Boundary modeling with uncertainty

by

Steven Allen Mancell

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

Master of Science

in

Mining Engineering

Department of Civil and Environmental Engineering

University of Alberta

© Steven Allen Mancell, 2020

Abstract

Mineral exploration and mining are capital intensive and carry significant environmental and societal considerations. The feasibility of prospective mining operations hinges on a series of engineering decisions. The mineral resource estimation of the quality and quantity of resources greatly influences these decisions. Boundary modeling for defining subsurface geology is an essential aspect of the mineral resource estimation workflow. The proper spatial distribution of geological domains for further estimation is integral to having accurate and precise models. Uncertainty in the boundary and resource is quantifiable and is a result of sparse sampling and complex geology. Consequences of poorly defined boundaries include dilution of ore material, smearing of grade into uneconomical rock, and increased uncertainty in the deposit tonnage. These consequences directly impact the economic, environmental, and societal feasibility of operations.

Boundary modeling workflows commonly use implicit techniques that automatically derive domain extents from data. These models are explicitly checked and edited to ensure the numerical model reflects known geological attributes. This deterministic approach generates a single model output and does not carry a measure of uncertainty. Stochastic approaches to boundary modeling capture short scale variability of the geology; however, imparting geological knowledge on the model is difficult.

This thesis develops a new implicit methodology for boundary modeling that provides

globally unbiased models with uncertainty. The indicator approach maps a field of probabilities and applies a threshold that results in an extracted boundary. Uncertainty assessment by varying the indicator thresholds provides eroded and dilated boundaries and a zone of uncertainty. Boundary definition is a critical and early step in resource estimation. The modeling of subsurface geology from sparse drill holes carries significant uncertainty. The results of this thesis provide a novel approach to boundary modeling with uncertainty. The geostatistical techniques and concepts are reviewed, the proposed framework outlined, and implementation, including case studies, are undertaken.

Dedication

For Pippa, Cleo, Macca, and V.

Acknowledgments

This work is not possible without the guidance and insight of my supervisor Clayton, so thank you. I also want to thank the Center for Computational Geostatistics (CCG) and its members for their help and advice throughout my studies. The fostering of ideas and friendships has been a highlight of my professional career, and the experience of working with such talented and capable people is a motivator for future endeavours. Moreover, thank you to the CCG industry partners who funded this study.

I would also like to thank my family. To my mother and father, who have unconditionally supported me throughout life. To my brother and sister and Amy and Blake: you are constant role models for living a successful and fulfilling life. Thank you.

Last thanks to the Lagoboys for the continuous unfettered global adventures.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Description
2-D	Two-dimensional
3-D	Three-dimensional
CDF	Cumulative distribution function
FN	False Negative
FP	False Positive
GSLIB	Geostatistical software library
MCC	Matthews Correlation Coefficient
MPS	Multi-Point Statistics
NN	Nearest Neighbour
PH	Probability-High Selection
PI	Probability Interval
PTC	Probability-Threshold Curves
PTh	Probability-Threshold Selection
RBFs	Radial Basis Functions
SDF	Signed Distance Function
SGS	Sequential Gaussian Simulation
SIS	Sequential Indicator Simulation
TN	True Negative
TP	True Positive
TPG	Truncated Pluri-Gaussian

Chapter 1

Introduction

Resource modeling and estimation is a critical aspect of project development in the mining industry. The basis for building a capital intensive mine hinges on a series of engineering decisions occurring from initial exploration phases to commercial production. An early and integral step is estimating mineral quantities and qualities, known as resource estimation. With the scarce and variable data inherent to mineral exploration, the accurate and unbiased estimation of resources carries significant uncertainty. The quantification of resource uncertainty aids in decision-making. The decision to build or not build a mine is significantly affected by the estimated resources and reserves, further risk assessments, and projected cash flows.

An important aspect of resource modeling is the definition of geological boundaries. A boundary exists between rock types with different properties or of unequal value. Before resource estimation may occur, the study volume must be defined. This is done early in resource quantification (Rossi & Deutsch, 2014). Inside of a domain, the rock is geologically homogeneous with similar mineralization controls and statistical characteristics (Rossi & Deutsch, 2014). The domain boundaries define the extents for the estimation of rock quantities and qualities in the subsurface. Therefore, the accurate and unbiased placement of domain boundaries is integral to resource estimation and decision making. The accurate modeling of geological features is also important. Faults, folds, lithological contacts, and other geological complexities are often host to mineralization. The modeling of geological characteristics influences resources and reserves, mine design and planning, and future profits (Dimitrakopoulos, 1998).

After defining domain boundaries, the grades at unsampled locations within can be estimated. Geostatistics is the statistical inference of spatially and/or temporally

correlated data using a set of techniques to numerically model subsurface geology and quantify uncertainty (C. Deutsch & Journel, 1998). The complex nature of geology and sparse data, implies that effective geostatistical techniques must be implemented to accurately estimate the value at unsampled locations. Kriging and simulation techniques are popular and well documented as in C. Deutsch and Journel (1998); Journel (1983b); Rossi and Deutsch (2014). The defined resources subsequently play a pivotal role in the decisions to further advance a project.

1.1 Problem Motivation

Accurate boundary placement and access to uncertainty are vital for resource modeling and subsequent decision making. Correctly assessing boundaries is crucial to accurately and precisely estimating the location, tonnage, and uncertainty of a resource (Manchuk & Deutsch, 2019). These models are essential in mineral exploration and mining through all stages of drilling, resource and reserve estimation, mine planning, and operations.

Implicit modeling is a data-driven automatic boundary construction technique. By using implicit functions inferred from data, computer software interpolates and generates boundary models (Manchuk & Deutsch, 2019). Computer-generated, implicit modeling for boundaries is fast and results are reproducible if parameterized equally (Carvalho, 2018). Current practice often models geological features using implicit techniques while geologists inspect the wireframe to confirm and control the model's adherence to known geological constraints. Often referred to as a hybrid process between explicit and implicit techniques, the process is critical in imparting geological knowledge on the model.

The underlying algorithm for many boundary modeling implicit techniques is the Signed Distance Function (SDF). The SDF is a Euclidean distance measure between the nearest two samples of different indicator values. Negative values indicate 'inside' and positive values 'outside' of the domain. The SDF values are interpolated and the

boundary is extracted from the zero isosurface. Careful calibration of a modification constant, known as the C-parameter, allows access to uncertainty by returning a bandwidth between a dilated and eroded case (Hosseini, 2007). The SDF in the presence of variably spaced data or asymmetries can return conservatively biased boundary placement resulting in underestimated tonnages. The simple additive nature of the C-parameter results in a relatively constant uncertainty bandwidth that does not honour the spacing of local conditioning data. The inherent SDF bias, combined with the constant C-parameter bandwidth, can affect vital aspects of the boundary geometry and domain volumes. Bias and inconsistencies enter into resource calculations ultimately affecting risk assessment and subsequent decision-making.

Explicit techniques, such as stitching of Two-dimensional (2-D) sections into Three-dimensional (3-D) space, impart useful geological knowledge on the model, but are often time-consuming, difficult to reproduce, and do not provide for uncertainty. Conversely, a wholly probabilistic approach using stochastic methods is computationally cheap and fast, yet imparting geological attributes to numerical models is difficult.

An alternative is to use indicators to interpolate a field of probabilities. By thresholding the probabilities, boundaries can be extracted that are globally unbiased in volume and shape. Access to accurate and precise uncertainty is attainable via additional low and high threshold values for boundary extraction. By defining a relationship between probability and thresholding, eroded and dilated boundaries are extracted that give realistic uncertainty that honours the local conditioning data structure.

1.2 Literature Review

1.2.1 Distance Functions

The initial step in many implicit boundary modeling workflows is a variant of the distance function algorithm. In its simplest form, the distance function is a distance measure between data of different indicators or domains. It is related to the distance

between two domains and their interface (Wilde, 2011). Originating from level sets in mathematics, distance functions are useful in various disciplines, including medical imaging, computer-generated imagery, and mining (Carr et al., 2001; Cowan et al., 2003; Li, Xu, Konwar, & Fox, 2006). The distance function output is a set of scalar values that are then interpolated exhaustively across a domain.

1.2.1.1 Signed Distance Function

The use of SDFs for assessing the position of geological boundaries and quantifying uncertainty is common in implicit modeling (Martin & Boisvert, 2017). The algorithm considers the Euclidean distance between data of differing indicators and is 'signed' because it is negative if the sample is inside of the geological domain, and positive if outside. The SDF considers the minimum distance between two data points that are of opposite indicators (Wilde, 2011). The indicator formalism for defining the data is:

$$i(\mathbf{u}_\alpha) = \begin{cases} -1, \text{ if } \mathbf{u}_\alpha \,\epsilon \text{ geology} \\ 1, \text{otherwise} \end{cases} \text{ for } \alpha = 1, ..., n \tag{1.1}$$

Where $i(\mathbf{u}_\alpha)$ is the indicator at location $\mathbf{u}_\alpha$, equals -1 inside the geology, 1 outside of the geology, and is applied to all n conditioning data. The SDF is then calculated as:

$$\text{df}(\mathbf{u}_\alpha) = i(\mathbf{u}_\alpha) \cdot \min_{\beta=1,...,n} \left\| \mathbf{u}_\alpha - \mathbf{u}_\beta \right\|, \text{ for } \alpha = 1, ..., n, \text{ and } i(\mathbf{u}_\alpha) \neq i(\mathbf{u}_\beta) \tag{1.2}$$

Where the distance function at location, $\text{df}(\mathbf{u}_\alpha)$, is equal to the indicator at that location multiplied by the absolute minimum distance to the nearest sample of opposite indicator value. Anisotropy can be accounted for by:

$$\left\| \mathbf{u}_\alpha - \mathbf{u}_\beta \right\| = \sqrt{\left(\frac{x_\alpha - x_\beta}{a_x} \right)^2 + \left(\frac{y_\alpha - y_\beta}{a_y} \right)^2 + \left(\frac{z_\alpha - z_\beta}{a_z} \right)^2} \tag{1.3}$$

The distance between sample locations in the presence of anisotropy is a function of the respective vector components divided by anisotropy ranges in the specified direc-

tion. The x, y, z are oriented in the principle directions with the a values specifying the anisotropy. When the SDF is equal to zero, the model is transitioning. The iso-zero surface is the interpolated contact between the domain and outside or vice versa. The SDF field gradient is normal to the boundary surface defined between the two samples in the calculation; however, as one moves away from sampled locations, uncertainty in the boundary location is inevitable (Manchuk & Deutsch, 2019).

1.2.1.2 Modified Distance Function

Having unbiased and fair models is essential to resource estimation and resulting decision making. Therefore, modifications to the SDF include a C-parameter for uncertainty bandwidth range and bias correction β-parameter for appropriately centring the uncertainty bandwidth (Munroe & Deutsch, 2008b). These parameters are generally calibrated by cross-validation or jackknife (Hosseini, 2009; Wilde & Deutsch, 2011). The resulting modified-SDF is a modification of the original distance function with correcting C-parameter added or subtracted and the bias β-parameter either inserted as a multiplier or denominator to the distance function.

C-parameter The C-parameter is an additive factor applied to the SDF. It is added to the conditioning data if outside the domain, and subtracted from the data if inside:

$$DF_{mod} = \begin{cases} \text{df}(\mathbf{u}_\alpha) - C, \ \epsilon \text{ geology} \\ \text{df}(\mathbf{u}_\alpha) + C, \ \text{otherwise} \end{cases} \tag{1.4}$$

The modification increases the difference between the function values of inside and outside data. The calibration of the constant entails a jackknife procedure whereby a subset of data is left out, estimation follows, and a comparison to the subset of actual values ensues (Wilde, 2011). The subset locations can be correct – inside or outside. Alternatively, the estimation can be incorrect –outside and truly inside or inside and truly outside. The misclassifications are of interest. The procedure is an iterative process whereby the modeler changes the C-value until an acceptable amount of

misclassification is achieved. A common method for defining acceptable limits is to use elbow plots and select the distance, after which little change in misclassification occurs. Once interpolated, the results include eroded and dilated boundary models between -C and +C (Figure 2.6). Any estimated locations greater than C are considered truly outside of the domain, and any estimated locations less than -C are qualified truly inside. The band between -C and C constitutes the range of uncertainty within the model, whereby the true boundary position lies somewhere in between (Wilde & Deutsch, 2011).

β-parameter Bias exists if the expected value of the estimate differs from the expected value of the truth. Unbiasedness is key to tonnage estimation because it controls the model so that it is not too restrictive or too liberal (Carvalho, 2018). In order to control for bias, the β-parameter constant is either inserted to multiply or divide the original SDF formulation:

$$DF_{mod} = \begin{cases} (\text{df}(\mathbf{u}_\alpha) - C) \cdot \beta, \ \epsilon \text{ geology} \\ (\text{df}(\mathbf{u}_\alpha) + C)/\beta, \ \text{otherwise} \end{cases} \tag{1.5}$$

The β-parameter shifts the uncertainty bandwidth location that is controlled by C (Munroe & Deutsch, 2008a). A β-value of greater than 1 causes the iso-zero surface to expand, and a β-value less than one results in a contraction of the iso-zero surface. Figure 1.1 shows the effect of the parameter on the estimates.

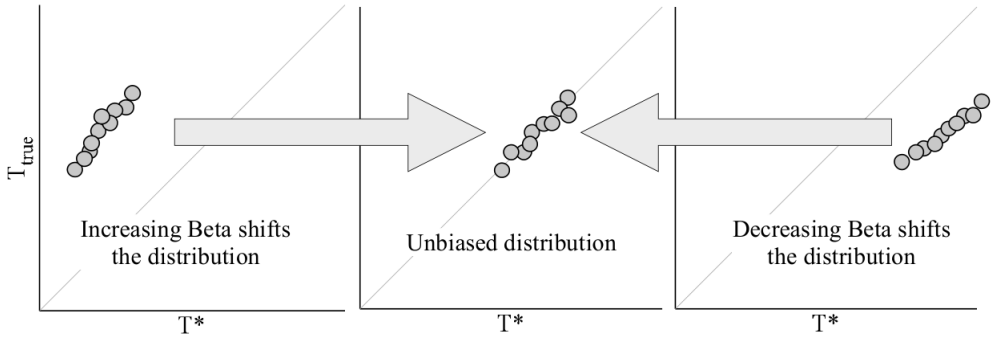


Figure 1.1: β-parameter and its effect on a set of interpolated estimates of tonnage (T*) compared to their respective True tonnages (T_{true}). The dots signify the values, while the 1:1 represents unbiasedness between estimates and truths. By changing β-parameter it shifts distribution to the right or left, for increasing and decreasing the β-parameter, respectively (Munroe & Deutsch, 2008a)

Multiple reference models are needed to calibrate the β-parameter. Optimization programs use reference models to arrive at a β-value that corresponds to a fair and unbiased estimate of uncertainty (Munroe & Deutsch, 2008a). The computational expense is significant because of the multiple reference models and optimization algorithms (Munroe & Deutsch, 2008b). Wilde and Deutsch (2011) outlines a calibration method only using the C-parameter and leaves β out of the workflow.

1.2.2 Other Boundary Definition Methods

There are varying techniques for modeling boundaries, with each having specific advantages and disadvantages in certain geological circumstances. In circumstances where data is too sparse for geological inference between sections, a nearest neighbour or geostatistical algorithm can be utilized. Boundary modeling algorithms can be categorized as deterministic methods– explicit or implicit– and stochastic methods.

1.2.2.1 Deterministic

Datasets comprised of large scale features in nature, such as exhaustive regional geophysics or mapping, lead to deterministic models. A deterministic model does not have a random component and expresses the deposit's large scale variability, but does

not measure uncertainty (Rossi & Deutsch, 2014). Deterministic models are unique, impart deposit specific interpretations, and are subdivided into explicit and implicit. Explicit methods impart useful geological knowledge upon the model and result from a modeler using 2D sections of geology to combine into a explicit 3D boundary model. In contrast, implicit methods are data-driven and automatically construct a boundary by fitting a function to the data.

Explicit Explicit boundary definition methods for geological modeling are based on manual digitization or stitching together of 2D sections and plan views into 3D wireframes(Carvalho, 2018). This manual process is a deterministic method that results in a single model. The manual approach imparts useful geological knowledge in the model; however, its application leads to long process times, difficult model updating, and no access to uncertainty. Moreover, because it relies heavily on a geologist's interpretation, it is subjective and unique to the modeler. Nonetheless, explicit controls on boundary models are crucial to resource modeling due to imparting valuable geological knowledge otherwise not included by implicit techniques. This knowledge is especially necessary for instances where data is sparse and, therefore, the importance of the modeler's geological knowledge increases. (Manchuk & Deutsch, 2019).

Implicit Implicit modeling refers to a data-driven automatic boundary construction. By using implicit functions inferred from data, computer software can interpolate and generate boundary models (Manchuk & Deutsch, 2019). Because it is computer-generated, implicit modeling for boundaries is fast and, combined with consistent parameterization, model repeatability is easily achieved (Carvalho, 2018). A current widespread practice automatically derives boundaries from distance function values. Techniques for the interpolation of distance function values map out a scalar field of data from which the extraction of boundaries can follow. A popular interpolation technique uses Radial Basis Functions (RBFs) with the extracted boundary transitioning from negative to positive values known as the 'isozero' line. The disadvantages of the technique are that volume uncertainty is not directly accounted for, and short-scale variations are problematic in sparse drilling areas.

Differences arise between explicit and implicit modeling when drilling is sparse; however, in the presence of tight definition drilling, the models tend to converge. Therefore, in the early exploration phases of a project, short term variability is often not captured resulting in smooth models. The smoothing often leads to biased mineral resources as the lack of variability on a local scale leads to optimistic boundaries that minimize ore loss and dilution and do not reflect reality (D. Silva, 2015). Other disadvantages are that uncertainty is not quantified, and subjectivity may lead to the same data resulting in varying models and resource calculations (D. Silva, 2015).

1.2.2.2 Stochastic

Variability exists at all scales in geology. The large scale variability and smooth models produced by deterministic methods do not adequately reflect the boundaries' small scale variability. A natural process evolution can be characterized by probability theory. This characterization is a stochastic process that shows the variability at all scales of geology (D. Silva, 2015). Stochastic methods of boundary modeling include process-based (Michael et al., 2010), object-based (Bridge & Leeder, 1979), Sequential Indicator Simulation (SIS)(Journel, 1983a), Truncated Pluri-Gaussian (TPG) (Galli, Beucher, Le Loch, Doligez, & Group, 1994), and Multi-Point Statistics (MPS) (Tahmasebi, 2018). Process-based techniques mimic a physical model of geological deposition or processes, such as channels, rivers, or deltas. Conditional simulation techniques are geostatistical methods which result in a probabilistic model for the random variables (Rossi & Deutsch, 2014). These include object-based modeling, SIS, and TPG. Object-based models rely on primitive objects from a library placed in a geological code matrix and used to simulate geological features. SIS estimates the probability of a particular geology code at unsampled locations using conditioning data. Monte Carlo simulation draws a value from a conditional CDF, and the value is added to the conditioning data set. Sequentially, all unsampled locations are visited as the process is repeated. TPS invokes precedence in the geological models by defining relationships between the units. Geological units are given a range in the Gaussian distribution and then simulated via Sequential Gaussian Simulation (SGS) (Rossi &

Deutsch, 2014). The reproduction of local data at their exact locations makes these techniques useful. The modeler has control, using variograms, on the objects' size, but noise and short-scale variability is often introduced, which appears unrealistic in the results and can affect engineering decisions. MPS invokes a training image representing the targeted spatial structure and reproduces complex geological features in the model (Tahmasebi, 2018). By exhaustively scanning the training image and identifying patterns derived from the conditioning data structure, the frequency of such configurations is used to build a categorical CDF that can be drawn from. The drawn value is added to the conditioning data, and the next node is visited, whereby the process is repeated. Difficulties using MPS arise from the selection of training image and dimensionality. The training image should represent the phenomenon being modeled while maintaining consistency with the data. The dimensionality drawback pertains to the vast combinatorial configurations being intractable.

The duality of macro-micro variability in geological bodies leads to combining deterministic and stochastic approaches, as done by D. Silva (2015). The importance of modeling variability at all scales in earth sciences is paramount to reflect the realities of complex earth systems. Downstream decisions on resource and reserve estimates are greatly influenced by the placement of boundaries and the inherent uncertainty.

1.3 Thesis Outline

Chapter 2 outlines the proposed framework for unbiased and fair boundary modeling using thresholded indicator estimates and highlights SDF modeling issues. It further develops the thresholding of indicator estimates to an unbiased global volume derived from a nearest neighbour (NN) model. Access to local and global uncertainty is essential in geological modeling, and the understanding of geometric and volumetric uncertainty is essential. Chapter 3 outlines unbiased boundary modeling in detail, showing the workflow for thresholding indicator estimates to achieve global unbiasedness and to further explain the intricacies of local unbiasedness. It further outlines the definition of boundary models that closely honour the data structure to give a

fair representation of the geological domain. Chapter 4 explores the workflows that target quantification of uncertainty in boundaries. The establishment of Probability-Threshold Curves (PTCs) for acquiring eroded and dilated boundary models gives access to eroded and dilated boundary models. The experimental PTC workflow and its results over thousands of models give a better understanding of the uncertainty involved with various model types and geometric shapes. Chapter 5 explains implementation details, including multiple category modeling and edge effects from over-extrapolation. Chapter 6 is a binary case study that uses a porphyry deposit dataset to illustrate the proposed workflow and compares the results to SDF models. Chapter 7 illustrates a multi categorical case study for the threshold indicator approach compared to an equivalent SDF model. Finally, Chapter 8 summarizes the conclusions of the study and describes areas for future work.

Chapter 2
Framework

Current implicit boundary modeling workflows often use the SDF to implicitly define the interfaces between domains (Martin & Boisvert, 2017). The modeling of uncertainty is through a modifying C-parameter to the distance function that creates an uncertainty bandwidth between -C and C. The result is a final model that consists of dilated, eroded, and base case boundaries. An alternative method to the SDF for boundary modeling is using indicators with an interpolator to map out a field of probabilities. These probabilities allow for immediate first pass assessment of uncertainty, and if thresholded correctly, give unbiased and fair boundary models to be used for further geostatistical estimation of rock properties. Further thresholding of the indicator estimates results in dilated and eroded boundaries with a bandwidth of uncertainty that follows the spatial configuration of conditioning data.

2.1 Signed Distance Functions Issues

The use of the SDF in implicit modeling workflows can introduce a conservative bias. The introduction of this bias occurs when the SDF algorithm reads between samples of different indicator values and assigns a value based on the Euclidean distance. Although the algorithm works as intended, it does not consider the structure of the surrounding data. Therefore, in the presence of data asymmetries, the algorithm will assign a 'distance' value to the closest datum without allowing for the consideration of where a boundary likely exists and, therefore, may disregard other more pertinent surrounding data. Additionally, the uncertainty C-parameter is an additive constant applied to the conditioning data, which results in unrealistic assessments of uncertainty that do not honour local conditioning data.

2.1.1 SDF in presence of asymmetry

In the presence of sparse data and variably spaced samples, the uncertainty inherent in the boundary of a geological domain is significant (Manchuk & Deutsch, 2019). As data is collected, the uncertainty in the boundary location decreases. Asymmetries in sampling and sparseness of data, both inherent in mineral resource models, can lead to situations where the SDF could introduce bias. The potential bias arises due to the function selecting the nearest sample of opposite indicator value with no regard to the structure and spacing of the surrounding data. A simple 1-Dimensional scenario illustrates this asymmetrical bias in Figures 2.1 & 2.3 and Figure 2.2 shows how the SDF should operate to be unbiased. The SDF algorithm incorrectly associates the outer negative data point with the nearest sample of opposite indicator value and subsequently creates a conservative bias. The estimated support in the RBF calculation in Figure 2.3 is the largest of the minimum circles that fit within the data configuration.

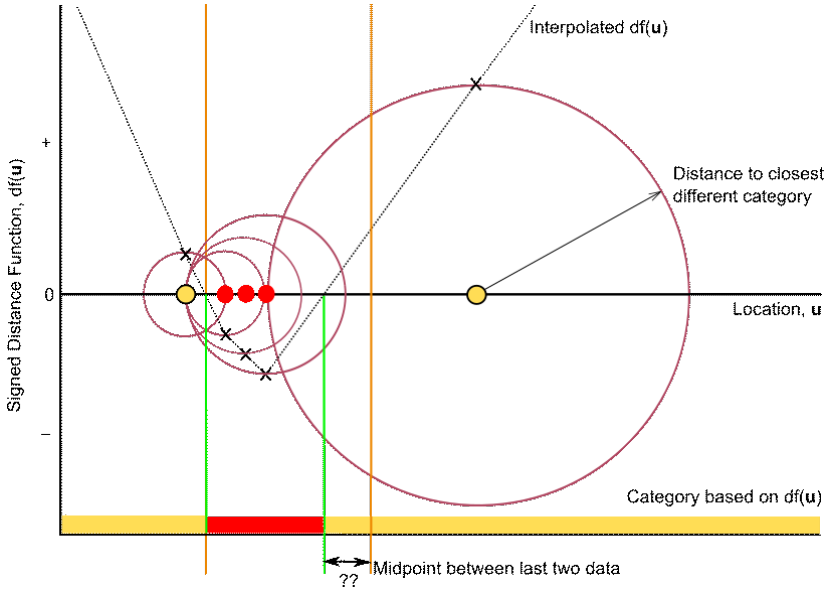


Figure 2.1: A 1-Dimensional SDF schematic showing two indicators inside (red) and outside (yellow) and the interpolated distance function. The circles show the SDF function algorithm selecting nearest indicator of opposite value. The dotted grey line shows the SDF value as the model transitions between positive and negative values. The X's on the circle represent the SDF value assigned to data points. The discrepancy between the last two samples shows a conservative bias in the value of the distance function notated by the gap with arrow. The third red inside data is assigned a value from the first yellow data instead of yellow value on oppsite right hand side.

