolutionary design of neural network architectures: A review of three decades of research. *Artificial Intelligence Review*, 55(3), 1723–1802. doi: 10.1007/s10462-021-10049-5
- Vincent, J., & Deutsch, C. (2021). MIK vs MG in Non-Gaussian Environments. , 13.
- Vincent, J. D. (2021). *Multiple-Indicator Kriging of Gaussian and Non-Gaussian Data* (Unpublished doctoral dissertation).
- von Mises, R. (1936). La distribution de la plus grande de n valeurs. *Rev. Math. Interbalcanic Union*, 1, 141–160.
- Wackernagel, H. (1988). Geostatistical Techniques for Interpreting Multivariate Spatial Information. In C. F. Chung, A. G. Fabbri, & R. Sinding-Larsen (Eds.), *Quantitative Analysis of Mineral and Energy Resources* (pp. 393–409). Dordrecht: Springer Netherlands. doi: 10.1007/978-94-009-4029-1_24
- Wang, H., Bah, M. J., & Hammad, M. (2019). Progress in outlier detection techniques: A survey. *IEEE Access*, 7, 107964–108000.
- Wilde, B. J. (2007). *Wide array declustering for representative distributions (The Ultimate DECLUS Program)* (CCG Annual Report 9). Edmonton AB: University of Alberta.
- Yan, Y., Jeong, J., & Genton, M. G. (2020). Multivariate transformed Gaussian processes. *Japanese Journal of Statistics and Data Science*, 3(1), 129–152.
- Yang, X.-S. (2018). *Nature-inspired algorithms and applied optimization* (Vol. 744). Springer.
- Yao, L., Dimitrakopoulos, R., & Gamache, M. (2020, July). High-Order Sequential Simulation via Statistical Learning in Reproducing Kernel Hilbert Space. *Mathematical Geosciences*, 52(5), 693–723. doi: 10.1007/s11004-019-09843-3
- Yao, L., Dimitrakopoulos, R., & Gamache, M. (2021, April). Learning high-order spatial statistics at multiple scales: A kernel-based stochastic simulation algorithm and its implementation. *Computers & Geosciences*, 149, 104702. doi: 10.1016/j.cageo.2021.104702
- Yao, L., Dimitrakopoulos, R., & Gamache, M. (2021, October). Training Image Free High-Order Stochastic Simulation Based on Aggregated Kernel Statistics. *Mathematical Geosciences*, 53(7), 1469–1489. doi: 10.1007/s11004-021-09923-3
- Zhou, H., Gómez-Hernández, J. J., & Li, L. (2014, January). Inverse methods in hydrogeology: Evolution and recent trends. *Advances in Water Resources*, 63, 22–37. doi:

10.1016/j.advwatres.2013.10.014

- Zimek, A., & Filzmoser, P. (2018). There and back again: Outlier detection between statistical reasoning and data mining algorithms. *WIREs Data Mining and Knowledge Discovery*, 8(6), e1280. doi: 10.1002/widm.1280

Appendix A

Software

The following Appendix presents the software developed for the methodologies in this thesis. Programs are developed as a series of FORTRAN 90 codes with geostatistical software library (GSLIB) style parameter files (Deutsch & Journel, 1992). `NMROPT` is developed for NMR parameter inference and `NMRIMP` for imputation of latent Gaussian factors. These programs together form the suite of programs for the NMR framework. The following sections present the parameter file for each program, summarizing the key parameters. All source code is available at the authors github (`NMROPT` and `NMRIMP`).

A.1 NMROPT

`NMROPT` provides functionality for the NMR parameter inference discussed in Chapter 4. The program takes two data files as inputs: (1) specifying the drillhole data and (2) specifying the covariance structure of each latent factor. Network architecture, thresholds, experimental variogram search parameters, and objective functions components are specified directly in the parameter file. The optimized network parameters are output as a text file and passed to `NMRIMP` for latent factor imputation. The parameter file is shown below:

```
1          PARAMETERS FOR NMROPT
2          *****
3
4  START OF PARAMETERS:
5  data.dat          - file with data
6  1 4 5 6 7 11     - columns for dh, x, y, z, var and wt
7  0                - normal score transform var? (0=no, 1=yes)
8  -1.0e21    1.0e21 - trimming limits
9  100              - number of unconditional realizations
10 0                - simulation type (0=LU, 1=sequential)
11 5841044          - random number seed
12 1 1              - debugging level, realization to output
13 nmropt.dbg       - file for debugging output
14 nmropt.out       - file for network output
15 nmrwts.out       - file for optimized network weights
16 nmrobj.out       - file for objective function value per ...
```