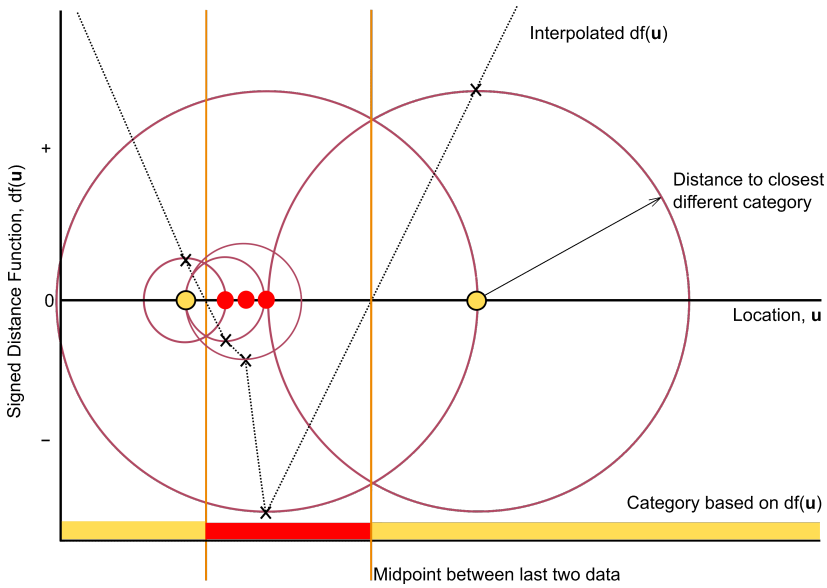


Figure 2.2: A corrected 1-Dimensional SDF schematic showing two indicators inside (red) and outside (yellow) and the interpolated distance function. The error observed in Figure 2.1, whereby the last two samples show a conservative bias, is mitigated by forcing the distance function to read the pertinent distal data. The circles show the SDF function algorithm selecting nearest indicator of opposite value. The dotted grey line shows the SDF value as the model transitions between positive and negative values. The X's on the circle represent the SDF value assigned to data points.

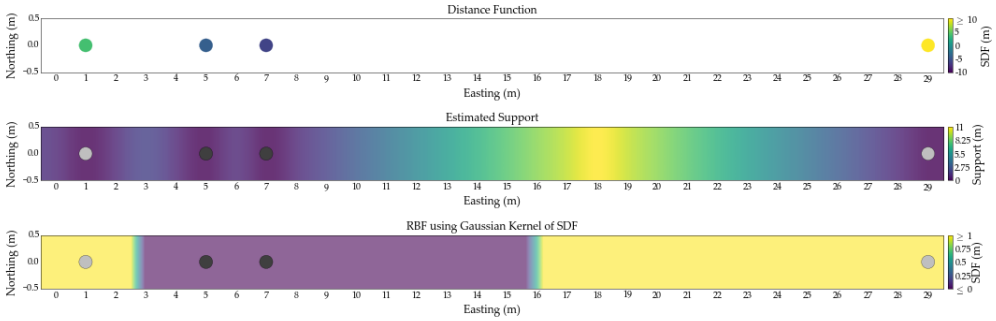


Figure 2.3: 1-Dimensional example of calculated SDF (top), estimated support (middle), and radial basis function interpolated SDF using Gaussian kernel highlighting the discrepancy between the known level of support and interpolated result. The support level of 18m is the midpoint between data yet is not recognized in the SDF interpolation where the conservative bias results in a contact at 16m–a global volume bias of -11%.

In a 2-D situation where inside domain data is tightly clustered along a specific trend, and the controlling outside data is concentrated on a particular side of that trend, the SDF will return conservatively biased results. The bias is due to the SDF algorithm

selecting the closest data on the clustered side while not selecting the relevant data further away on the sparsely concentrated side. The 2-dimensional schematic in Figure 2.4 exemplifies this phenomenon and illustrates a correcting assessment of the sparsely concentrated boundary.

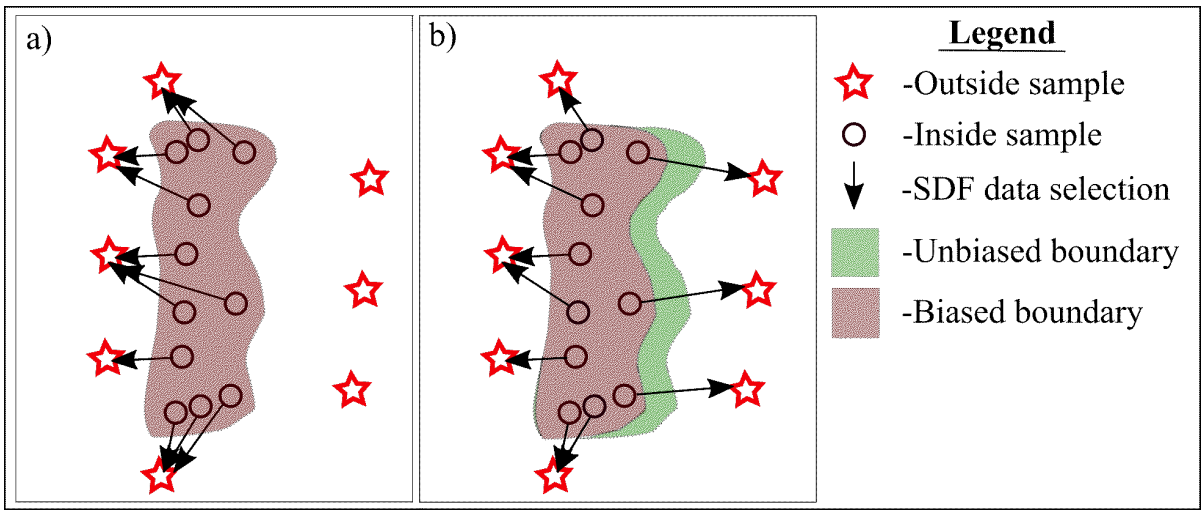


Figure 2.4: 2-dimensional schematic showing the SDF calculation with arrows indicating reading of data (*left*). In contrast, a more realistic formulation of a iso-zero boundary surface (*right*) where the sparsely concentrated inside data reads from relevant nearby data accounting for the structure of the data.

The insertion of control points to bound the external limits helps account for data asymmetry and the SDF's subsequent fallibility. The decision of control point placement is often a function of the inferred resources limits (Carvalho, 2018). Although this cosmetically alleviates apparent error in the boundary, the subjective nature of where to add control points is a challenge.

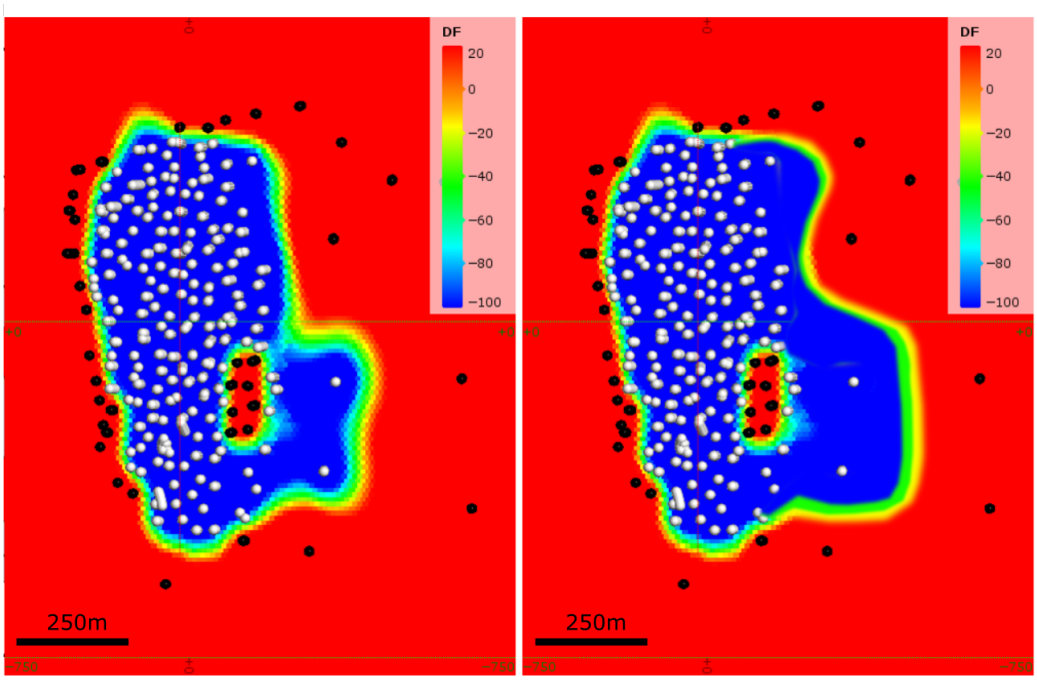


Figure 2.5: Plan view interpolation of distance function with the concentration of outside data (dark) along left-hand margin and hole in the domain (*left*) that shows an asymmetry issue in the calculation of the SDF. An edited version showing an unbiased boundary (*right*). From (Carvalho, 2018)

2.1.2 SDF & Uncertainty

The SDF C-parameter controls the bandwidth of uncertainty as discussed in Section 1.2.1. A jackknife workflow determines the constant, which is subsequently added to outside SDF values and subtracted from inside SDF values before interpolation. Because it is simply additive and global, the result does not consider local features in the data structure. This leads to a nearly uniform bandwidth of uncertainty regardless of local conditions. The β-parameter centers the uncertainty zone. The determination of the β-parameter comes from a reference model workflow that optimizes the parameter until a fair and unbiased estimate is reached. The arduous nature of the workflow combined with the excessive computational expense often results in the β-parameter being left out of the final boundary model (Wilde & Deutsch, 2011).

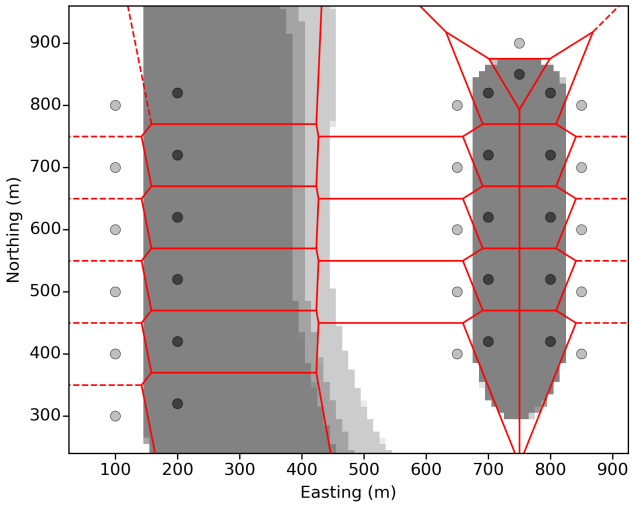


Figure 2.6: SDF interpolations with C-parameter of eroded -C case (black), base case C=0 isozero (dark grey) and dilated +C case (lighter grey). The uncertainty bandwidth shows conservative bias compared to a nearest neighbour globally unbiased model(red lines). The feature on the right has no uncertainty bandwidth on its southern extent where there should be given no outside conditioning data

2.2 Interpolation

The interpolation of the SDF values across the grid precedes the boundary extraction. The correct interpolator choice is imperative to return a reasonable boundary model that varies smoothly (Wilde, 2011). Computational cost, acceptable anisotropy types, and parameterization requirements must be considered (Martin & Boisvert, 2017). Algorithms considered include inverse distance weighting (Hosseini, 2009), kriging (C. Deutsch & Wilde, 2013), locally varying anisotropy (Lillah & Boisvert, 2012), or radial basis functions (Cowan et al., 2003; Martin & Boisvert, 2017). These interpolation techniques vary in their properties and parameters. Global methods are ideal because they utilize all conditioning data, are relatively fast, and result in artifact-free models (Martin & Boisvert, 2017). Global kriging and Radial Basis Functions (RBF) give identical results if parameterized equally (Fazio & Roisenberg, 2013). By eliminating artifacts that arise from nearby data searches, the distance function models using global methods vary smoothly throughout (Wilde, 2011). A Gaussian model for variograms and RBF kernels helps preserve the short-range con-

tinuity of the model (Martin, 2019). Moreover, parameterization of the variogram and kernel should include a nugget effect for mathematical stability (Wilde, 2011). Issues arising from global interpolation methods limit the allowable amount of conditioning data due to CPU and RAM constraints (Martin & Boisvert, 2017). The SDF is highly non-stationary, which is problematic for kriging methods that rely on first and second-order stationarity (Martin, 2019). RBFs' advantages over Global kriging are that RBFs do not rely on first-order stationarity assumptions, there is no need for variogram calculations, and the RBF does not depend on a modeled covariance function, thus honouring arbitrary shapes (Cowan et al., 2003; Hillier, Schetselaar, Dekemp, & Perron, 2014; Knight, Lane, Ross, Abraham, & Cowan, 2007).

2.3 Proposed Indicator Threshold Approach

The proposed framework for boundary modeling using indicator thresholding is explained in the following subsections. There are three main steps: Indicator estimation, NN-thresholding, and uncertainty assessment. Indicator estimation maps a field of probabilities for the occurrence of a variable. The probability of the category being present, or 'inside'. The values range between [0,1], and a specific value within that range corresponds to a threshold for a unique unbiased boundary model. A novel approach is to use the Nearest Neighbour (NN) model to determine the threshold. The value is used in conjunction with the CDF of the indicator estimate to extract a reasonable, unbiased boundary by considering all nodes higher than the threshold as being inside. Thresholding above and below the NN-volume ratio results in an eroded and dilated boundary that gives access to a bandwidth of uncertainty.

2.3.1 Indicator Estimation

Indicators for categorical modeling provide a straightforward, intuitive approach to estimating the occurrence of an inside value at an unsampled location. The estimation is of the distribution at a particular node, not the distribution parameters as continuous variable kriging (Rossi & Deutsch, 2014). The drill hole lithological

data is coded as being inside or outside the geology (Equation 2.1). Journel (1983b) outlines the methodology for predicting the conditional probability distribution at a given location by indicator kriging. The result is a field of probabilities mapped over the domain. Radial Basis Functions (RBF) are a popular technique for interpolating, and the results are the same as Global Kriging if parameterized equivalently (Fazio & Roisenberg, 2013). The following equations outline the RBF methodology used throughout this thesis.

The initial step is to codify the data to the appropriate categories as done in Equation 2.1:

$$i(\mathbf{u}_\alpha) = \begin{cases} 1, \text{ if } \mathbf{u}_\alpha \ \epsilon \text{ geology} \\ 0, \text{otherwise} \end{cases} \text{ for } \alpha = 1, ..., n \tag{2.1}$$

an RBF is used to interpolate the values across the grid. The indicator values at (x, y, z) location $\mathbf{u}$ is the weighted linear combination of all conditioning data evaluated on a radial kernel:

$$i^*(\mathbf{u}) = \sum_{\alpha=1}^{N} \lambda_\alpha \phi(|\mathbf{u} - \mathbf{u}_\alpha|) \tag{2.2}$$

where λ_α is a vector of the weights and ϕ is a chosen radial kernel. This linear estimator leads to system of equations given as:

$$\mathbf{D} \cdot \boldsymbol{\lambda} = \mathbf{b} \tag{2.3}$$

where $\mathbf{b}$ is a column vector of indicator values at locations $\mathbf{u}_\alpha$ for $\alpha = 1, ...N$, $\mathbf{D}$ is a square and symmetric interpolation matrix with $D_{\alpha,\beta} = \phi(|\mathbf{u}_\alpha - \mathbf{u}_\beta|)$ for $\alpha, \beta = 1, ..., N$. The $\boldsymbol{\lambda}$ is a column vector of weights from solving the normal system of equations. The solving of the weights leads to the ability to extract the value of the interpolation for any location.

The parameterization of the kernel, ϕ, is specific to the problem. Generally, for boundary modeling a Gaussian kernel is used as it best preserves the short-scale continuity. The Gaussian kernel is defined as:

$$\phi(r) = exp^{-\epsilon^2 r^2} \tag{2.4}$$

It is globally supported and positive definite (Martin, 2019). r-values are optional anisotropy ratios. ϵ is the support parameter or range of the kernel, and in the domain is the radius of the largest circle or sphere that can fit between sample locations (Fasshauer, 2007). A small nugget effect is often introduced for mathematical stability (Martin & Boisvert, 2017).

The indicator estimation results in the probabilities for each node being inside of the geology. The probability field allows for easy first-pass assessment of uncertainty as the estimation is of the conditional probability distribution at each unsampled location. The extraction of a boundary model follows from the thresholding of the estimate at a specific value.

2.3.2 Nearest Neighbour Thresholding

Nearest Neighbour (NN) volumes are reasonable to check an interpreted model's global volume (Rossi & Deutsch, 2014). The algorithm assigns a geological attribute to a given node by finding the nearest sample and assigning its value. The NN-model gives a unbiased spatial representation of the conditioning data and is a reasonable interpolator to use in sparse data areas with little geological knowledge. In 2-D, NN models begin by connecting all data points by Euclidean distance, forming Delaunay triangulations. Bisectors of the Delaunay lines extrapolate to form a Voronoi diagram forming cells for each conditioning data. Any location within a Voronoi cell is closest to the conditioning data of the cell. Figure 2.7 shows Delaunay triangulations and Voronoi cells for gridded and irregular data configurations. In practical 3-D situations, a search algorithm is applied at each grid location and the closest data value is assigned. A NN model is a reasonable reference for checking global volumes

of boundary models. The algorithm is based on the structure of the data without subjective parameterization that can result in biases. A boundary between two data in a domain with little other conditioning information would be reasonably separated by a boundary halfway and equidistant. The summation of cell volumes from NN models results in a global volume that is non-subjective and reasonable.

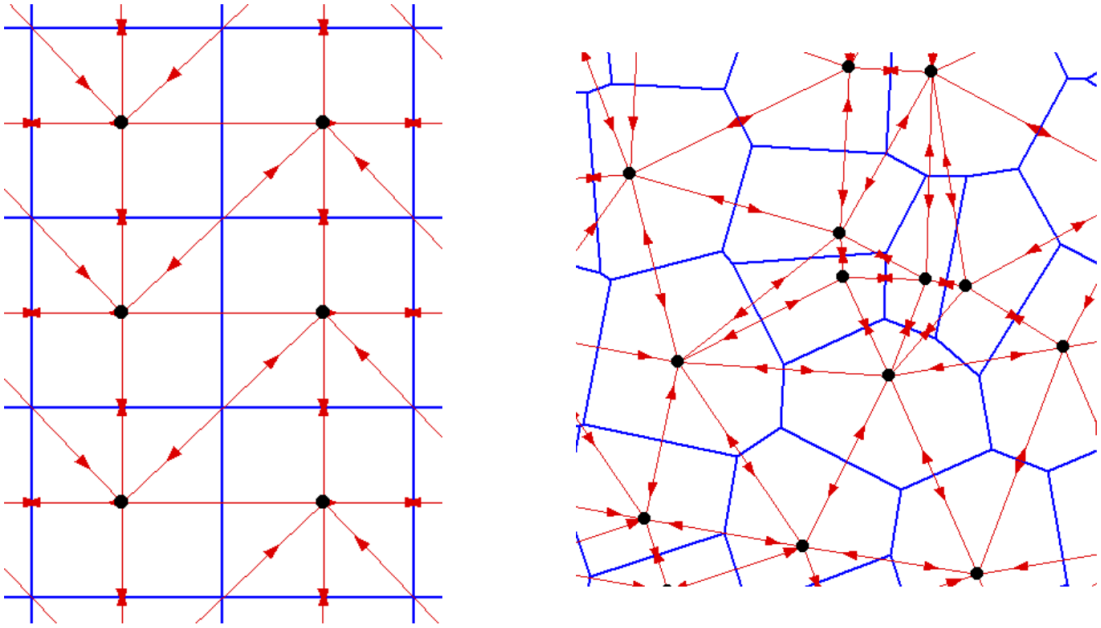


Figure 2.7: 2-D plan view Voronoi diagram showing gridded data spacing (*left*). 2-D plan view with irregular drill spacing and resulting Voronoi diagrams (*right*). The black dots are data, the red lines form the Delaunay triangulations, while the blue lines bisect the Delaunay lines and extrapolate to form Voronoi polygons. From (Naus, 2008)

The Voronoi polygon, or NN cell, containing a datum assigns all the nearby nodes with the appropriate conditioning indicator. A global volume is obtainable from the sum of all the inside cells. The inside volume ratio to outside volume results in a unique value between [0,1].

The NN-threshold value corresponds to a cumulative frequency of the indicator estimate's Cumulative distribution function (CDF). The NN-thresholding value between [0,1] extracts from the CDF the thresholding z-value. The z-value is the indicator estimate's extraction value that corresponds to a volume equal to the NN-model volume. Any indicator estimate value above the z-value is considered inside the boundary,

and the nodes with z-value probability form the interface.

2.4 Boundary Uncertainty

In geological boundary modeling there are generally two types of uncertainties: geometric and volumetric. The uncertainties influence one another and are ubiquitous due to the sparse sampling inherent to mining and mineral exploration (Lillah & Boisvert, 2012). Geometric uncertainty relates to the lack of knowledge of the orientation, domain extent, and domain continuity and shape in the deposit's unsampled areas interpreted by the geomodeler (Martin, 2019). Volumetric uncertainty pertains to the uncertainty involved between sample locations and the volume contained therein (Martin, 2019). Both uncertainties are crucial to final resource calculations and downstream decision making.

2.4.1 Geometric Uncertainty

The quantifying of geometric uncertainty in deterministic modeling workflows is difficult. The position of boundaries and corresponding thicknesses define geometric uncertainty. Orientations of structures and boundaries flow from the geological interpretations of the deposit type and measurements made in the field. Often, this control for geometric uncertainty comes from orientation measurements, which carry uncertainty due to error (Martin, 2019).

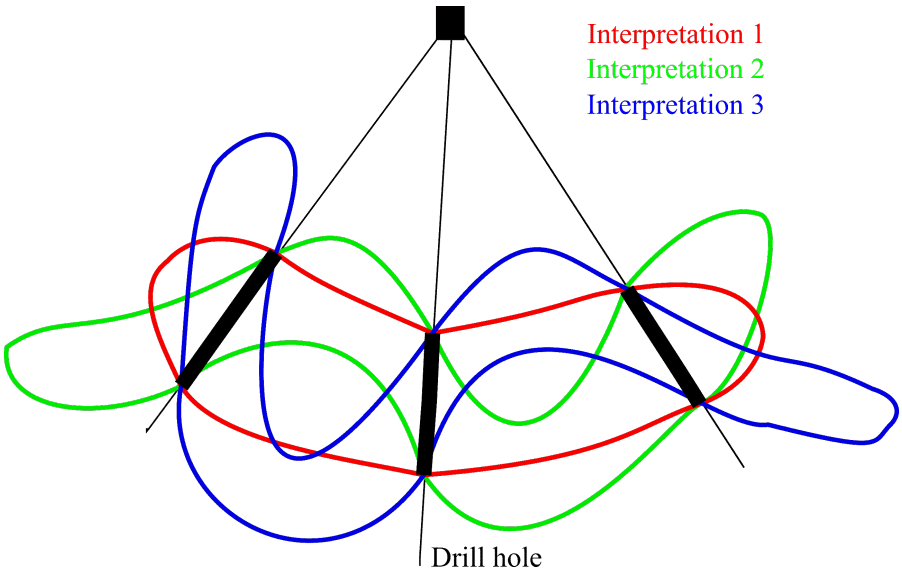


Figure 2.8: Three geometric interpretations derived from the same data in section view. The black square is the drill, the black lines are the drill hole traces, and the thicker traces indicate 'inside' geology. The interpretations are all vastly different given the same intercepts of inside geology and represent geometric uncertainty

Orientation data is useful for constraining the boundary surface. Simulation of orientation data and subsequent 3D geological modeling can lead to the location, quantification, and visualization of geometric uncertainty (Lindsay, Ailleres, Jessell, Dekemp, & Betts, 2012). However, in implicit geological modeling, these data are generally secondary to the indicator data. In most exploration and mining scenarios, the coordinate system used in the field is not congruent with the deposit geometry. Because of this, estimation or simulations may contain errors due to the techniques not following the geometry of the underlying geology. Carvalho (2018) proposed to mitigate these uncertainty issues in tabular vein deposits by shifting to local coordinates that follow the strike and dip of geological features. By rotating, translating, or fitting a plane to the structure and imputing data, there is an improvement in resource evaluation and modeling (Carvalho, 2018). Indicator interpolation using the dual form of kriging or RBFs maps out a field of probabilities based on the surrounding local data. In contrast to SDF modeling, the indicator approach does not rely on the scalar output data from a distance function that can be biased. The indicator methodology results in an estimate that adheres strictly to the structure of the surrounding conditioning data

while minimizing bias. Moreover, this translates into threshold models that not only have unbiased volumes but also follow the structure of the data effectively. Often, the geological model of deposition defines the geometric uncertainty in a boundary model and only volumetric uncertainty is quantified (Martin, 2019).

2.4.2 Volumetric Uncertainty

The volume of a boundary model directly affects the resource tonnages that follow from geostatistical estimation. Therefore, understanding the geological domain volume uncertainty is critical and will affect downstream financial and engineering decisions. Volumetric uncertainty relates to the bandwidth of uncertainty that exists between sample locations (Martin & Boisvert, 2017).

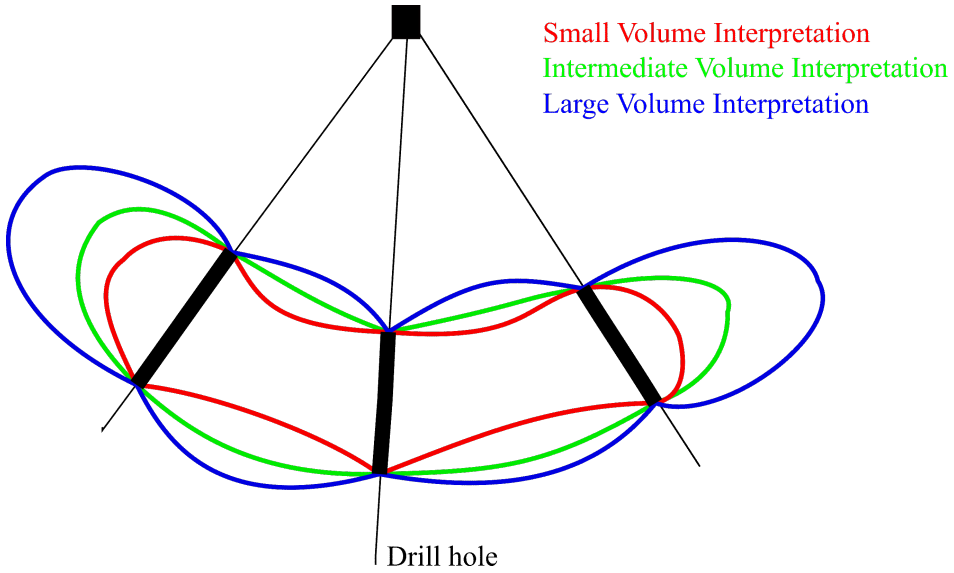


Figure 2.9: Three volumetric examples derived from the same data. The black square is the drill, the black lines are the drill hole traces, and the thicker traces indicate 'inside' geology. All interpretations are possible with the data given, but have vastly different volumes creating uncertainty.

The volumes from NN models are often used to check geostatistcal models for global reasonableness, and their proportions offer an avenue to threshold indicator estimates to result in equivalent global volumes. Defining uncertainty around the thresholded indicator estimate is important.

2.4.3 Probability-Threshold Curves

Synthetic datasets with known simulated true values are sampled and modeled in order to understand boundary uncertainty. The study of these models allows for the assessment of uncertainty over thousands of different scenarios. The uncertainty assessment is investigated through the development of Probability-Threshold Curves (PTC). These curves explain the nature of uncertainty as indicator estimates are thresholded through varying increments and compared to the true scenarios volume. In reality, access to PTCs is not possible due to the unknown Truth. Data in earth sciences are widely spaced and scarce. Therefore, the inference of the PTC shape of a particular deposit comes from the in-depth study of PTCs derived from simulated Truths and subsequent thresholding workflows.

The construction of PTCs begins with the simulation of true deposits. The deposits are generated by sequential indicator simulation (SIS). The true values represent reality as if everything about the deposit structure is known. The PTC workflow provides uncertainty in the boundary. Sampling the true deposits and global interpolation of the indicator values results in an estimate for the probability of a given node being inside the geological domain. The indicator estimate is thresholded from $p100$ to $p0$. Each threshold is a boundary model that is compared to the volume of the true deposit. A threshold is reached to find the reference true volume. Figure 2.10 shows the experimental PTC workflow for a single drill hole spacing.

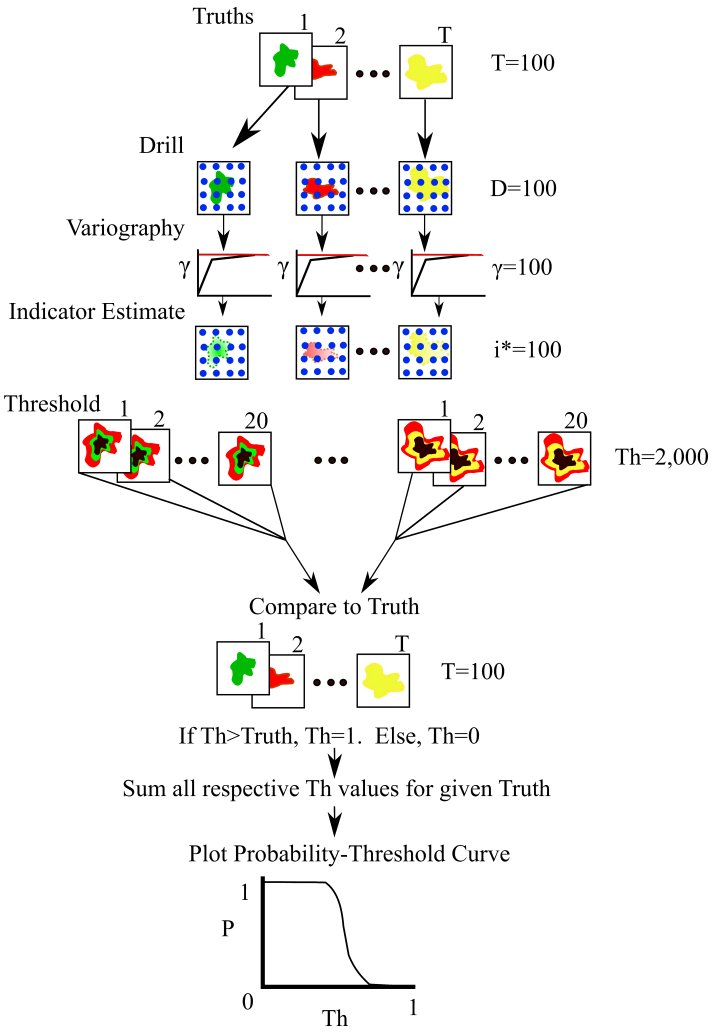


Figure 2.10: Probability-Threshold Curve construction for single drill hole spacing

The 100 Truths are drilled. Variography is done for each synthetic drill hole dataset, and indicator interpolation maps the probabilities of being inside the grid's domain. The models are incrementally thresholded and compared to their respective Truths. The sum over all 100 Truths gives a probability for a given threshold to be larger than the true values. The process is carried out on multiple geological scenarios with varying characteristics and drill hole spacings in section and plan view.

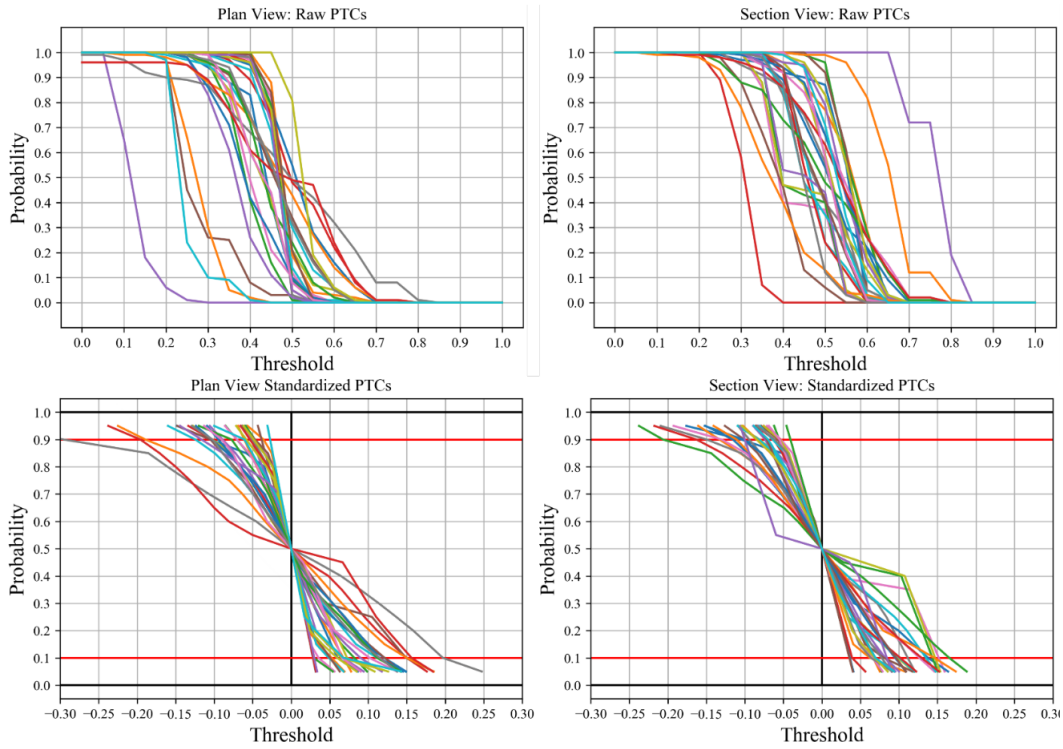


Figure 2.11: Probability-Threshold Curves for plan and section view with varying drill hole spacing and geology(*above*). Standardized PTCs for same scenarios trimmed to show slight variation at $p90$ & $p10$ (*below*)

The proposed workflow for uncertainty assessment with indicator thresholding is based off extensive studies of PTCs seen in Chapter 4. In reality, access to the true deposit is not possible. Therefore, the PTC for a given deposit has to be inferred. The findings suggest that further thresholding of the indicator estimate occur at +/-0.15 of the base case NN-threshold model. The resulting dilated and eroded boundaries give access to an uncertainty bandwidth that follows the structure of the local data and expands and contracts relative to the data configuration.

2.5 Implementation Details

The indicator threshold method is straightforward, although there are specific implementation details to note. Edge effects are a common issue with boundary modeling. In places near the model edge where there is sparse conditioning data, the tendency

for higher probabilities and boundary extension may occur. In SDF modeling, edge effects are controlled by the range of variograms, the insertion of controlling points, and applying the SDF in a hierarchical manner (D. Silva, 2015). The indicator threshold approach mitigates over extrapolation by the same avenues, but also, controlling the maximum distance radii in NN models allows the modeler to limit the resulting global volumes for thresholding.

In the multicategorical modeling of the SDF, each category is modeled separately. Overlap between the binary categorical models exists. Post-processing multi categorical models in the SDF workflow is accomplished by assigning the node with the most negatively predicted category (D. Silva, 2015). For the indicator threshold approach, the same problem arises. All categories are independently modeled, and overlaps between nodes may exist. An algorithm selecting the category with the highest differential between estimated probability and corresponding NN threshold is implemented to mitigate node overlap.

Interpolation of the SDF or indicators is needed to predict values at unsampled locations. Choosing an appropriate interpolator should be based on the interpolator's functionality and the information essential for parameterizing the technique (Martin & Boisvert, 2017). Global interpolators use all conditioning data and are advantageous as the models are artifact-free, and the algorithms are fast (Carvalho, 2018). The SDF is highly non-stationary (Martin, 2019). For multi categorical modeling, the ease of using one estimated support rather than multiple variograms results in a faster, easier workflow for the modeler. For data sets over N>30,000, the interpolation techniques become strenuous with CPU time and RAM constraints (D. Silva, 2015). Therefore, the global interpolator algorithms can be impractical for large datasets as the N X N matrix is dense. Specific solutions to this problem are simplifying the dataset, using sparse direct or iterative solvers, implementing domain decomposition, or using search-restricted Ordinary Kriging (Martin, 2019).

The methodology is deterministic and provides boundaries from conditioning data. If

the data spacing is large and geological units modelled are small, then any boundaries produced and derived from samples. For disjointed geological units and sparse data, the effectiveness of providing boundaries from this methodology will be questionable. Variability in earth sciences exists at all scales, and geology should be modelled stochastically for disjointed geological units in the presence of wide spaced sampling.

Chapter 3

Unbiased Boundary Modeling

Unbiased numerical models are essential to resource extraction because of the need to quantify and qualify the underlying material effectively (Martin, 2019). If there is a favour to a particular outcome, then it is said to be biased. Unbiasedness occurs if the expected value of the model tonnage (T*) matches the expected value of the real tonnage (T):

$$E\{T^*\} = E\{T\} \tag{3.1}$$

When estimates are greater on average than the truth, it would indicate a bias towards overestimation (Munroe, 2012). For statistical models to minimize bias, the input geological boundary models ideally will be unbiased. Therefore, geological boundary placement's influence has a direct impact on downstream reserve and resource calculations for mineral deposits. Moreover, the bias in the boundary model will not only pass to the final resource calculations, but it will also affect engineering and financial decisions that aim to maximize profit while minimizing costs and environmental impacts (Martin, 2019). Boundary models should be checked visually and statistically (Rossi & Deutsch, 2014). Local bias is the preferential outcome at a unsampled location and is related to the informing data. Overall, the global bias is the entirety of the model trending towards a particular outcome.

3.1 Global Unbiasedness

In the context of boundary modeling, global unbiasedness pertains to how the overall model compares to reality. If a preferred outcome arises globally, then the model is globally biased. Often interpolation techniques such as Inverse Distance Weighting or Nearest Neighbour methods are employed to check a geostatistically derived boundary

model's global volumes for validation.

3.1.1 Indicator Threshold Workflow

A workflow for attaining a globally unbiased boundary model follows. In conjunction with a NN model volume threshold, global indicator estimates result in a globally unbiased volumetric boundary model.

3.1.1.1 Indicator Estimation

The indicator estimation can be carried out with various interpolators, as discussed in Chapter 1. Global kriging limits artifacts by using all conditioning data and is computationally fast (Carvalho, 2018). Moreover, Ordinary kriging is a popular interpolator as it ensures global unbiasedness (Rossi & Deutsch, 2014). The example uses all ten data in a Global Ordinary Indicator kriging framework to map a field of probabilities throughout the domain seen in Figure 3.1

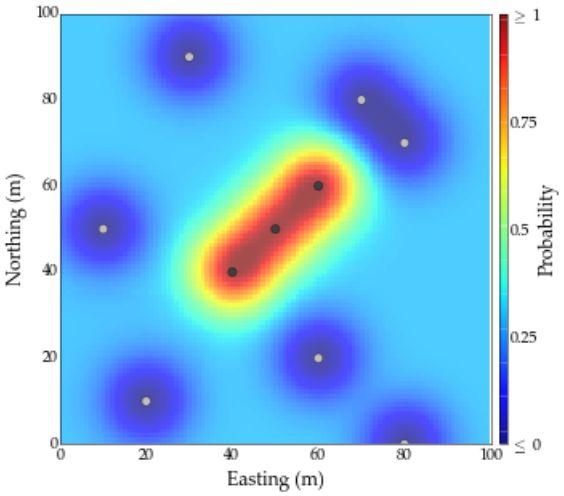


Figure 3.1: Global Ordinary kriging estimate of indicators

The probabilty of occurrence defines the discrete distribution at each unsampled location.

3.1.1.2 Nearest Neighbour

Nearest neighbour models are unbiased geometric representations of their conditioning data. Often modelers will check geostatistical models with NN model volumes for validation. The NN interpolation is a product of Delaunay triangulations and its dual, Voronoi plots. Delaunay triangulations connect the data with a line of Euclidean distance. The bisector of the connecting lines gives the mid-distance between the two points in question. The extrapolation of the bisected lines in space yields an unbiased spatial representation of the data (Iles, 2009)(Figure 3.2). The cells' closure yield volumes that each contain a single data point with cell walls that are halfway between neighbouring data points. These cell boundaries define the data's spatial representations, are unbiased, and form the transition point to the closest proximal data point. Therefore, the closest data point in space is within the cell.

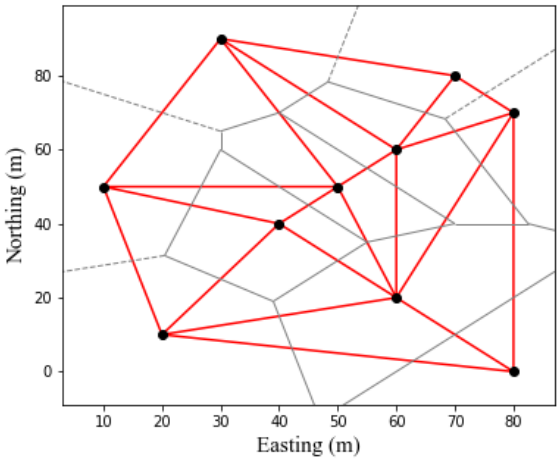


Figure 3.2: Data points in black, Delaunay triangulations in red, and bisectors joined to form Voronoi cells (grey). The Delaunay triangulations join data by Euclidean distance. The grey bisectors of the Delaunay red lines extrapolate to form the Voronoi cells

The Voronoi cells are the unbiased representation of the input data as a spatial decomposition.

3.1.1.3 Nearest Neighbour Volumes

Using a nearest neighbour model in conjunction with an indicator formualism, a suitable global volume is reached. The first step in the process is to define the indicators:

$$i(\mathbf{u}_\alpha) = \begin{cases} 1, \text{ if } \mathbf{u}_\alpha \, \epsilon \text{ geology} \\ 0, \text{otherwise} \end{cases} \text{ for } \alpha = 1, ..., n \tag{3.2}$$

Where $i(\mathbf{u}_\alpha)$ is the indicator at location $\mathbf{u}_\alpha$ and is 1 inside the geology and 0 otherwise calculated for all α conditioning data.

The implementation of NN workflow using Delaunay triangulations is seen in Figure 3.2. An unbiased volume for the domain is reached by flagging the Voronoi cells with the indicator constraints and calculating the inside cell volumes.

3.1.1.4 Thresholding

The resulting cumulative distribution function (CDF) of the indicator estimate is in Figure 3.3. Using the ratio of inside to outside cell volumes (Equation 3.3), a single frequency– the volume ratio– is determinable and coincides with a threshold value, z_{i^*}, for the indicator estimate CDF (Equation3.4).

$$th_{NN} = 1 - \frac{V_{NN}}{V_{Total}} \tag{3.3}$$

$$z_i^* = F_{i^*}(th_{NN}) \tag{3.4}$$

The z_{i^*}-value demarcates where the boundary exists. All nodes with probabilities above the z_{i^*}-value are qualified inside. Locations with estimates below the z_{i^*}-value are considered outside. The model transitions at nodes with predictions equal to the

z_{i^*}-value. The NN trimmed indicator estimate boundary model is in Figure 3.3.

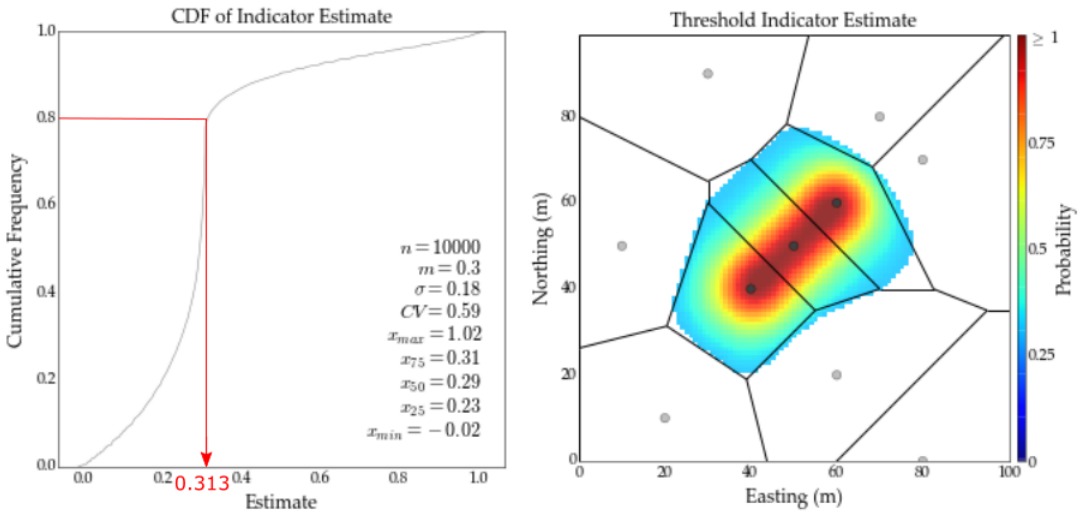


Figure 3.3: Cumulative distribution function of Global Ordinary kriged estimate with corresponding value of 0.313 to be thresholded to the indicator estimate (*left*). Thresholded estimate which represents the global unbiased volume (*right*)

.

The boundary from the indicator threshold approach closely follows the structure of the conditioning data. For comparison purposes, the SDF estimate interpolated using Global Ordinary kriging and isosurface is seen in Figure 3.4. The SDF boundary returns a conservative bias of 68% of the NN model volume. The indicator threshold model is 100% of the NN model.

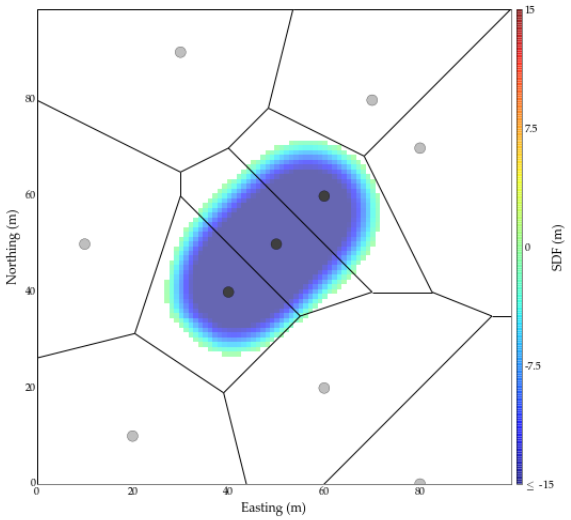


Figure 3.4: Global kriging of SDF trimmed to show iso-zero line location. NN model for comparison illustrates that the SDF interpolated returns a conservative volume of 68% of the unbiased global volume. Using the same variogram range as the indicator kriging, the volume was 42% that of the NN volume

.

3.2 2-D Case Study

Drill hole samples from a tabular vein deposit are in Figure 3.5. The indicator data is from Carvalho (2018) and exemplifies issues arising from SDF modeling. The roughly constant green-yellow bandwidth does not fluctuate in width between areas of varying sample concentrations. Moreover, the northeastern and southeastern side of the domain, where data spacing increases, shows significant conservative bias in the boundary. The boundary to the west is informed by close data spacing and is reasonable.

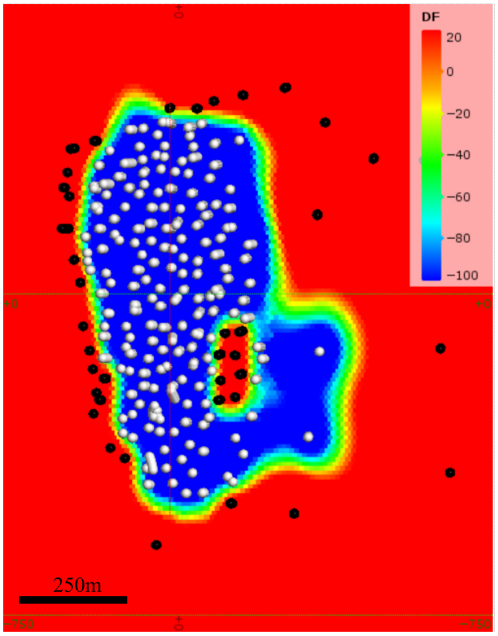


Figure 3.5: Original globally kriged estimate showing white inside vein drill samples, black outside vein samples from Carvalho (2018). The example show conservative bias resulting from local data asymmetry and unrealistic constant uncertainty bandwidth

Figure 3.6 shows the digitized indicator data which is ran through the Geostatistical software library (GSLIB) `DFcalc` executable to return the SDF values for the conditioning data (Wilde, 2011).

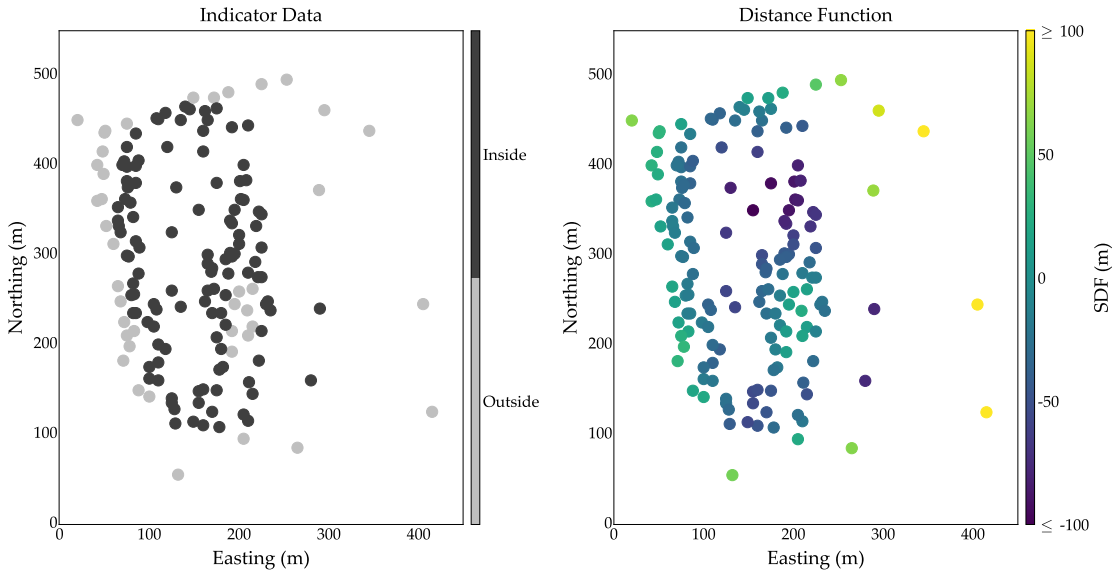


Figure 3.6: Digitized categorical data showing black inside vein drill samples, grey outside vein samples (*left*). The calculated distance function values for indicator data (*right*)

The SDF values are interpolated across the grid using an RBF with a Gaussian kernel. The indicator data also utilizes an RBF framework with an equivalent Gaussian kernel. Both estimates are seen in Figure 3.7. Note the relatively constant colour bandwidth around the entire vein irrespective of data configuration in the SDF estimate. In contrast, the indicator estimate has a variable colour bandwidth with tighter spacing giving smaller bandwidth on the western boundary and wider spacing, giving larger bandwidths on the eastern boundary.