17	nmr_	- prefix for target/experimental output
18	3	- number of network layers
19	5 5 1	- network layer dimensions
20	1 0.1	- network wt regularization (0=none, ...
21	pool.dat	- file with cov. structs. of Gaussian pool
22	0	- consider factor precedence? (0=no, 1=yes)
23	1 1 1 1	- objective components: varg, ivarg, ...
24	1 1 1 1	- objective weight: varg, ivarg, ...
25	3	- number of indicator thresholds
26	-1.28 0.0 1.28	- Gaussian indicator thresholds
27	1 1 1	- threshold weights
28	1	- runs above or below threshold? ...
29	30	- max number of runs to consider
30	0	- runs target from file? (0=no, 1=yes)
31	target_runs.out	- runs target file
32	1	- npoint above or below threshold? ...
33	30	- max number of connected steps to consider
34	0	- npoint target from file? (0=no, 1=yes)
35	target_npoint.out	- npoint target file
36	0.8 0.5 1.0 15 1000	- DE parameters: F, CR lo, CR hi, pop ...
37	0.0 1.0	- DE bounds: lower, upper
38	omega.out	- file with factor omega bounds
39	1	- num. threads for parallel DE ...
40	2	- number of experimental variogram ...
41	0.0 22.5 1000 0.0 22.5 1000 0.0	- dir 01: azm,azmtol, ...
42	8 1000.0 500.0	- number of lags,lag distance,lag ...
43	90. 22.5 1000 0.0 22.5 1000 0.0	- dir 02: azm,azmtol, ...
44	8 1000.0 500.0	- number of lags,lag distance,lag ...
45	1	- number of target variogram models
46	3	- number of target indicator variogram ...
47	1.0	- IDW power for variogram optimization ...
48	999999	- max number of exp. variogram pairs ...
49	1	- standardize sill? (0=no, 1=yes)
50	1 0.1	- nst, nugget effect
51	1 0.9 0.0 0.0 0.0	- it,cc,ang1,ang2,ang3
52	10.0 10.0 10.0	- a_hmax, a_hmin, a_vert

53	1	0.1				- inst, nugget effect
54	1	0.9	0.0	0.0	0.0	- iit,icc,iang1,iang2,iang3
55	10.0	10.0	10.0			- ia_hmax, ia_hmin, ia_vert
56	1	0.1				- inst, nugget effect
57	1	0.9	0.0	0.0	0.0	- iit,icc,iang1,iang2,iang3
58	10.0	10.0	10.0			- ia_hmax, ia_hmin, ia_vert
59	1	0.1				- inst, nugget effect
60	1	0.9	0.0	0.0	0.0	- iit,icc,iang1,iang2,iang3
61	10.0	10.0	10.0			- ia_hmax, ia_hmin, ia_vert

Each line in the parameter file is summarized below:

- Lines 5-8 are standard GSLIB-style data inputs.
- Line 9 is the number of unconditional realizations to simulate for optimization; the final objective value is the expectation across all realizations.
- Line 10 is the simulation type, either LU or sequential Gaussian simulation. LU is recommended for less than ≈ 2500 data.
- Line 11 is the random number seed.
- Line 12 is debugging options; some or all unconditional realizations are written to the file specified in line 13.
- Line 14 is the output file for the final network mixture model.
- Line 15 is the output file for the optimized network weights or the best population vector.
- Line 16 is the output file for the objective function value per iteration.
- Line 17 is the prefix for output files for the target and experimental (optimized) objective components.
- Line 18 is the total number of network layers, including input and output layers.
- Line 19 are the corresponding layer dimensions (number of nodes per layer). The number of layers must match line 18, or an error is thrown.
- Line 20 indicates the type of regularization (L1 or L2) and the regularization constant.
- Line 21 is the input file containing all elemental factor variogram models. Only one structure is permitted per model.
- Line 22 is the option to consider factor precedence. Precedence is specified in the file on line 38.
- Lines 23 and 24 select the objective function components and their weighting, respectively. Each component is automatically scaled internally; however, a user-defined weighting is permissible via Line 24.
- Line 25 specified the number of thresholds to consider, and line 26 specifies the threshold

values. The number of indicator thresholds must match the number of indicator variogram targets, or an error is thrown.