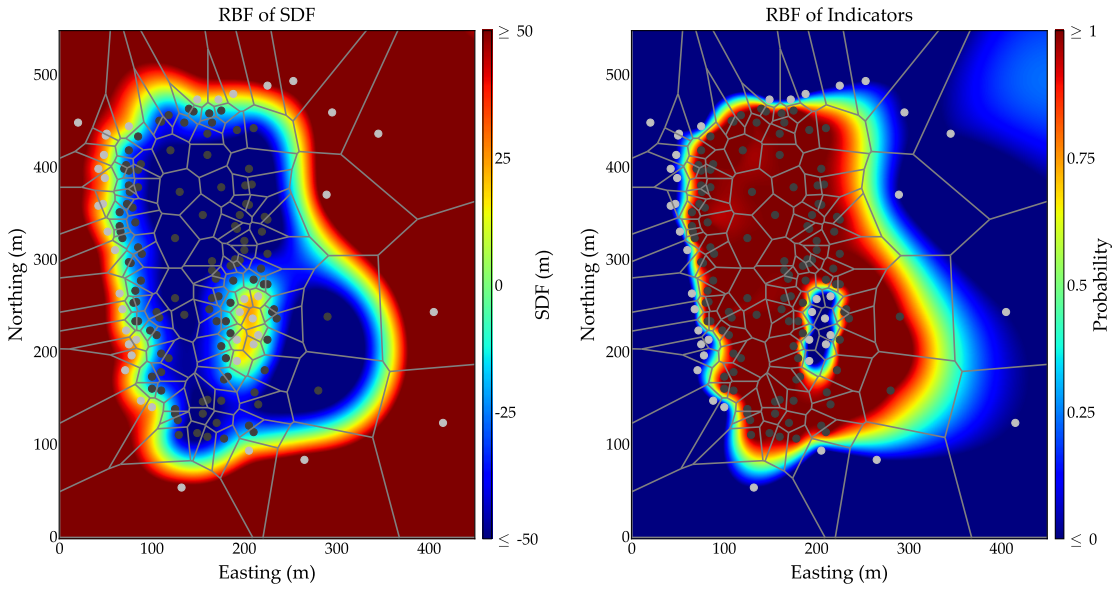


Figure 3.7: RBF interpolated estimate for SDF (*left*). Indicator estimate (*right*)

A NN model is constructed for the indicator data, and the indicator estimate CDF is thresholded to the unbiased NN volume (Figure 3.8 & 3.9). The NN model volume has a threshold value of 0.655, which corresponds to a z-value for the indicator estimate CDF of 0.37. Therefore, any estimate location with a probability above 0.37 is inside the vein; any probability below 0.37 is outside the vein.

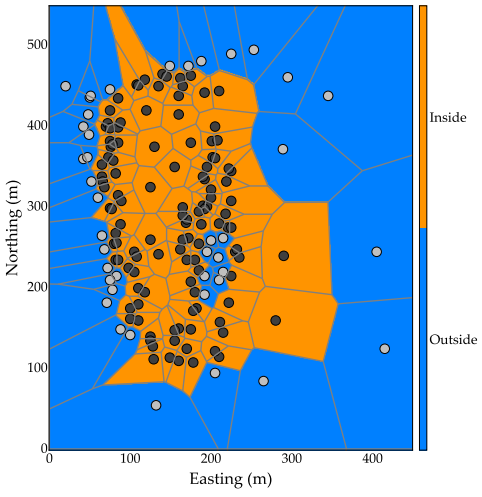


Figure 3.8: Nearest Neighbour model for indicator data

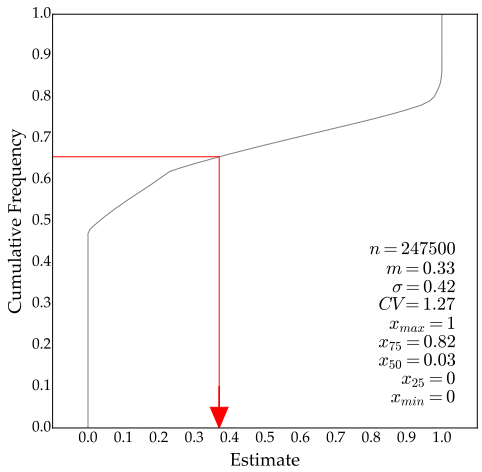


Figure 3.9: CDF of indicator estimate with red arrow showing threshold to NN model volume. NN volume ratio of 0.655 corresponds to z-value threshold of 0.37

The final SDF and indicator threshold boundary models are in Figure 3.10. The indicator threshold model returns an unbiased global volume and effectively follows the structure of the data. Moreover, first-pass uncertainty assessment from the field of probabilities in the indicator estimate is intuitive. However, the SDF modeling shows a uniform transition from the boundary inwards, not compensating for the conditioning data configuration.

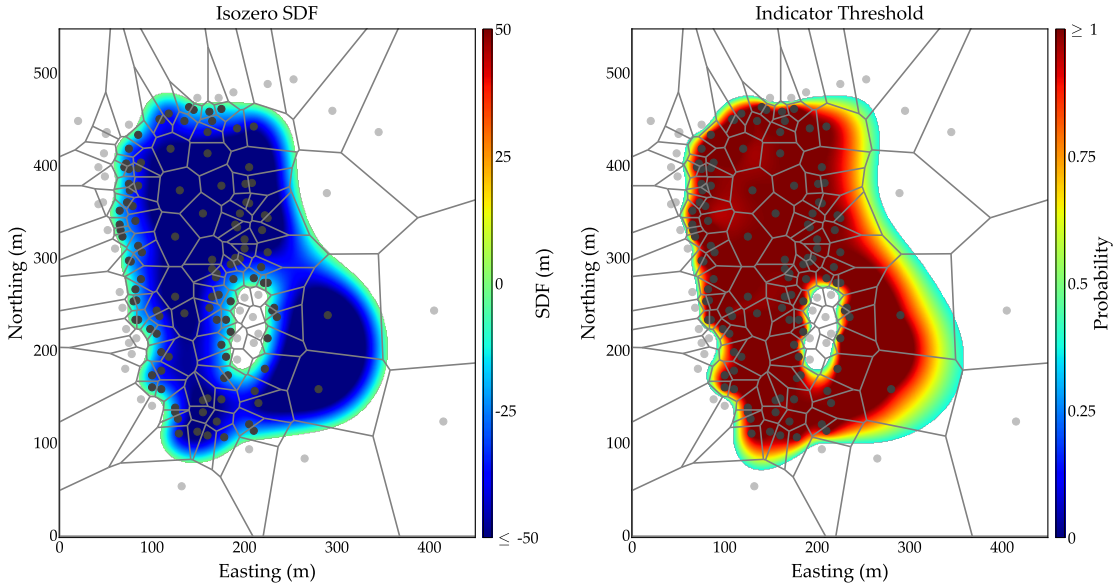


Figure 3.10: SDF model iso-zero interface. The total volume inside is 6.6% conservative(*left*). Indicator thresholded to NN model returning unbiased global volume inside(*right*)

In areas with dense sampling, such as the western boundary, the models perform similarly. In contrast, the models behave differently in zones of variable data spacing. The north-east boundary illustrates conservative bias in SDF modeling. The inside datum receives its SDF value from the sample immediately north of it. The northern boundary is close to halfway between the two data points of opposite indicators and appears reasonable. However, to the east, the conditioning outside data is further away, yet the inside datum's SDF value is smaller, leading to a conservative boundary. The same phenomena exist in the south-east boundary of the models. Apparent asymmetry in the configuration of the data is evident. The SDF inside data are closer to and read from the data near the vein's hole. However, the modeled boundary is on

the opposite side. Therefore, the SDF value from the western data is informing the eastern boundary. The proper SDF value for the eastern boundary would derive from distal data near the model area's edge. In contrast, the indicator threshold approach results in a boundary model that closely respects the configuration of the eastern data. Moreover, the indicator threshold model has a distinct gradational change in colour in this increased uncertainty zone. In both instances, the models need to have uncertainty bandwidth workflows completed better to ascertain the local and global uncertainty in the model.

Chapter 4
Boundary Uncertainty

Uncertainty is inherent and ubiquitous to any boundary in earth sciences. The uncertainty is due to data sparsity and randomness in the geological phenomena (Caers, 2011). Uncertainty in boundaries increases as one moves further from known contacts and samples. Global uncertainty relates to estimates of the volumes or tonnages of different domains. Local uncertainty relates to how the estimates vary at specific locations. A proper boundary model will account for both global and local uncertainty.

The uncertainty in boundary modeling is volumetric and geometric. Volumetric uncertainty pertains to the between-sample location bandwidth and volume. Geometric uncertainty relates to the orientation, extent, and continuity of the boundary between data (Martin & Boisvert, 2017). The volumetric bandwidth uncertainty can be modeled for a single geometric interpretation (Martin & Boisvert, 2017). In practice, volumetric uncertainty is often quantified, while the geometric uncertainty is thought to be satisfactorily accounted for by the geological deposit model. Boundary modeling and the resulting volumes play a significant role in the tonnage of a mineralized domain (Rossi & Deutsch, 2014)

In reality, access to the Truth is not possible, as there are no replicates of the exact geological setting. Therefore, the global uncertainty is always from a numerical model. Access to local uncertainty in an indicator model is available. The indicator estimate predicts the conditional distribution of uncertainty at each unsampled location. Local uncertainty is checked by various cross validation methods such as K-fold, jackknife or leave-n-out techniques. By leaving out a random, but uniform, subset of conditioning data and running the model workflow, the modeler checks the estimated locations that correspond to the subset data and compares the predictions to the

known values. Repeating the process gives access to local uncertainty. Moreover, the use of accuracy plots to visualize local uncertainty is effective (J. Deutsch & Deutsch, 2012). The accuracy methodology is a test for accuracy and precision of the indicator estimate and its closeness to the underlying Truth. By showing an indicator estimate is accurate, it follows that the local uncertainty estimated from the interpolator is reasonable.

4.1 Global Uncertainty

Precision is the variability of the models about the mean. The accuracy of the model relates to the correctness of the probabilities, that is, there should be half of the values above the P50, 10 percent below the P10 and so on. In the proposed framework, the integration of local uncertainty from indicator estimation is combined with global uncertainty for boundaries by further thresholding the estimate. The following workflow for synthetic datasets uses the local uncertainty quantification from indicator interpolation, in conjunction with a thresholding technique, to arrive at an accurate measure for global uncertainty.

4.2 Probability Threshold Curves

Probability Threshold Curves (PTCs) show the relationship between thresholding values and probability for an indicator estimate. By finding the appropriate threshold, the extraction of an unbiased and fair boundary model is possible. Uncertainty assessment from further studies of the PTC relationship and corresponding acceptable low and high thresholds is an essential aspect of the final boundary model. These thresholds correspond to eroded and dilated boundary models, which give access to uncertainty. The only direct access to the true PTC is from experimental PTCs constructed from synthetic data. The relationship and nature of the uncertainty is understood by simulating multiple possible Truths from the data configuration, sampling the Truths, indicator estimating with synthetic samples, and comparing

the volumes of the estimates to the corresponding Truths. Through construction of experimental PTCs over a wide array of model types, the study of patterns and characteristics of the PTCs is possible. The behaviour of the PTCs over thousands of models gives an understanding of the function that describes uncertainty. Moreover, a database of various models and their characteristics summarizes the simple shape of the function and its five parameters.

4.2.1 Experimental Workflow

The experimental workflow for producing PTCs begins with the creation of synthetic Truth models. By simulating true reference realizations of boundaries, sampling the realizations, and interpolating those samples, one can build a set of models. The comparison of the models and their corresponding Truths over incremental threshold values allows for the determination of the true PTC. The inspection of the PTC and associated thresholds allows for assessing boundary uncertainty for the final model.

The workflow begins with the simulation of the true reference models from a given data set. The use of the GSLIB executable `BLOCKSIS`, in conjunction with global proportions and conditioning data, allows for the creation of multiple true models (Hadavand & Deutsch, 2016). Manually specified local data constrain the models to be more realistic. The T=100 Truths represent possible realities given conditioning data and global proportions.

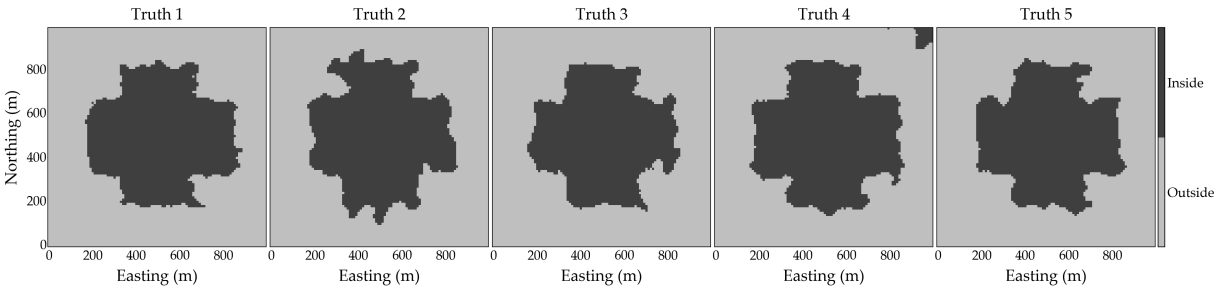


Figure 4.1: Simulated Truths from input data and global proportions using `BLOCKSIS` executable.

After the construction of the simulated Truths, the `GSLIB` executable `GSAMPLE` is run

for gridded drilling at varying drill hole spacings (e.g. d= 10m, 25m, 50m, 100m, 150m, 200m, 250m)(Pinto & Deutsch, 2015). The total drilled data comprises 700 sets of data from the original T=100 Truths. The omnidirectional variography is carried out for each set of synthetic drill data, and the indicator kriging estimation using `KT3DN` CCG software renders $i^*_{T,d}$=700 models.

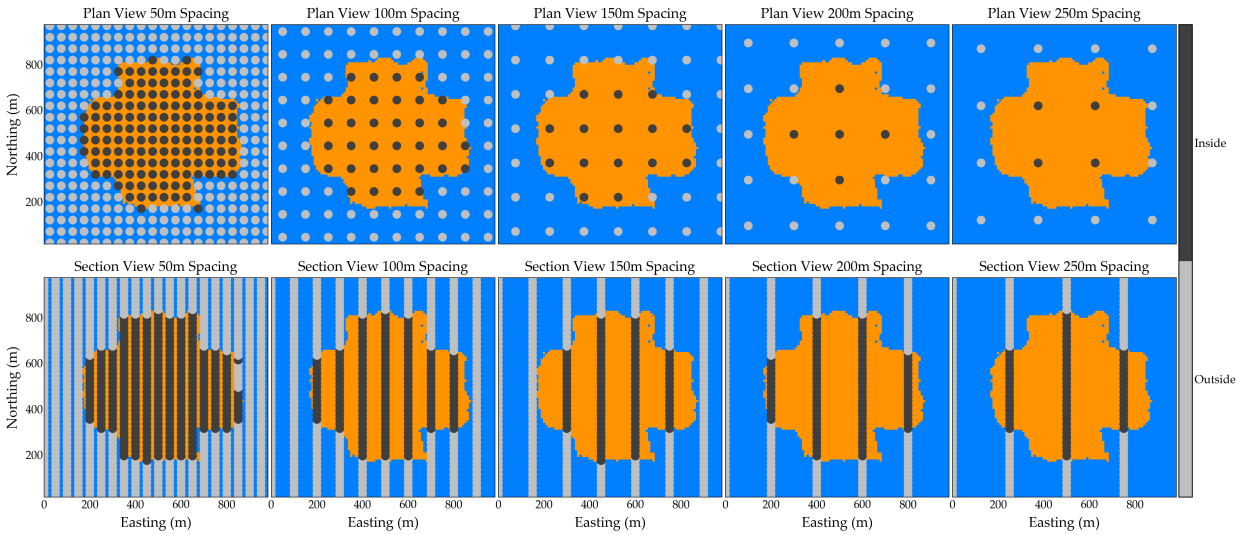


Figure 4.2: Truth Realization 1 with drill hole spacings 50, 100, 150, 200, 250m overlain in plan (*top*) & section (*bottom*). The 10m and 25m spacings are tight at resolution

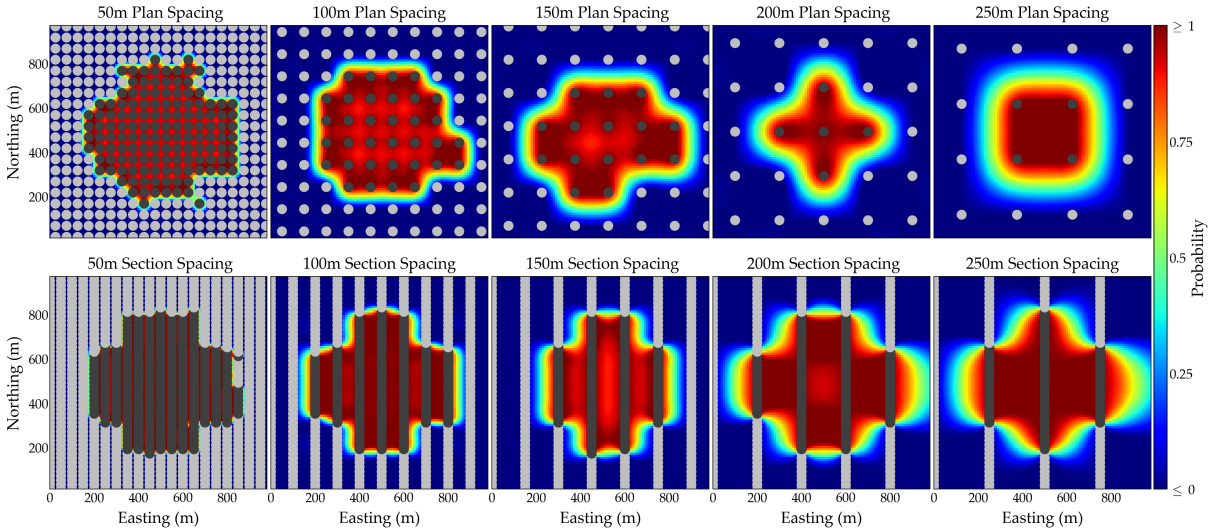


Figure 4.3: Truth 1 indicator estimates at respective drill hole spacings 50, 100, 150, 200, 250m in plan (*top*) & section (*bottom*). The 10m and 25m spacings are tight at resolution

A thresholding process of the $i^*_{T,d}$=700 indicator estimates at p05 increments from p0

to p100 ensues combining for a total of $i^{*,p}_{T,d}$=14,000 models. Comparison between the $i^{*,p}_{T,d}$=14,000 threshold model volumes and their respective T=100 Truth realizations follows whereby if the threshold model at a particular drill hole spacing is larger in volume than the corresponding Truth, it returns a value of 1; otherwise, a value of 0 is assigned.

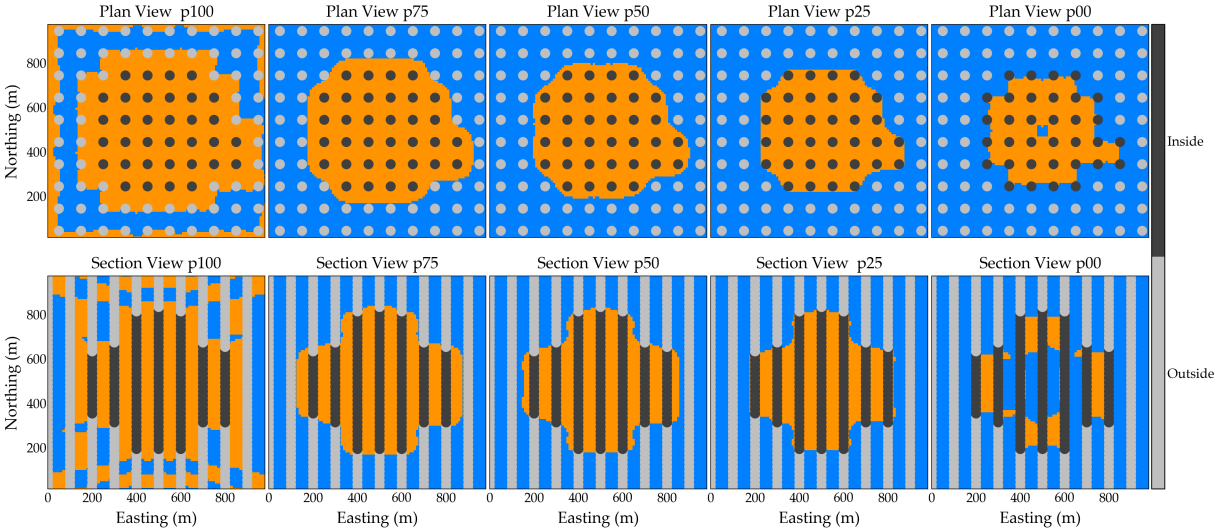


Figure 4.4: Indicator kriged models with 100m drillhole spacing and thresholds in plan (*top*) & section (*bottom*). The resulting models are compared to the corresponding Truth volumes

The cumulative number of models at specific thresholds larger than corresponding Truths are summed and divided by the number of T=100 Truths. The final data input to the PTC curve is an extrapolation of the threshold model comparison data for specific drill hole spacings plot against the thresholding values. The value of 1 on the ordinate probability axis signifies that all of the threshold models are larger in volume than the Truths at that particular spacing. Conversely, a value of 0 indicates that all the models were smaller than the corresponding Truths. The nature of the curve between probabilities of 1 and 0 varies as it transitions and defines the zone of uncertainty. The behaviour of the curves lead to building a database of varying input model characteristics such as shape, size, drill orientation, and structure.

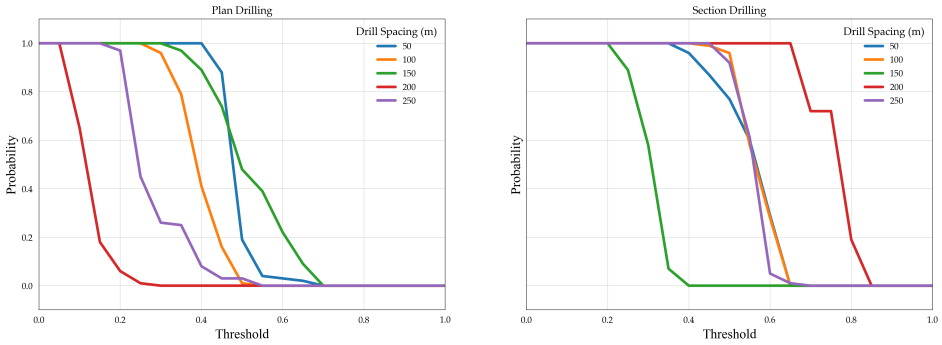


Figure 4.5: Probability Threshold curves for varying drill hole spacings in plan (*left*) and section view (*right*). A probability of 1 signifies that all threshold models at that spacing are larger than the Truths. Conversely, the probability of 0 indicates that all Truths are larger than the threshold models

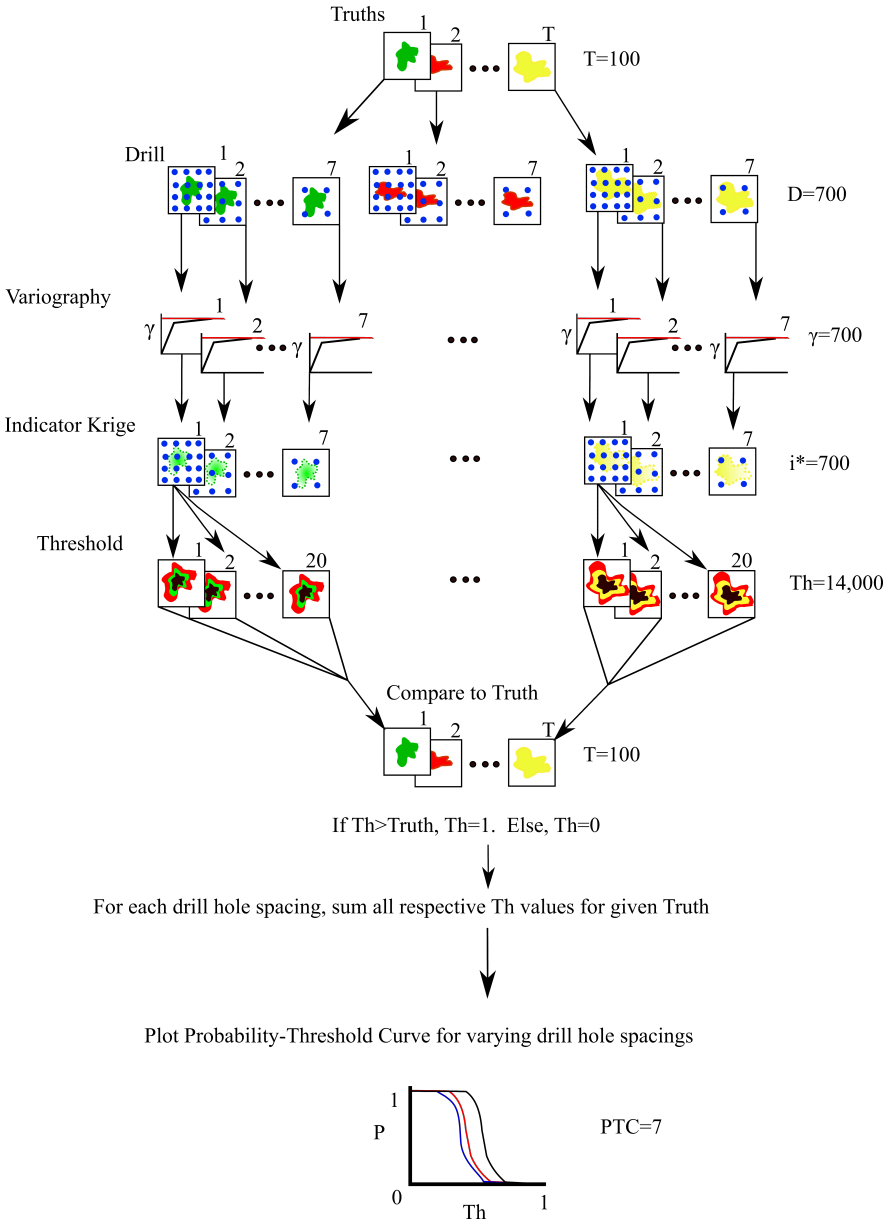


Figure 4.6: Schematic for the construction of Probability Threshold curves for varying drill hole spacings. Simulated Truths at various spacings and variography are carried out on each synthetic data set and fed into `KT3DN` software. The krige models are thresholded from p100 to p0 in p05 increments and compared to their respective Truths. If the models are larger than the Truth, the model is assigned a value of 1; otherwise, the value is 0. The model values are summed and divided by the number of Truths to return a probability of a threshold model being larger than the Truth for a specific drill hole spacing. The end product is an experimental PTC for each drill hole spacing showing the probability for a model to be larger than the Truth

4.2.2 Multiple Scenarios

To quantify and understand uncertainty in the context of complex and variable geological realities, there must be a variety of scenarios and PTCs that span a realistic space of geological model circumstances. The behaviour of PTCs as they relate to changes in the size of a domain relative to geology, varying geological shapes, drill spacing ratio to geology size and other geological attributes give insight. A subset of shapes can be seen in Figure 4.7 that represent synthetic Truths for testing. The experimental shapes are run through the PTC workflow over varying drill spacings in plan and section view (Figure 4.8). Plotting multiple scenario curves together illustrates the variation in model uncertainty. Standardization of the resulting PTCs centers the curves for closer observation and comparison purposes in Figure 4.8.

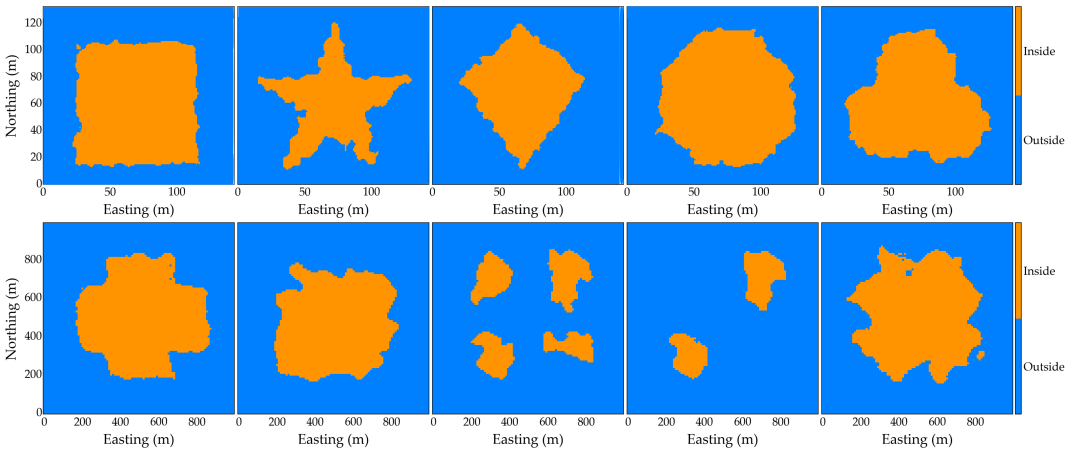


Figure 4.7: Truth 1 values for 10 example scenarios testing varying geological shapes

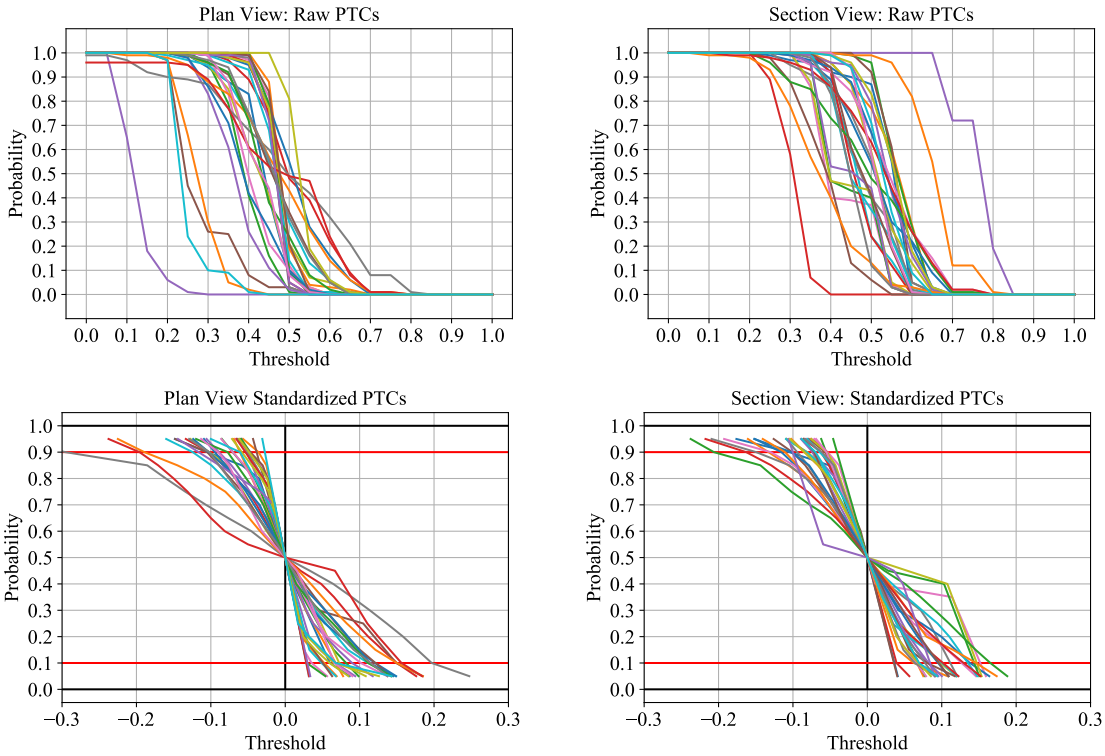


Figure 4.8: Raw PTC curves for a subset of 5 different geological scenarios at varying drill spacings in plan and section view (*top row*). Standardized PTCs of same scenarios and drill spacings (*bottom row*)

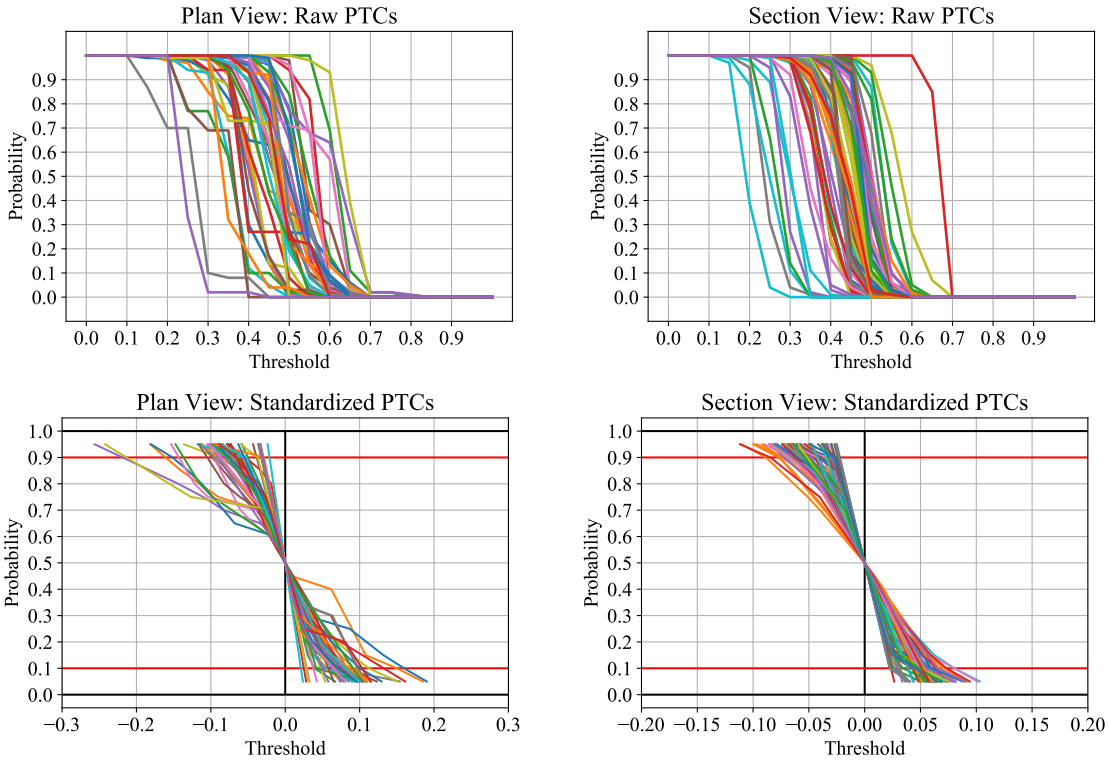


Figure 4.9: Raw PTC curves for a subset of 15 different geological scenarios at varying drill spacings in plan and section view (*top row*). Standardized PTCs of same scenarios and drill spacings (*bottom row*)

The tails of the PTCs are trimmed as the behaviour at the extremities is erratic and represents highly unlikely large or small models. Moreover, the threshold values for final uncertainty in the model will likely be within a more acceptable range centred around the unbiased NN model threshold. Figure 4.4 illustrates how the high and low threshold values ($p100$ & $p00$) correspond to unrealistic models and justifies trimming the curves for comparison purposes. The standardized PTCs show that the uncertainty bandwidth at the p95, p05 follows similar linearity with values ranging between 0.05 and 0.15.

4.2.3 PTC Function Fitting

Access to the Truth and replicates is not possible in reality. There is one true boundary that is unknown and must be estimated. Therefore, following the experimental PTC

workflow is impossible in reality. However, by building an experimental PTC database of different boundary model scenarios with differing characteristics, one gains insight into boundary uncertainty. The experimental PTCs for the varying drill hole spacings are fit with a function that is parameterized by five variables: an upper limit, base case probability, lower limit, upper exponent, and lower exponent. The curve is fit with an upper and lower function separated by the p50, where half of the models are larger than their corresponding Truths. Equation 4.1 shows the function parameters and a non-linear PTC is in Figure 4.10.

$$t = \begin{cases} \left(\frac{P-0.5}{0.5}\right)^{\omega_h}(b-h)+h, & for\ P > 0.5 \\ \left(\frac{P}{0.5}\right)^{\omega_l}(l-b)+b, & for\ P < 0.5 \end{cases} \tag{4.1}$$

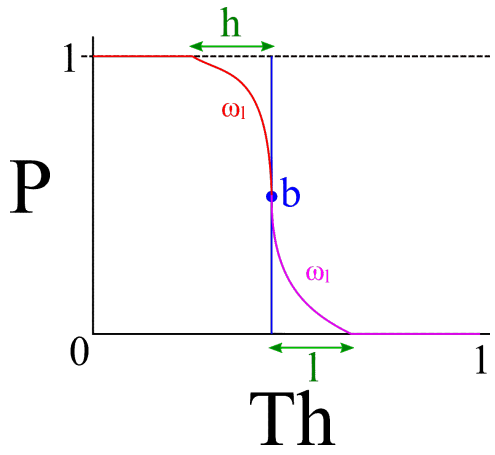


Figure 4.10: Parameterization of PTC function form. The upper (h) and lower (l) limits, and base case point, b, control the extents and centering, respectively. The w values control the curvature of the function

Where the threshold curve value (t), for probability greater than 0.5, is defined by the probability (P), an upper curvature parameter (ω_h), the base case probability (b), and the upper limit (h). For probabilities below 0.5, the threshold value is a function of the probability value (P), lower curvature parameter (ω_l), lower limit (l), and base case probability parameter (b).

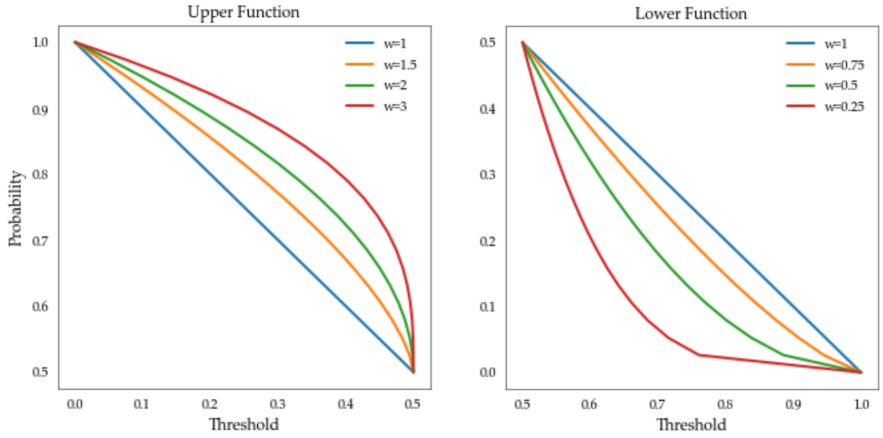


Figure 4.11: Curvature parameterization of PTC function form. The w_h values control the curvature of the upper function, while the w_l controls the curvature of the lower function

By fitting the functions to the PTCs from synthetic data, constructing a database of parameterized function values can help identify specific model characteristics and their potential effects on the curves. With changing input parameters and simulating new Truths, the process of creating PTCs and fit functions to build a substantial dataset of function parameters gives insight to boundary uncertainty and the relationship to PTCs.

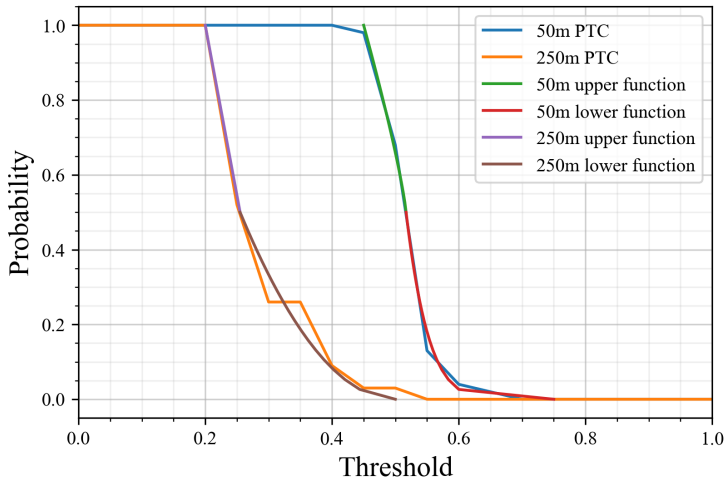


Figure 4.12: Fitted functions for 50m & 250m drill hole spacings. A total of 5 parameters would be produced for each curve: base case p50 (b), upper limit (h), lower limit (l), upper curve exponent (w_h), and lower curve exponent (w_l).

The function fitting for further analysis assumes the function is linear. The b-value for the PTC is the $p50$ model, where half of the models are larger than the Truths and half smaller. The final model's base case will come from the NN model threshold workflow described in Chapter 3. The upper and lower limits given by l & h define the linear function's slope. A linear model for the PTC is in Figure 4.13.

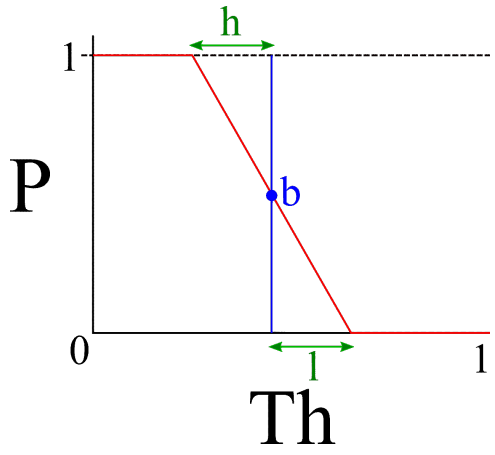


Figure 4.13: Parameterization of PTC function form. The upper (h) and lower (l) limits, and base case point, b, control the extents and centering, respectively

Increasing h and l decreases the function's slope, resulting in increasing uncertainty as the threshold models transition below the corresponding true model volumes.

4.2.3.1 Delta Values

The linear function for PTCs is parameterized by the upper and lower limits at $p95$ and $p05$ for h and l, respectively. The delta values, the difference between b and h & l, for the different scenarios gives insight into the uncertainty threshold for dilated and eroded boundary extraction. The upper and lower limits are considered equal and categorized together as the delta values. Figures 4.14 & 4.15 show the delta h & l values over n=311 geological scenarios.

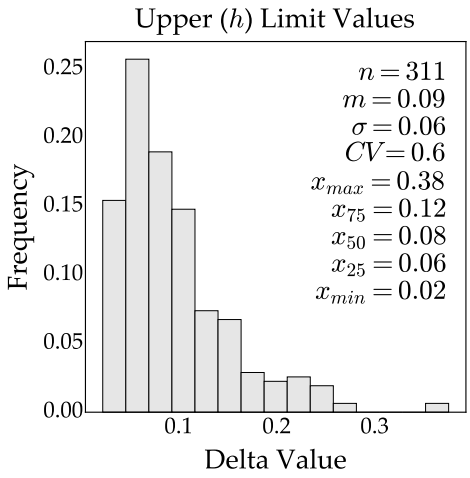


Figure 4.14: Upper (h) limit delta values for multiple scenarios

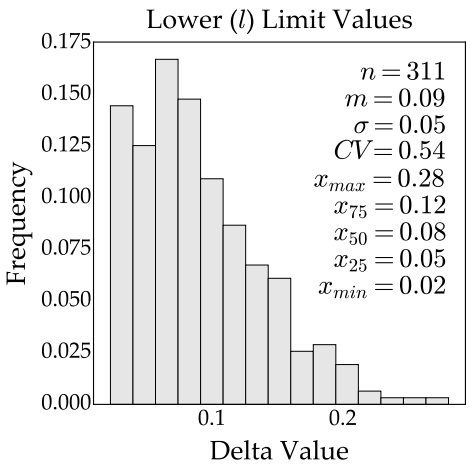


Figure 4.15: Lower (l) limit delta values for multiple scenarios

The means for the delta values are the same at 0.09. The standard deviations are near at 0.06 and 0.05 for upper and lower delta values, respectively. The higher delta values signify higher uncertainty in the scenarios. For delta value comparisons between plan view and section view drilling, the results are in Figure 4.16 & 4.17.

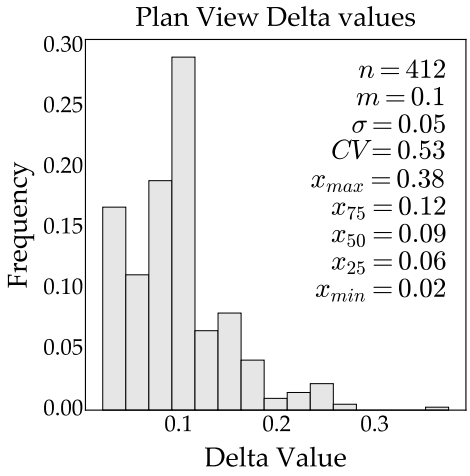


Figure 4.16: Delta values for all plan view drilling scenarios

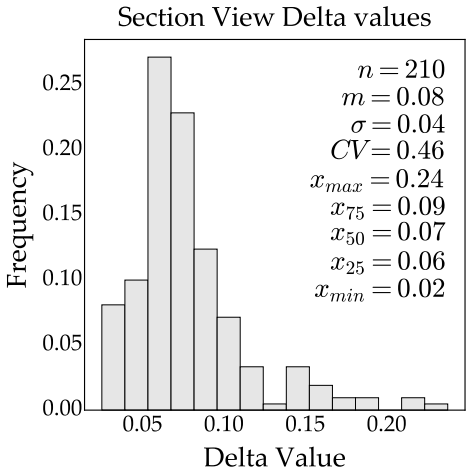


Figure 4.17: Delta values for all section view drilling scenarios

The section view scenarios have lower mean values and standard deviations. Referring to Figure 4.4, the data informing the section estimates are tightly spaced downhole.

In contrast, plan view estimates have sparse data leading to higher delta values and uncertainty.

Figure 4.18 shows the delta values over all geologies, drill hole spacings, in both plan and section view. The delta values describe the uncertainty relationship for the indicator estimates to the corresponding thresholds in a linear relationship.

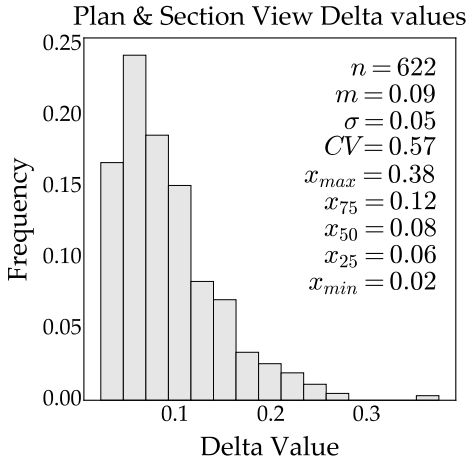


Figure 4.18: Delta values for all geological scenarios

The mean for delta values trimmed at $p95/p05$ is 0.09, with a standard deviation of 0.05. The distribution is skewed to lower delta values, with only 25% of values over 0.12. Understanding the controls on the delta values is crucial for determining appropriate uncertainty thresholds. Distinct patterns emerge when assessing the experimental standardized PTCs. A prominent predictor for increased uncertainty is the amount of informing data. The increasing drill hole spacings result in less conditioning data for the indicator estimation. The higher delta values associated with increasing drill hole spacings lead to more uncertainty in the boundary model. The higher uncertainty is reflected in the PTCs by a shallower slope. Figure 4.19 shows a scenario with Truths 1-5 drilled with increasing data spacing and the corresponding $p50$ models. The PTCs for the example are in Figure 4.20.

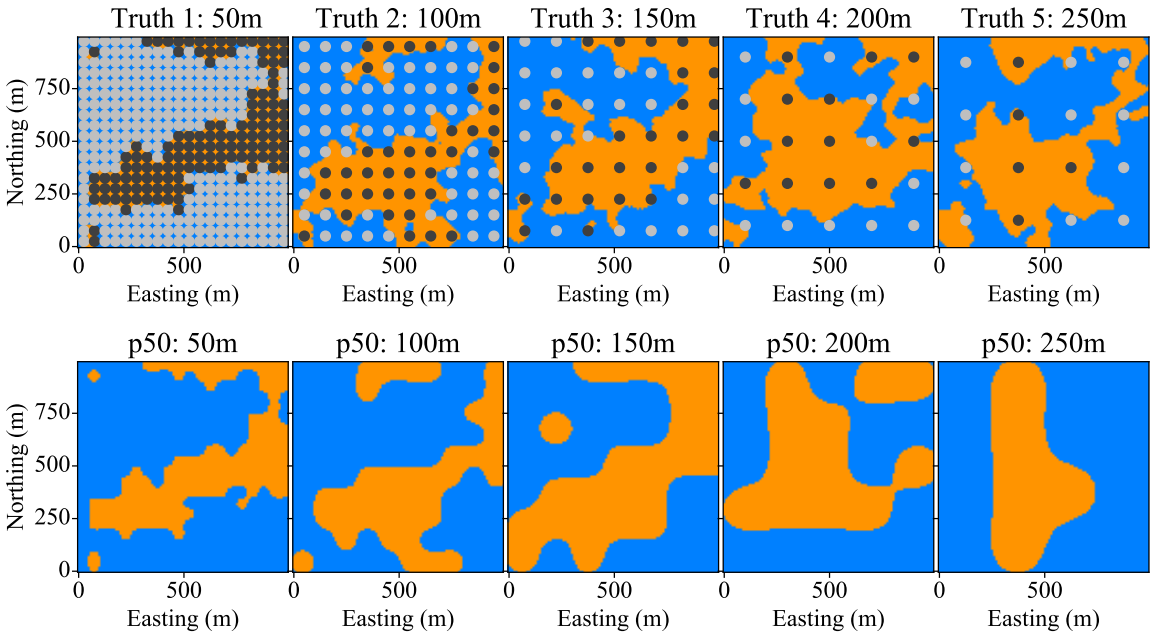


Figure 4.19: Truths 1-5 with increasing data spacing (*top row*). The 10m and 25m drill hole spacings are tight at resolution and are not included. Corresponding $p50$ indicator threshold models (*bottom row*)

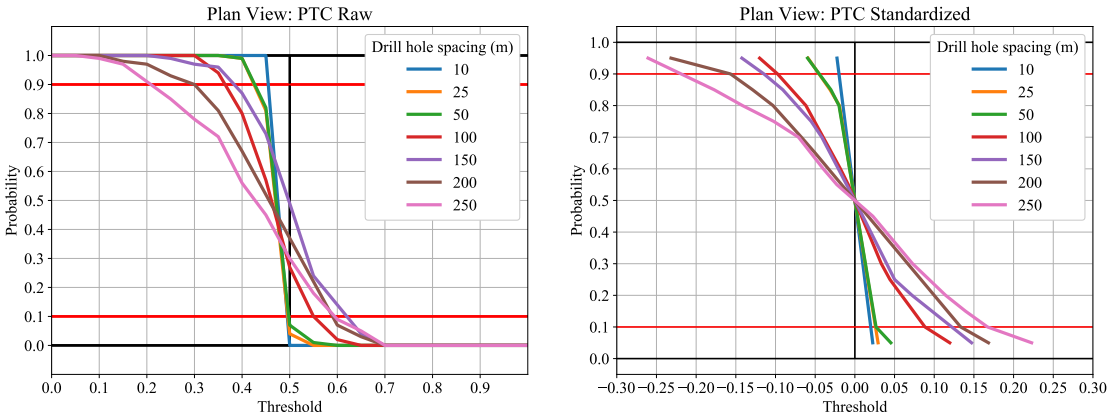


Figure 4.20: Raw PTC curves of geological scenario with seven drill spacings in plan view (*left*). Standardized PTCs of same scenario and drill spacings trimmed at $p95$-$p05$ (*right*)

The example illustrates how increasing sample spacing leads to larger delta values, shallower curves, and increased uncertainty in the boundary model. The approximation model is also appropriate as the PTC curvature is reasonably linear. Section view drilling contains significantly more samples compared to the plan view. The synthetic drill holes provide sufficient information for the estimation algorithm. The

results of section drilling are observed in Figure 4.21 with corresponding PTCs in Figure 4.22.