- Line 27 specifies a user defined weight to each threshold.
- Line 28 is an option to consider runs either above or below the threshold. Above the threshold amounts to $1 - I(\mathbf{u}_i; z_k)$. Line 29 is the maximum number of runs to consider.
- Line 30 specifies whether the target runs are calculated internally, or from a file. If 1, then the file on line 31 is read.
- Line 32 is an option to consider n -point connectivity either above or below the threshold. Above the threshold amounts to $1 - I(\mathbf{u}_i; z_k)$. Line 33 is the maximum number of connected steps to consider.
- Line 34 specifies whether the target n -point connectivity function is calculated internally, or from a file. If 1, then the file on line 35 is read.
- Line 36 contains DE parameters. F=mutation factor; CR=cross-over probability; pop. size=population size; its=total number of algorithm iterations.
- Line 37 are upper and lower bounds for each population vector.
- Line 38 specifies the file with upper and lower bounds of the ω parameters.
- Line 39 is an option for parallel DE.
- Line 40 is the number of directions for experimental variogram calculation. These experimental variograms are used to calculate MSE relative to the targets.
- Lines 41-44 are standard `varcalc` experimental variogram calculation parameters.
- Lines 45 and 46 are the number of continuous and indicator variogram model targets to consider.
- Line 47 is inverse distance weighting for variogram optimization. A larger power places more weight on optimizing shorter-range lags.
- Line 48 is the maximum number of experimental pairs to consider. This option can speed up optimization but results in greater error and potentially unstable results.
- Line 49 is the option to standardize variogram sills if not already.
- Lines 50-61 are standard GSLIB variogram model specifications for the targets outlined in lines 45 and 46. The continuous model is first followed by some number of indicators. The number of models here must equal line 45 + line 46, or an error is thrown.

The Gaussian pool file specified on line 21 contains a single structure variogram model for each factor in the input layer. An example of a pool file with three input factors is shown below:

1	1	0.0				<code>-nst, nugget effect</code>
2	1	1.0	-40	0	0	<code>-it,cc,ang1,ang2,ang3</code>
3	129.6	75.6	25			<code>-a_hmax, a_hmin, a_vert</code>
4	1	0.0				<code>-nst, nugget effect</code>

```

5   1   1.0  -40   0   0       -it,cc,ang1,ang2,ang3
6   232.2 124.2 50       -a_hmax, a_hmin, a_vert
7   1   0.0              -nst, nugget effect
8   1   1.0  -40   0   0       -it,cc,ang1,ang2,ang3
9   334.79 172.8 75       -a_hmax, a_hmin, a_vert

```

The structures can be permissible GSLIB variogram structures (spherical, experimental, Gaussian). White space is permitted between the models. `nst` may not be greater than 1. The program allocates arrays and reads data based on the specified input layer dimension; the number of structures must match the number of input nodes, or an error is thrown.

A.2 NMRIMP

NMRIMP provides latent imputation functionality discussed in Chapter 5. The program takes two data files as inputs: (1) the same drillhole data provided to NMROPT, and (2) the file containing the optimized network weights generated by NMROPT. The same network architecture must be specified in both programs or errors allocating arrays will occur. The program outputs all imputed realizations to a file with the same structure as the input data; each imputed factor becomes a new variable in the file. The parameter file is shown below:

```

1           PARAMETERS FOR NMRIMP
2           *****
3
4   START OF PARAMETERS:
5   data.dat           - file with data
6   1 4 5 6 7 11      - columns for dh, x, y, z, var and wt
7   0                 - normal score transform var? (0=no, 1=yes)
8   -1.0e21   1.0e21  - trimming limits
9   100              - number of realizations
10  5841044          - random number seed
11  1                - debugging level
12  nmrinp.dbg       - file for debugging output
13  nmrinp.out       - output file with imputed realizations
14  3                - number of network layers (input to ...
15  5 5 1            - network layer dimensions (input + ...
16  nmrwts.out       - input file with optimized network weights
17  0 1 -1.5         - consider factor precedence? (0=no, ...

```

18	0	0.9	1	- seed values? (0=no, 1=yes), thresh ...
19	20			- exclusion radius for seeded values ...
20	40			- maximum previously simulated nodes
21	50000	10000		- maximum iterations for step 1 and step 2
22	0.1	0.01		- rejection tolerances for step 1 and step 2
23	pool.dat			- file with cov. structs. of Gaussian pool

Each line in the parameter file is summarized below:

- Lines 5-8 are standard GSLIB-style data inputs.
- Line 9 is the number of realizations to impute.
- Line 10 is the random number seed.
- Lines 11 and 12 are the debug flag and the output file for debugging.
- Line 13 is the output file for imputed realizations.
- Lines 14-15 specify the architecture of the NMR. These parameters must match those used in NMROPT.
- Line 16 is the input file with optimized network weights. This file is output by NMROPT.
- Line 17 is an option to consider factor precedence; this must match that used by NMROPT.
- Line 18 is the option to seed values, the threshold above which to seed, and the factor index to seed. Line 19 is the exclusion radius for seeding and setting the semi-random path.
- Line 20 is the number of previously simulated nodes to consider in sequential simulation. 40-60 is typical for a 3D problem.
- Line 21 are the number of iterations for the coarse search (step 1) and solution polishing (step 2), respectively.
- Line 22 are the rejection tolerances for the coarse search (step 1) and solution polishing (step 2).
- Line 23 is the input file containing all elemental factor variogram models. Only one structure is permitted per model. This file should be the same as in NMROPT.