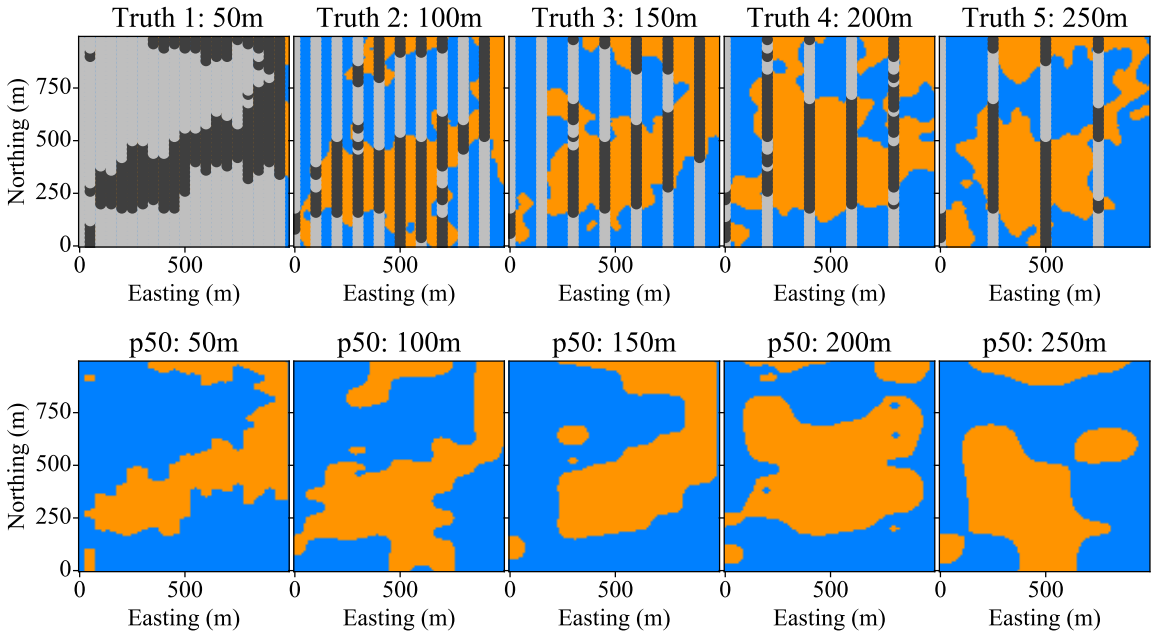


Figure 4.21: Truths 1-5 with increasing data spacing (*top row*). The 10m and 25m drill hole spacings are tight at resolution and are not included. Corresponding $p50$ indicator threshold models (*bottom row*)

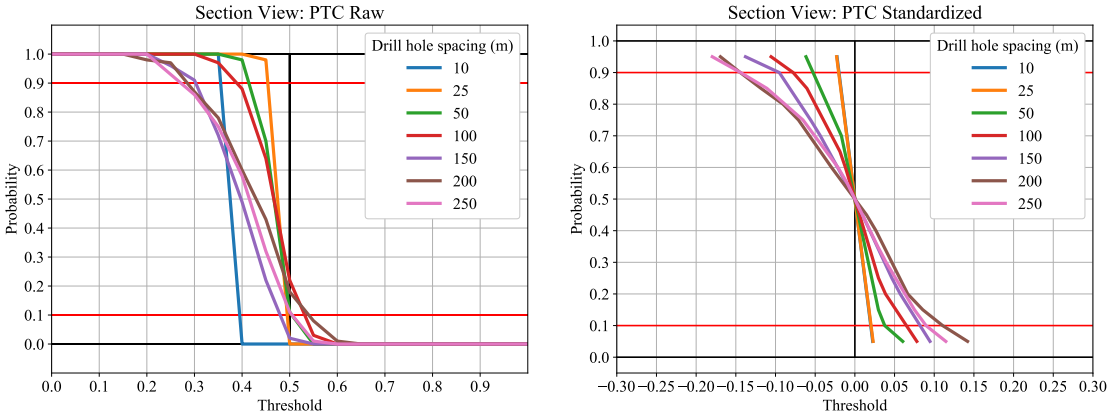


Figure 4.22: Raw PTC curves of geological scenario with seven drill spacings in section view (*left*). Standardized PTCs of same scenario and drill spacings trimmed at $p95$-$p05$ (*right*)

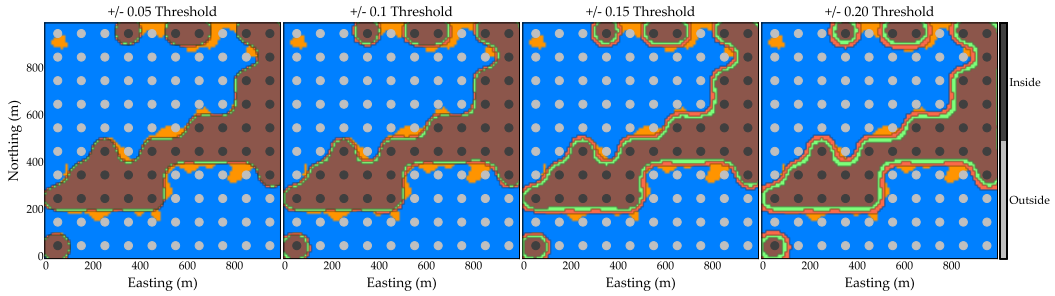
The delta values in the section view are considerably less than in the plan view for wider spacings. The lower uncertainty is attributable to an increase in conditioning

data down the drill holes. The tighter drill hole spacings have similar delta values; however, with increasing drill spacing, the delta values diverge.

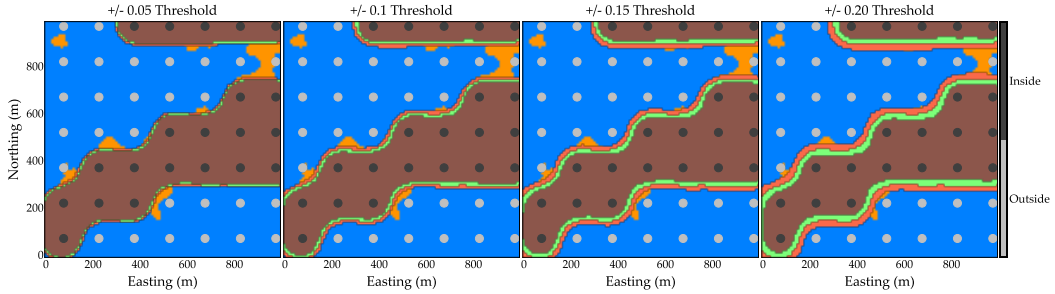
4.3 Uncertainty Thresholds

The delta values threshold the indicator estimates above and below the NN thresholded model. Eroded and dilated cases, in conjunction with the NN threshold model, form the final boundary model with uncertainty. The delta values over all scenarios are generally between 0.05-0.20 in section and plan view. To understand the effect of increasing delta values, the indicator threshold workflow, including uncertainty assessment, is visually inspected.

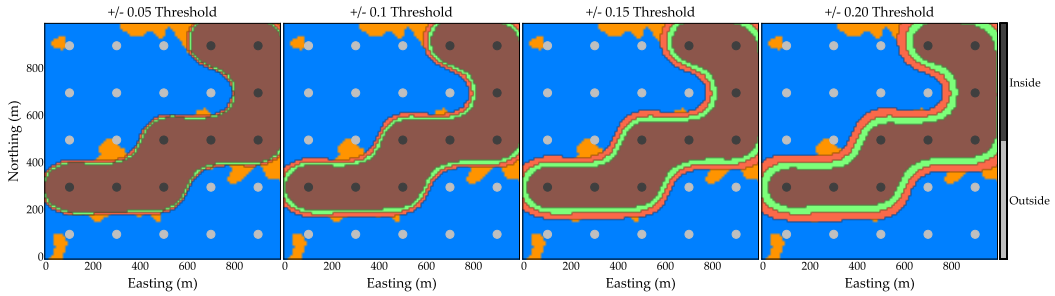
Figure 4.23 shows varying drill spacings, and delta values used to threshold for boundary uncertainty. The 10, 25, & 50m drill spacing models are tight at resolution with indiscernible uncertainty bandwidths and thus not shown. Uncertainty bandwidths grow with increasing delta values and drill spacing. A delta value of 0.05 corresponds to a narrow bandwidth with unrealistic uncertainty. The remaining threshold models appear realistic, given the conditioning data. The gridded nature of the drilling results in a relatively uniform bandwidth; however, noticeable increases in uncertainty are evident at estimation locations between diagonal inside data. The 250m spacing models exemplify this as the eroded cases are disjointed, the base cases thinly center between the data, and the dilated cases expand to give reasonable uncertainty. The methodology's adherence to the spatial configuration of the conditioning data results in reasonable bandwidths for uncertainty.



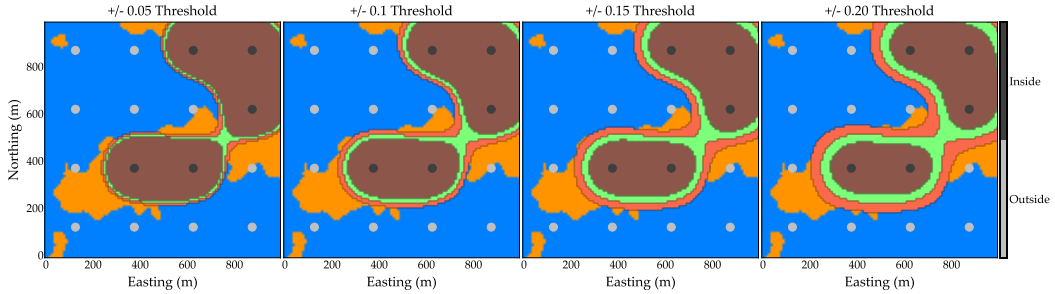
(a) Threshold models for 100m gridded drill spacing models showing uncertainty bandwidths



(b) Threshold models for 150m gridded drill spacing models showing uncertainty bandwidths



(c) Threshold models for 200m gridded drill spacing models showing uncertainty bandwidths



(d) Threshold models for 250m gridded drill spacing models showing uncertainty bandwidths

Figure 4.23: Boundary models with uncertainty bandwidths from threshold values of +/- 0.05, 0.1, 0.15,& 0.2. Eroded models in *brown*, NN thresholded base case models in *green*, and dilated models in *red*. The plan view models overlie the corresponding Truth (*orange*) and underly the conditioning data informing the thresholded indicator estimates.

The volumes of the eroded and dilated boundaries as a percentage of the underlying

Truth volume are in Tables 4.1 & 4.2. The percentages for 10m spacing hold constant in two instances showing the minimal uncertainty between higher thresholds at tight spacings.

Table 4.1: Eroded boundary threshold volumes as percentage of true volume for different plan view drill spacings

Spacing (m)	Threshold Volumes (%)			
	0.05	**0.10**	**0.15**	**0.20**
10	-1.2	-5.0	-5.2	-5.2
25	-0.1	-1.0	-1.7	-2.6
50	-4.3	-6.8	-8.1	-9.8
100	-8.2	-11.4	-17.1	-22.1
150	19.0	12.9	7.7	0.9
200	3.9	-2.3	-9.2	-17.0
250	-12.1	-19.2	-25.7	-32.8

Table 4.2: Dilated boundary threshold volumes as percentage of true volume for different plan view drill spacings

Spacing (m)	Threshold Volumes (%)			
	0.05	**0.10**	**0.15**	**0.20**
10	4.4	4.6	4.6	4.8
25	3.1	3.8	4.2	5.7
50	4.4	5.6	7.4	9.7
100	1.0	4.6	10.4	14.3
150	29.8	35.9	41.4	47.9
200	16.1	22.2	28.7	34.8
250	3.3	10.3	17.4	24.7

The 150m spacing eroded model is larger than the Truth, resulting from having the highest percentage of inside conditioning data. Moreover, the conditioning data configuration is proximal to the true boundary leading to the estimate expanding beyond (Figure 4.24).

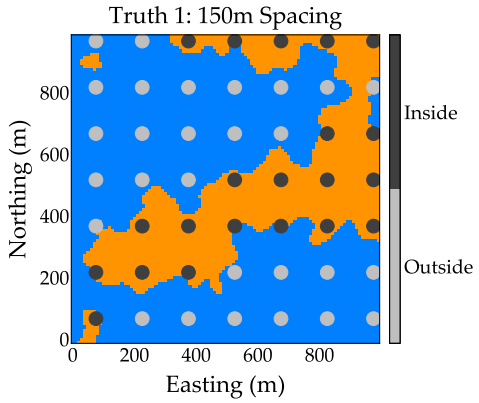
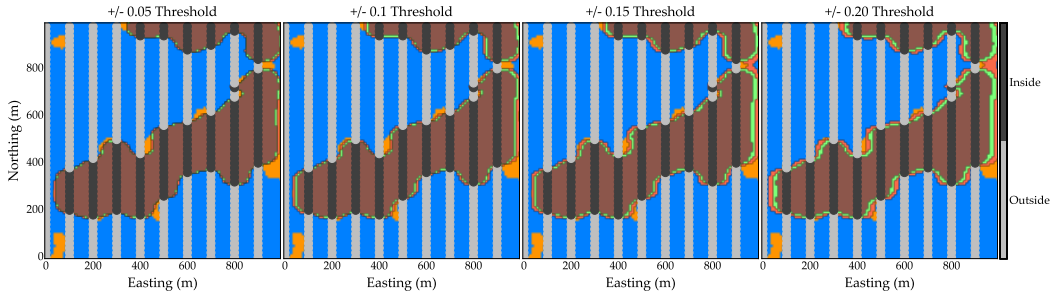


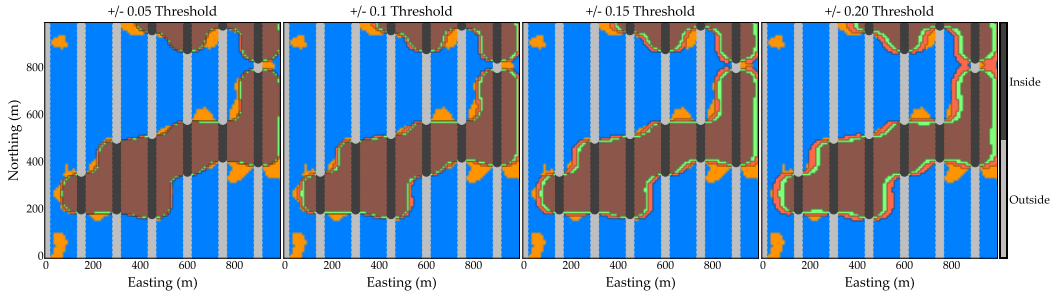
Figure 4.24: Truth realization 1 with 150m drill spacing

For section view spacing, replicating downhole drilling, the models are informed by more data. Figure 4.25 shows the same Truth, but with section drilling instead of

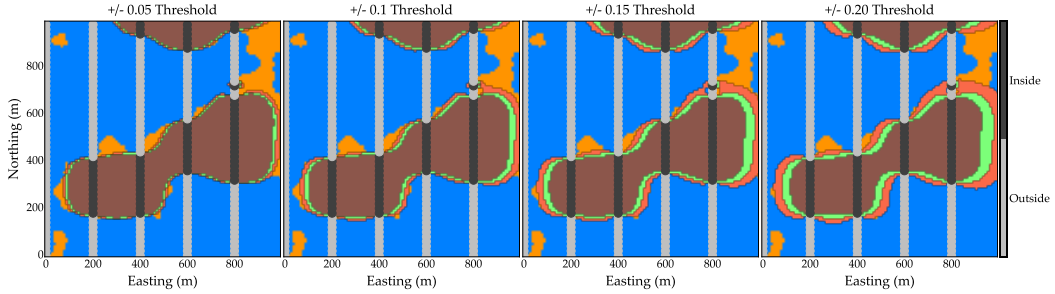
plan view sampling.



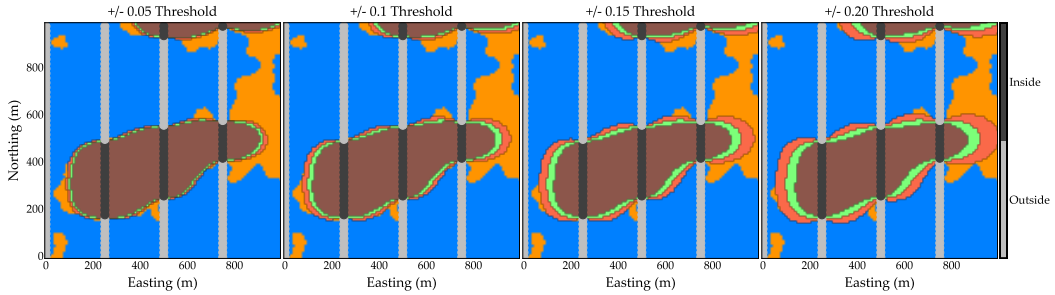
(a) Threshold models for section view 100m drill spacing models showing uncertainty bandwidths



(b) Threshold models for section view 150m drill spacing models showing uncertainty bandwidths



(c) Threshold models for section view 200m drill spacing models showing uncertainty bandwidths



(d) Threshold models for section view 250m drill spacing models showing uncertainty bandwidths

Figure 4.25: Boundary models with uncertainty bandwidths from threshold values of +/- 0.05, 0.1, 0.15,& 0.2. Eroded models in *brown*, NN thresholded base case models in *green*, and dilated models in *red*. The section view models overlie the corresponding Truth (*orange*) and underly the conditioning drillholes informing the thresholded indicator estimates.

The section view drilling shows two disparate sample densities in orthogonal directions. The data is collected densely downhole with sparse data existing along the strike of the collars. The asymmetry is accounted for in the uncertainty thresholding as the bandwidths vary smoothly and honour local data configurations. The most prominent example of this is the +/-0.20 threshold for 250m seen in Figure 4.25d. The uncertainty bandwidth expands between drill holes when distal to informing data and contracts proximal to the drill holes. Where the boundaries near the model edges, and away from conditioning data, expansion into the zone of higher uncertainty occurs. The fluctuating uncertainty bandwidth contrasts with SDF modeling where the additive C-parameter for uncertainty results in relatively constant bandwidths (Mancell & Deutsch, 2019).

Table 4.3: Eroded boundary threshold volumes as percentage of true volume for different section view spacings

Spacing (m)	**Threshold Volumes (%)**			
	0.05	**0.10**	**0.15**	**0.20**
10	-6.5	-6.5	-6.5	-6.5
25	-2.0	-3.5	-6.3	-8.9
50	-0.4	-4.4	-6.3	-9.2
100	-0.4	-4.8	-8.0	-11.0
150	-11.4	-16.5	-20.1	-23.8
200	-13.1	-18.8	-23.6	-28.0
250	-30.3	-34.9	-39.0	-42.6

Table 4.4: Dilated boundary threshold volumes as percentage of true volume for different section view spacings

Spacing (m)	**Threshold Volumes (%)**			
	0.05	**0.10**	**0.15**	**0.20**
10	4.6	4.6	4.6	4.6
25	2.9	4.7	6.5	7.5
50	4.0	8.4	10.1	11.3
100	6.5	9.9	12.7	15.6
150	-2.5	0.4	4.1	8.9
200	-2.2	3.7	9.3	15.9
250	-21.1	-16.1	-9.7	-2.5

The section view volume difference of the models to the Truth as a percentage is shown in Table 4.3. The 10m spacing holds constant bandwidths across all thresholds, a function of closely spaced samples resulting in static uncertainty. The 250m dilated boundary volumes are all below the true volume of the model. The wider spacing misses the north-south structure to the east of the model, leaving a significant amount of geology unsampled. Across the entirety of the models, the correct thresholding for uncertainty coincides with a value between 0.1-0.2.

4.4 Volume Uncertainty

Comparing volume uncertainty among many geological scenarios, varying thresholding values, over thousands of Truths, at multiple spacings in plan and section view is difficult. An avenue to aid in differentiating uncertainty bandwidths is to standardize the global volumes. For each thresholded indicator model with uncertainty, the model volume is standardized by subtracting and dividing by the corresponding true volume. A value of zero implies that the model volume matches the Truth volume exactly. A negative value indicates that the model is smaller, while a positive value means the model is larger than the underlying true volume. The plan view n=6000 standardized volumes over 1500 geological scenarios with four sample spacings are in Figure 4.26.

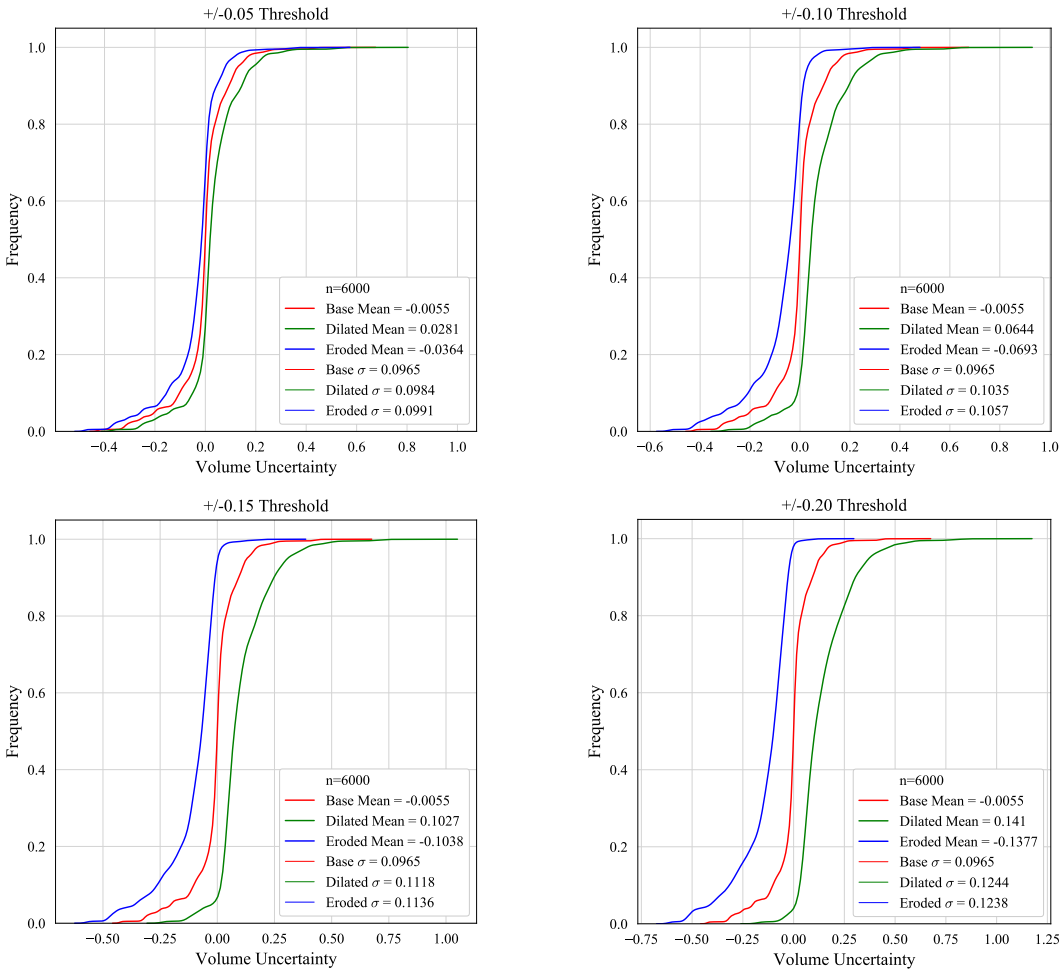


Figure 4.26: Volume Uncertainty from standardized boundary models for varying thresholds in plan view

The base case volume uncertainties are all equal with the same mean and standard deviation. Recall the NN indicator threshold model forms the base boundary model, and uncertainty is thresholded above and below to extract dilated and eroded boundaries. Therefore, the base case volume uncertainty is static across the different thresholds. The base case mean is near zero, indicating the NN thresholding is overall closely predicting the true underlying volume. The dilated boundary volume means increase as the threshold value increases, shifting the curves to the right. The opposite occurs with the eroded boundary volumes; as the threshold value increases, the curve shifts left, and volumes decrease. The standard deviations are similar between dilated and eroded distributions of a particular threshold; however, as threshold values increase,

the respective standard deviations are elevated. Figure 4.27 shows the standardized volume uncertainty for the same underlying Truths with section view drilling.

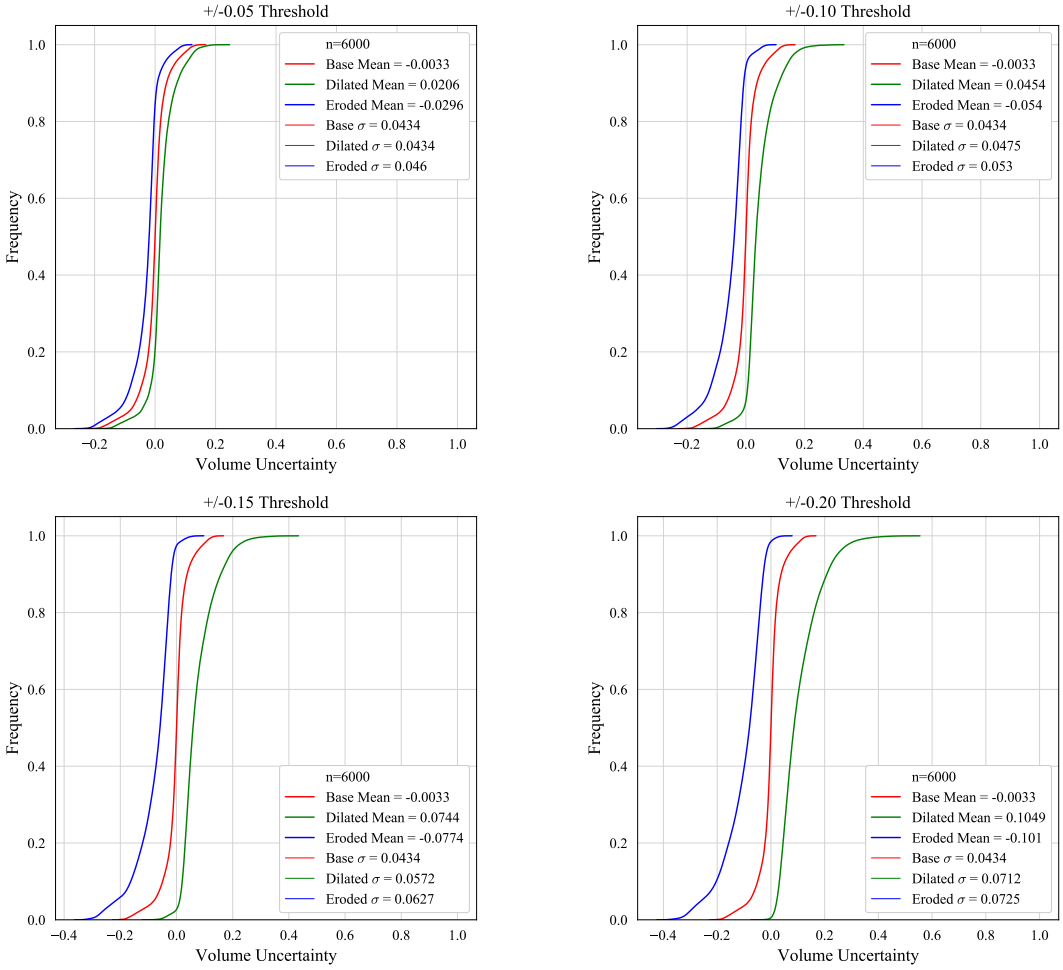


Figure 4.27: Volume Uncertainty from standardized boundary models for varying thresholds in section view

The section view distributions show the same trends and patterns as the plan views, however, with smaller mean values and standard deviations overall. The increase in sampling intuitively leads to smaller uncertainty. For eroded distributions, the upper tails transition more abruptly than the lower tails. This occurs as the models are exceedingly smaller than the Truths. The inverse is valid for the dilated distributions – the shift by the threshold value results in the lower tail transitioning faster than the upper.

The inference of thresholding values for parameterizing probability-threshold curves and assessing uncertainty is based on extensive synthetic dataset modeling. Verification of the workflow efficacy is accomplished by comparing results to the true underlying model. A thresholding value of +/-0.15 above and below the NN indicator threshold model is recommended. The bandwidth between resulting dilated and eroded boundaries gives realistic uncertainty that honours the spatial configuration of the conditioning data. For a modeler wanting a tighter, more conservative bandwidth, a threshold value of +/-0.1 may be appropriate. Wider bandwidths can be managed using a higher threshold of +/-0.2 to increase the probability of the boundary falling within the uncertainty zone.

4.5 Local Uncertainty

Local uncertainty quantification is accomplished through indicator estimation. The indicator interpolant framework defines the uncertainty distribution at a predicted node. Therefore, the field of probabilities mapped out by the interpolant gives direct access to local uncertainty. The local conditional CDF is a measure of uncertainty for the location, conditional to the surrounding data. A visual check of local uncertainty is possible from the interpolated probability field. Cross validation workflows, such as K-fold, leave subsets of data out of the modeling workflow that are then reconciled to measure if the model is accurate and precise. If the models are accurate and precise, then the local uncertainty prediction from the indicator estimation should be reasonable.

4.5.1 Accuracy Plots

Cross validation is effective for tuning modeling parameters and to compare algorithms (J. Deutsch & Deutsch, 2012). Probability intervals are binned in accuracy plots and the probabilities in each bin are averaged and compared to the actual fraction of each category. The goal is to have average predicted probabilities near the true fraction (J. Deutsch & Deutsch, 2012).

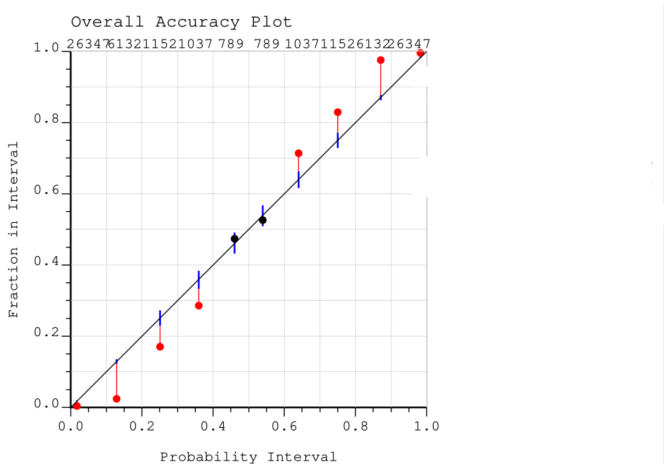


Figure 4.28: Accuracy plot for indicator model tested against the underlying Truth

The 1:1 line signifies that the average estimate probabilities match the fraction in that interval from the underlying Truth. The numbers at the top are the numbers of nodes in each bin and the red and black dots are the average probability in the bin. The methodology uses a bootstrap like approach for Probability Interval (PI). For each PI, or bin, the probability values are defined from the estimate. A random number is then generated and if the number is lower than the probabilty value then a value of one is assigned, otherwise zero is assigned. The fraction of random indicators are calculated for 1000 iterations. The 5% and 95% quantiles are extracted to arrive at the inside 90% probability interval (J. Deutsch & Deutsch, 2012). The global interpolator to estimate results is conditionally unbiased for the variables, but will not be independent.

Ten thousand accuracy plots from indicator estimates and their underlying Truths are difficult to manage. Therefore, a simplifying modification is implemented. The accuracy is parameterized by weighting the data, preserving the sign, and taking the fraction in the interval subtract the probability mean for all bins. The formulation is seen in Equation 4.2.

$$A = \frac{\sum_{bins_{i=1}}^{N} num\ bin_i \cdot (fraction_i - prob\ mean_i)}{\sum_{bins_{i=1}}^{N} num\ bin_i} \tag{4.2}$$

The accuracy, A, is the sum over the N-bins of the number in the bin multiplied by the difference of the fraction in that bin, $fraction_i$, from the mean, $prob\ mean_i$ standardized to the number in the bin, $\sum_{bins_{i=1}}^{N} num\ bin_i$. The result is a single accuracy value per model representing the average distance from the red and black dots to the 1:1 line in Figure 4.28. Over the ten thousand models the accuracy values build a distribution as in Figure 4.29.

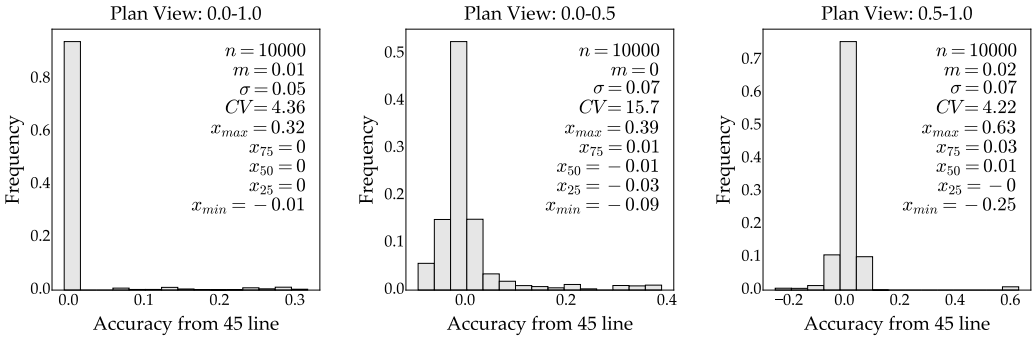
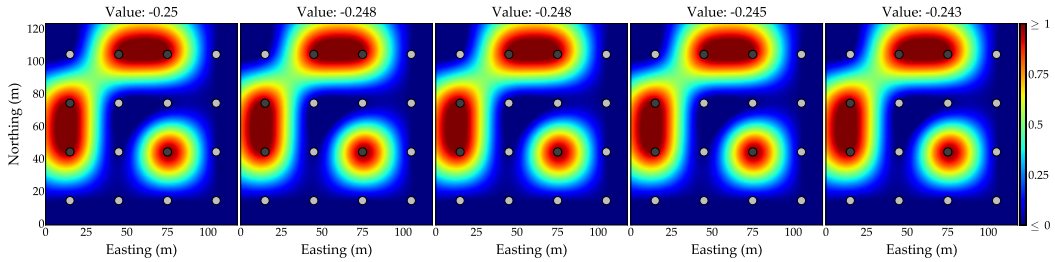


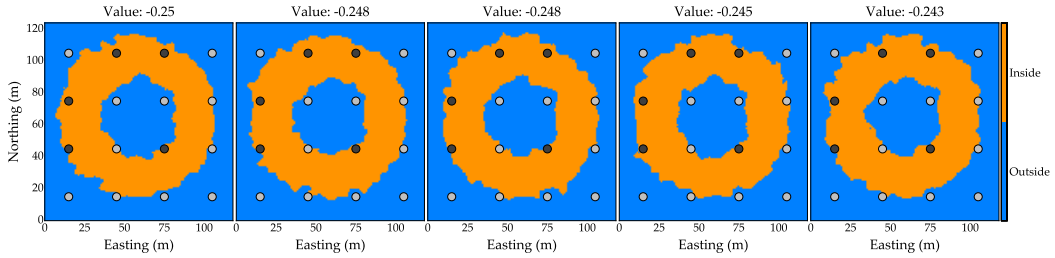
Figure 4.29: Distribution of accuracy values from 10,000 models. Over all probability intervals (*left*), from 0 to 0.5 probabilty intervals (*middle*), and from 0.5-1.0 probability intervals (*right*)

The values over all probability intervals average to 0.01; however, the first half of the PIs are lower than the second half on average. The values indicate that the fractions in the interval are near the 1:1 line, with the first half being on or slightly below and second half being above. Therefore, higher probability of high values being estimated and lower probability of low values being estimated showing minor systematic bias.

Checking outliers from the distributions gives insight into the metric. The worst accuracy values –the lowest and highest values– are sorted and the underlying Truths and models for the most negative values are in Figure 4.30.



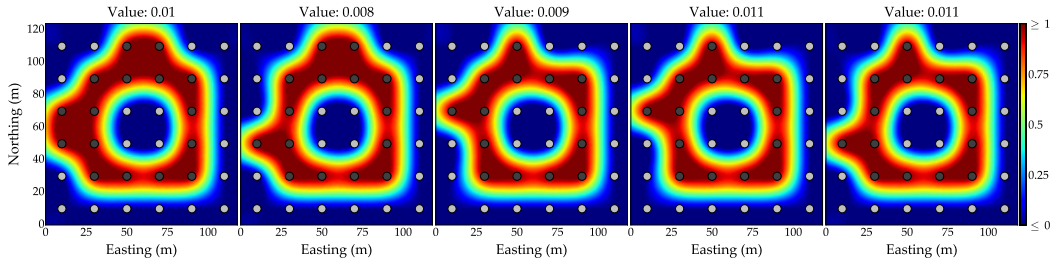
(a) Lowest Accuracy values corresponding models



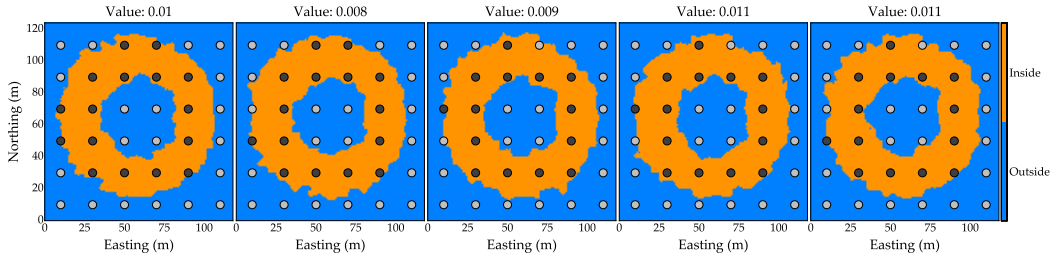
(b) Lowest Accuracy values and respective underlying Truths

Figure 4.30: Lowest Accuracy values, all from same scenario, and from the 0.5-1 distribution

The modeling is difficult for complex structures as above. The drill holes nearly miss on the eastern and southern extents leading to divergence between model and underlying Truth. Figure 4.30a shows five models that are identical as the drilling intersected the same geology. Therefore, the difference in accuracy values is from the slight variations at a finer scale in the underlying Truths as seen in Figure 4.30b. In contrast, the 20m spacing for the same scenarios illustrates the accuracy change by simple tightening of spacing in Figure 4.31.



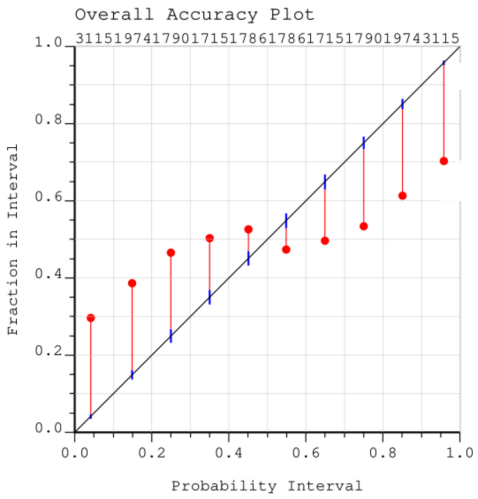
(a) Accuracy values corresponding models



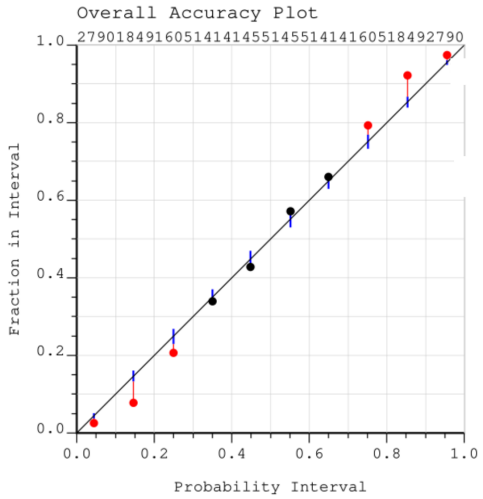
(b) Accuracy values underlying Truths

Figure 4.31: Accuracy values showing change in drill spacing and improved modeling

The values for the 20m spacing are near zero indicating proximity to 1:1 and probabilities matching the fraction in interval. In the 20m instance, the drill holes increase the data informing the model increasing the accuracy values. With the accuracy values near zero, the model for uncertainty generated from the indicator interpolation is deemed reasonable. The underlying accuracy plots for the lowest accuracy in the 30m spacing and corresponding 20m spacing near zero are in Figure 4.32.



(a) Accuracy plot for lowest accuracy value (-0.25) from model 73 in 30m spacing



(b) Accuracy plot for model 73 in 20m spacing with corresponding accuracy value of 0.01

Figure 4.32: Accuracy plots for lowest recorded accuracy value (*top*) and same realization drill at 20m spacing showing better accuracy (*bottom*)

The 30m plot shows high discrepancies from the 1:1 line reflecting poor accuracy in the model. In contrast, the 20m spacing shows the black and blue dots near the 1:1 line indicating the additional data informing the model is significantly increasing the accuracy of the indicator estimate and, therefore, the local uncertainty.

4.6 2-D Uncertainty Assessment

Competing uncertainty workflows are implemented for the tabular vein data from Chapter 3. The base case model is not modified. The SDF C-parameter is 15m, which is applied to the modified distance function. The C-value is added to positive and subtracted from negative DF values. The modified SDF values are interpolated across the domain. The isosurface is used as the boundary with the -15m boundary and the 15m boundary forming the eroded and dilated cases. The indicator threshold model is defined from the NN volume ratio, while eroded and dilated boundaries are extracted by taking +0.15 and -0.15 from the base threshold. The uncertainty models for the competing workflows are in Figure 4.33.

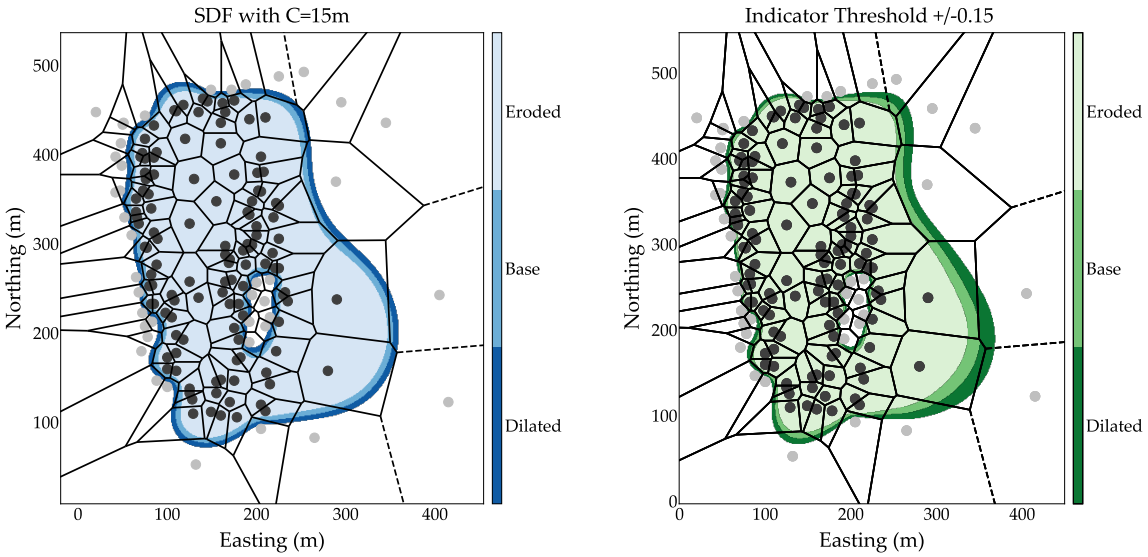


Figure 4.33: SDF uncertainty model using RBF framework and C-parameter uncertainty value of 15m (*left*). Indicator threshold uncertainty model using RBF framwork and +/- 0.15 uncertainty thresholds (*right*)

The SDF bandwidth of uncertainty is conservative in the vein's lower right due to data structure asymmetry (Mancell & Deutsch, 2019). The constant C-parameter simplicity in the vein's upper-right contact is also evident– a consequence of the global additive nature of the parameter and its lack of information derived from local data structure (Mancell & Deutsch, 2019). In contrast, the indicator approach

shows a fluctuating uncertainty bandwidth conditioned to the surrounding data. In areas where data is sparse, the bandwidth expands; in zones of tight spacing, the uncertainty contracts. An issue arising in boundary modeling is edge effects from over extrapolation. The indicator threshold eroded boundary is 9.5% conservative to the base case NN threshold model. The dilated boundary model for the indicator approach is 12.1% larger than the base case giving a total volumetric uncertainty of 21.6%. In contrast, the SDF model for uncertainty is conservative. The base case is 6.6% smaller than the NN model volume. The dilated SDF boundary is 5.1% larger, while the eroded case boundary is 18.6% conservative for a total volumetric uncertainty of 23.7%.

K-fold analysis shows the efficacy of the modeling approach. The SDF overall five folds has an error rate 71% that of the indicator approach at 0.0625 and 0.0875, respectively. On a per fold basis the indicator approach bests or equals the SDF error values 60% of the time. For the 20 fold analysis, the SDF overall error rate is 83% of the indicator threshold workflows. The indicator approach bests or equals the SDF modeling on a per fold basis 85% of the time.

4.7 Results and Considerations

Uncertainty quantification for boundary modeling is critical to understanding the underlying geological structure and to make informed decisions. The use of indicators to interpolate probabilities and assess uncertainty is straightforward. Indicator estimates predict the conditional distribution of uncertainty at unsampled locations. Cross validation techniques test the accuracy of the estimates in order to ensure reasonable local uncertainty. Geometric uncertainty is captured through the deposit model type. Global volumetric uncertainty is understood through the relationship between probabilities and thresholds illustrated by the use of experimental PTCs from synthetic data. Thresholding of indicator estimates 0.15 above and below the NN model threshold gives access to eroded and dilated boundaries. A conservative uncertainty bandwidth coincides with a +/-0.1 threshold value with the NN

model threshold forming the base case. A liberal bandwidth, capturing additional uncertainty, derives from a threshold value of +/-0.2. These boundaries form an uncertainty bandwidth that honours the conditioning data structure. The uncertainty bandwidth expands and contracts relative to the sample density. The final boundary model consists of a base, eroded, and dilated case that captures global volumetric uncertainty while honouring local data configurations.

Chapter 5

Implementation Details

The implementation of the proposed indicator thresholding approach is straightforward. The construction of a NN model and calculation of global volume give a ratio for inside to outside volume to threshold the indicator estimate. For uncertainty, the estimate is thresholded +/-0.15 of the base case z-value to extract a zone of uncertainty between dilated and eroded boundaries. The choice of interpolator and subsequent parameterization is essential. Edge effects from extrapolation to model extents may occur must be managed. In a multi categorical dataset, the independently modeled data will often have overlap at estimated nodes. For determining the dominant category at a given node, a selection algorithm is used to obtain a single final categorical model that can be further thresholded for uncertainty.

5.1 Interpolation

Indicator interpolation is a critical aspect of the proposed workflow. Often in boundary modeling, global methods of interpolation are desired. Global kriging and RBF interpolation use all conditioning data to inform the estimate. The algorithms are fast and artifacts are minimized (Carvalho, 2018). The speed of global dual kriging is based on the weights not depending on the location being estimated; therefore, the equations are solved only once. The results are artifact-free because no search for data is needed –all conditioning data informs each estimate. A disadvantage to global methods is for datasets with N>30,000; the computational power is strenuous due to the dense N X N matrix. One solution is to use Ordinary Kriging with a restricted search (C. Deutsch & Journel, 1998). Moreover, for RBFs, sparse iterative or direct solvers may assist in streamlining the CPU time. The sparse iterative solvers utilize the fact that perfect precision is not necessary for the computations allowing for pre-

determined accuracies to be set, leading to decreased computational time (Carr et al., 2001). Sparse direct methods consider that the data beyond a specific distance has a negligible effect on the estimation(Martin, 2019). Domain partitioning into smaller subsets of overlapping data and solved independently within the subsets may also be utilized for more massive datasets (Martin, 2019).

5.2 Edge Effects

Uncontrolled extrapolation of estimated domains to the extents of the model is a common problem in boundary modeling. These edge effects are a result of the influence of samples stretching into areas of low-density drilling. The spatial structure of the conditioning data has a fundamental control on edge effects. Mitigation of edge effects is generally done by introducing controlling samples, restricting the variogram range, or applying the distance function in a hierarchical manner (D. Silva, 2015). All of these measures make the boundary modeling algorithm more complex. Controlling for edge effects requires the modeler to know the geological conceptual model (D. Silva, 2015). Leaving edge effects in a boundary model may introduce bias; however, fixing the problem can often lead to artifacts, the introduction of other biases, and the potential for the further degradation of the model. The weights associated with the local informing data are the cause of over extrapolation. When using Global Ordingary kriging, the estimations at locations distal to conditioning data are relatively constant near the data's mean. Because the distal locations are relatively constant in predicted probability, if the NN threshold value is below the estimated probability at the location, it will be considered inside. The problem is more pronounced in dilated models whereby the thresholding value is -0.15 of the NN threshold. Figure 5.1 shows the thresholding issue where the dilated boundary extrapolates near model edges where there are no conditioning data.

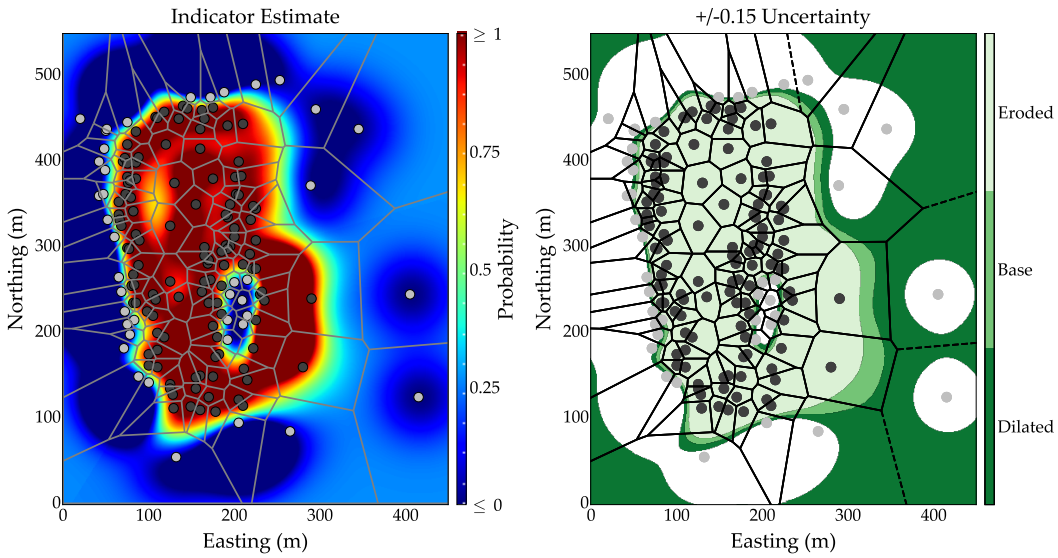
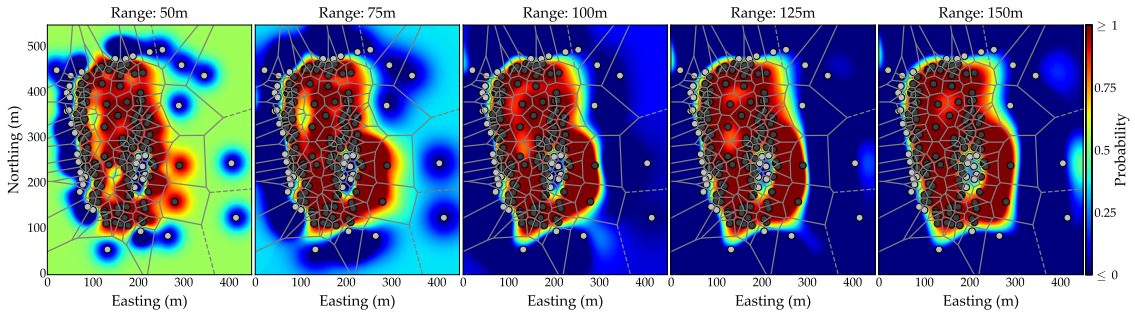
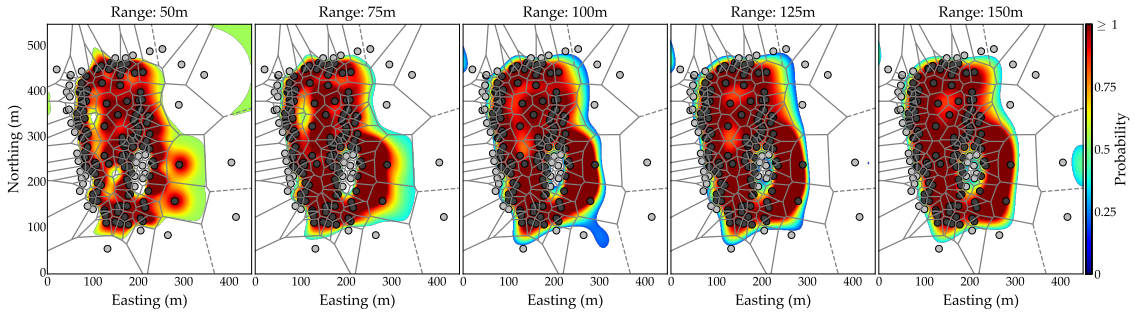


Figure 5.1: Global kriged indicator estimate using Gaussian variogram showing constant probabilities distal to conditioning data (*left*). Global kriged and NN thresholded model with +/-0.15 uncertainty bandwidth (*right*)

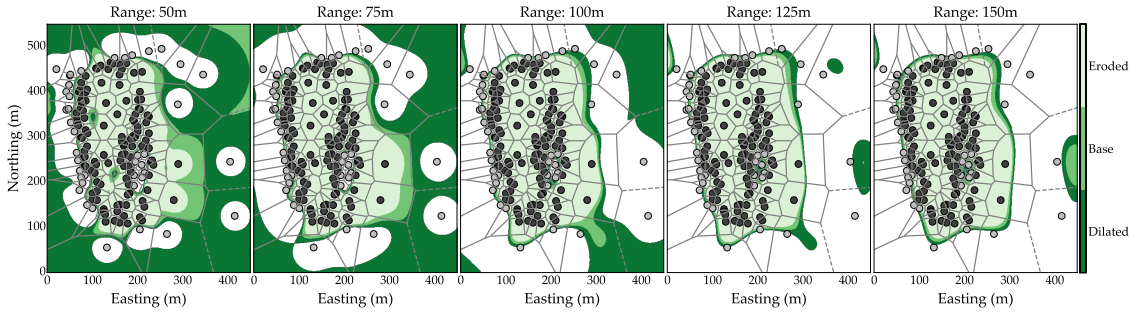
Typical avenues for properly handling edge effects is to restrict the variogram range and add control points. The effects of changing the variogram range are seen in Figure 5.2. The variogram range does not mitigate the extrapolation issues associated with the dilated boundary model (Figure 5.2c)



(a) Global Ordinary kriging using Gaussian variograms with different ranges



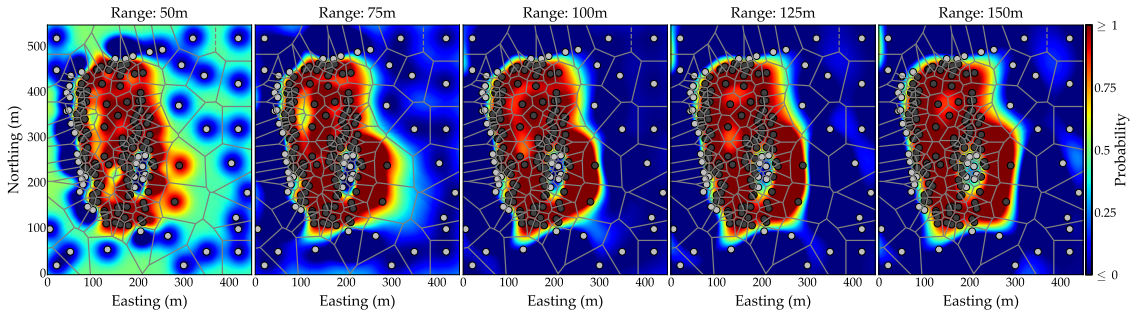
(b) Global Ordinary kriging using Gaussian variograms with different ranges thresholded to NN model



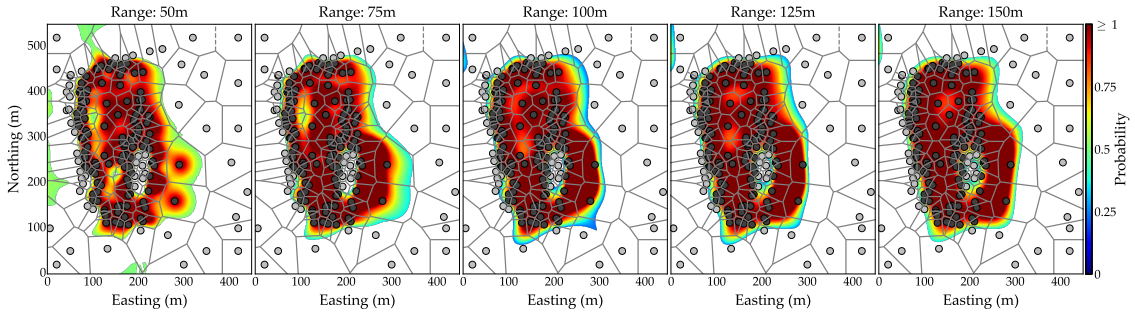
(c) Global Ordinary kriging using Gaussian variograms with different ranges thresholded to NN model with +/-0.15 uncertainty thresholds

Figure 5.2: Boundary models using Gaussian variograms with uncertainty bandwidths from threshold values of +/- 0.15 showing issue with over extrapolation and edge effects from dilated models

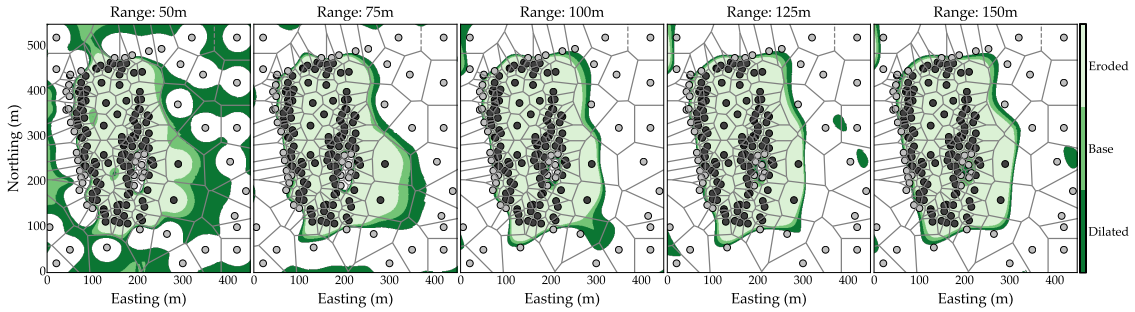
The effects of adding controlling points are seen in Figure 5.3. The Gaussian variogram with range 75m mitigates most of the over extrapolation issues; however, there are additional artifacts.



(a) Global Ordinary kriging using control points and Gaussian variograms with different ranges



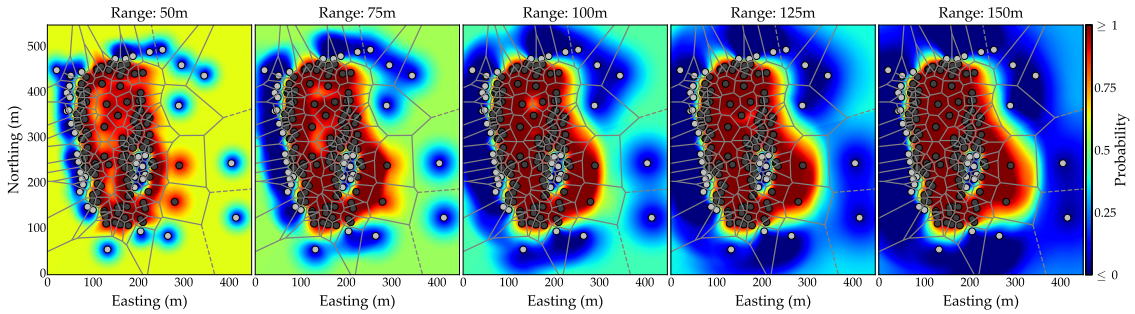
(b) Global Ordinary kriging using control points and Gaussian variograms with different ranges thresholded to NN model



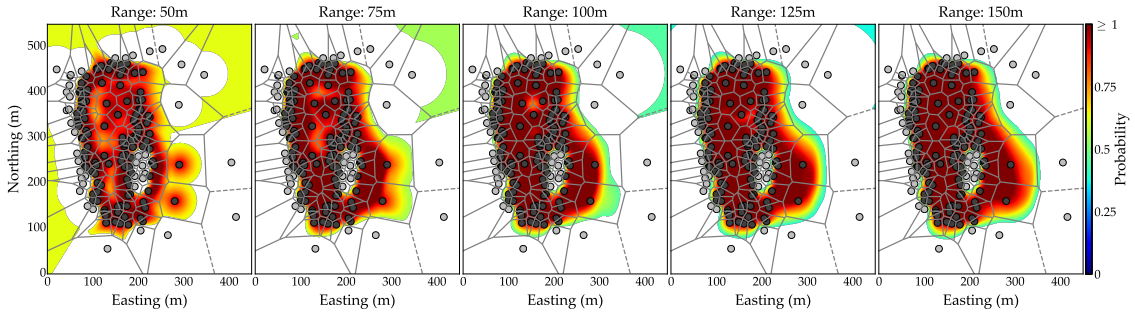
(c) Global Ordinary kriging using controlling points and Gaussian variograms with different ranges thresholded to NN model with +/-0.15 uncertainty thresholds

Figure 5.3: Boundary models using Gaussian variograms with uncertainty bandwidths from threshold values of +/- 0.15 showing issue with over extrapolation and edge effects from dilated models

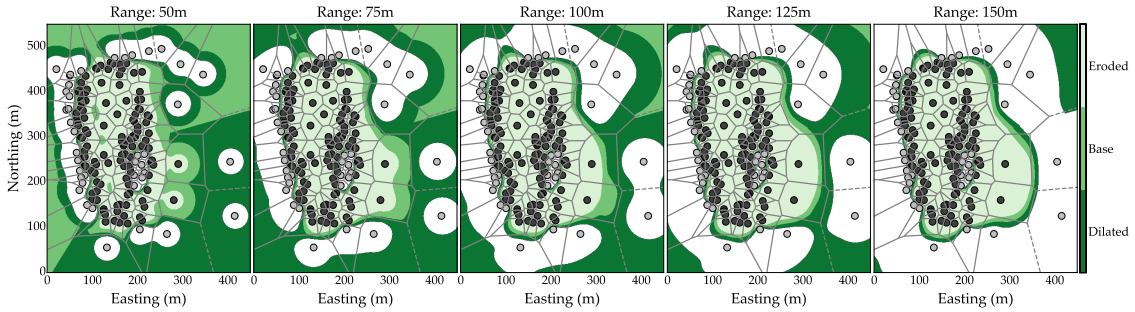
The results with a spherical variogram are shown in Figure 5.4. The uncertainty model for spherical variogram with range 150m in Figure 5.4c shows a better behaved boundary model.



(a) Global Ordinary kriging using spherical variograms with different ranges



(b) Global Ordinary kriging using spherical variograms with different ranges thresholded to NN model



(c) Global Ordinary kriging using Spherical variograms with different ranges thresholded to NN model with +/-0.15 uncertainty thresholds

Figure 5.4: Boundary models with spherical variograms with uncertainty bandwidths from threshold values of +/- 0.15 showing issue with over extrapolation and edge effects from dilated models

In contrast, the RBF framework fits a function to the conditioning data. The equivalent RBF model with uncertainty for the above scenario is in Figure 5.5. The RBF model appears geologically reasonable and does not show extrapolation issues.

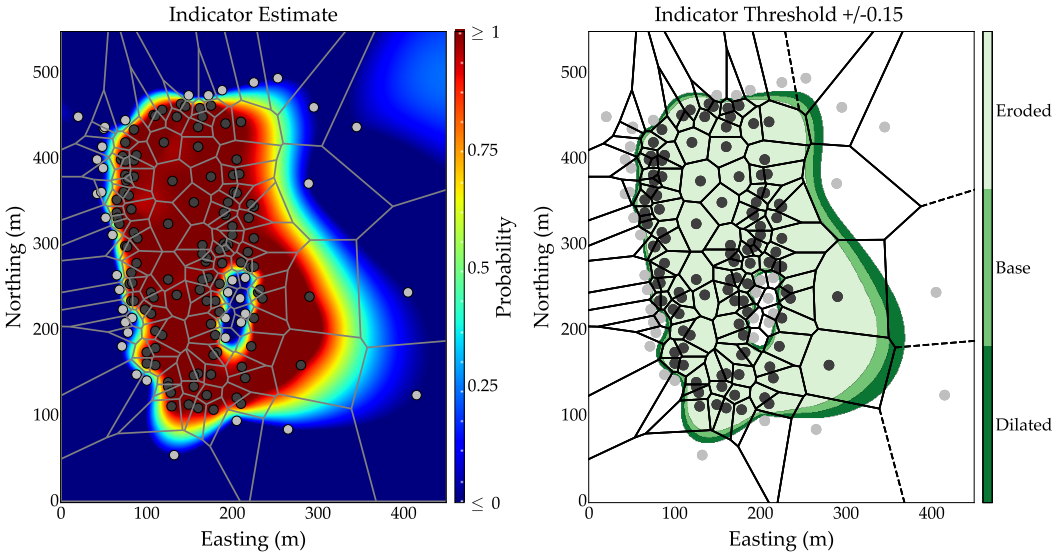


Figure 5.5: RBF indicator estimate showing constant probabilities distal to conditioning data (*left*). RBF and NN thresholded model with +/-0.15 uncertainty bandwidth (*right*)

The indicator thresholding approach introduces another control for extrapolation. A modeler with site-specific knowledge can restrict edge effects by limiting the global volume used for thresholding the indicator estimate by altering the maximum distance parameter for search radii in the NN modeling. Multiple NN models are generated with varying maximum distance parameters, then a realistic NN model is selected based on knowledge of the deposit. The NN volume ratio is applied to the indicator estimate. A NN model is generated by using Ordinary Kriging with one data. Ordinary kriging restricts the sum of the weights to one, so the reuslt is a NN model. Inverse distance with a high power will also result in a NN model. The closest sample receives the only weight, thus making it equivalent to the nearest neighbour estimation (Rossi & Deutsch, 2014). Visual inspection of the thresholded indicator model is essential to confirm the model adheres to known geological characteristics. Table 5.1 shows maximum distance parameters and resulting model volumes which is plotted in Figure 5.6.

Table 5.1: Maximum distance and NN volumes

Max Distance (m)	Volume (blocks)
50	11001
100	27811
200	40145
300	46990
500	54068
1000	55898
1500	55898
2000	55898
3000	58898

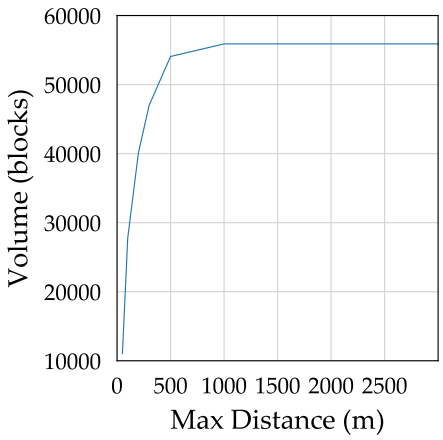


Figure 5.6: NN volumes as a function of maximum distance radii

The volumes from NN models stabilize at a maximum distance of 1000m. The 1000m distance roughly corresponds to the extents of the drilling area. Figure 5.7 shows the four different models with maximum distance parameterizations of 50, 300, 500, and 1000m. The 50m maximum distance model shows a restrictive model with no edge effects, and the volume corresponds to 11,001 blocks. The 500 and 1000m maximum distance models are stabilizing around 54-56,000 blocks. The 300m model is 46,990 blocks and is medial to the other models. A modeler with geological understanding of the deposit may choose the 300m maximum distance parameter to tune extrapolation issues. Figure 6.2 shows the indicator thresholded models for 50, 300, 500, and 1000m maximum distance parameters.

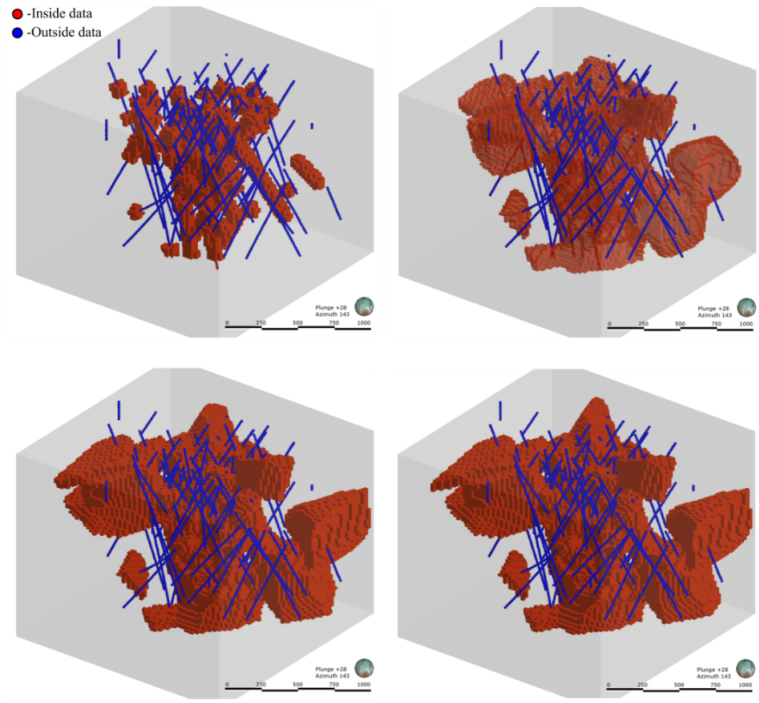


Figure 5.7: Nearest Neighbour models with maximum distance 50m (*top left*), 300m (*top right*), 500m (*bottom left*) & 1000m (*bottom right*)

.

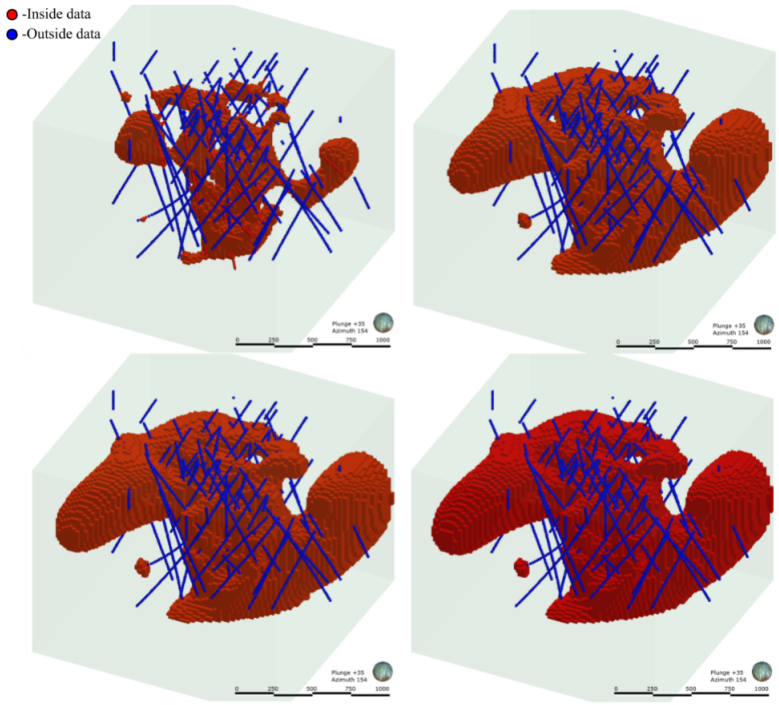


Figure 5.8: Indicator Threshold models with maximum distance 50m (*top left*), 300m (*top right*), 500m (*bottom left*) & 1000m (*bottom right*)

.

Extrapolation effects are present; however, the NN model volume approach reduces the extrapolation effects in a reasonable manner. NN volume control can help mitigate issues arising from extrapolation in conjunction with changing variogram ranges and introducing additional controlling data.

5.3 Multiple Categories

The indicator threshold workflow is adaptable to multicategory datasets. The process follows the same general steps as for single domain boundary modeling. The multi categorical dataset is split into $k = 1, ..., K$ datasets with each category having a conditioning dataset with inside values where that category was present and all other data are considered outside. The K datasets are independently indicator estimated, and corresponding NN models are constructed. The individual models are thresholded to their respective NN model volumes arriving at an indicator threshold model for

each category.

Overlap exists between the indicator estimate models, and they must be combined to a single final multicategory model. Nodes with multiple categories present select the category with the largest differential between the indicator estimation and its corresponding NN threshold to prevail at that location.

For all sampled locations $\{z(\mathbf{u}_\alpha), \alpha = 1, ..., n\}$, the categories are coded as:

$$i_k(\mathbf{u}_\alpha) = \begin{cases} 1, \text{ if } z(\mathbf{u}_\alpha) = 1 \\ 0, \text{otherwise} \end{cases} \quad \text{for } k = 1, ..., K \tag{5.1}$$

Each category is independently interpolated using an RBF framework. The indicator estimate for category k at (x, y, z) location $\mathbf{u}$ is the weighted linear combination of all conditioning data evaluated on a radial kernel:

$$i_k^*(\mathbf{u}) = \sum_{\alpha=1}^{N} \lambda_\alpha \phi(\left|i_k(\mathbf{u}) - i_k(\mathbf{u}_\alpha)\right|) \text{ for } k = 1, ..., K \tag{5.2}$$

And the selection algorithm for a given node chooses the highest differential between the estimated probability, $i_k^*(\mathbf{u})$, and the corresponding nearest neighbour z-value, $z_{nn;k}$:

$$i^*(\mathbf{u}) = k' \text{ such that } i_k^*(\mathbf{u}) = \mathbf{max}\left\{i_k^*(\mathbf{u}) - z_{nn;k}\right\}_{k=1}^{K} \tag{5.3}$$

The result is a single multi categorical model with no overlapping nodes.

A modeler with knowledge of the deposit could set geological precedence for the domains. Geochronology, cross-cutting relationships, and the laws of superposition can be incorporated into a site-specific selection method creating a realistic model that adheres to known geological attributes. The result is a single multi categorical boundary model. The individual domains' volumes do not precisely match their corresponding NN volumes due to the post-processing selection method mitigating

overlap.

Uncertainty in multi categorical boundary models is difficult due to overlap between adjacent dilated boundary models. Therefore, uncertainty assessment should be based on a domain of particular interest or domains that are distal to one another and do not come into contact. Once the domain for uncertainty is decided, thresholds 0.15 above and below the z-value are taken to further threshold the indicator estimate into dilated and eroded boundaries. The uncertainty bandwidths take precedence over surrounding domains, and the models combine to form a single multi categorical model with uncertainty. In tight drill spacing areas, or along contacts in drill holes, the base case or eroded boundary may be in contact with a foreign domain. For the eroded case, the uncertainty is within the size of the block.

Chapter 6
Single Domain Case Study

Data from a porphyry deposit serves to demonstrate the proposed indicator workflow. The drill hole samples contain no geographical locations or assays in order to preserve the anonymity of the project and there is no further geological information or interpretations given. The data has previously been studied by D. A. Silva and Deutsch (2015) in an SDF workflow paper and is useful for comparison purposes. The point data consists of five domains: an oxide, a sulphide, and three intrusions. For a single domain study, the three intrusions combine to form one and constitute the 'inside' data with the oxide and sulphide domains combining to form the 'outside' data. The indicator threshold workflow using the two indicators in a binary example shows robust and effective results. Furthermore, the data is passed through the SDF workflow in order to assess the differences between modeling methods. The results of the different workflows are discussed and the efficacy of the methodology explored.

6.1 Boundary Modeling

A single intrusion combining the three known intrusions from the porphyry drill hole dataset is used for a univariate boundary modeling case study. The coded data in Figure 6.1 shows the inside data in red and outside data in blue comprised of codes 4,5,6 and codes 2,3, respectively. There are 902 conditioning inside data out of a total of 3276 data. The data is within a domain 1750mN by 1750mE and 1425mZ with a block dimension of 25mX25mX15m, for a total of 465,500 blocks.

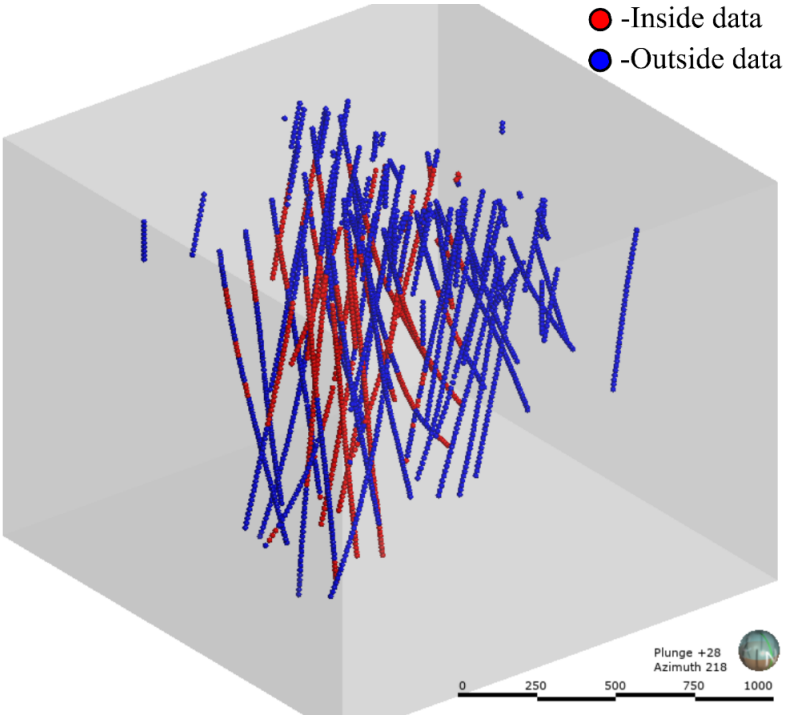


Figure 6.1: Drill data from porphyry dataset with single domain comprised of inside and outside data. Codes 4,5,6 are inside domain whilst Codes 2,3 are outside

.

6.1.1 Nearest Neighbour Model

Nearest neighbour models, as discussed in Chapter 3, give a spatial representation of the conditioning data. The summation of the NN model volumes for individual cells gives an unbiased global volume for a domain. Extrapolation issues arise near edges in nearest neighbour models in areas not bounded by 'outside' conditioning data. Therefore, nearest neighbour models with varying maximum distances are built in order to understand and control edge effects created from extrapolation. Different distances are input to the maximum search radii parameter in `KT3DN GSLIB` software and run. Ordinary kriging with a maximum one data informing results in a NN model. With only one data informing, and the sum of the weights having to equal one for Ordinary Kriging, the effective result is a NN model. Figure 6.2 shows the relationship of maximum search radii to NN model volumes. The chosen distance for the model is 1000m, whereby the curve flattens, and volumes stay constant at 55,898 blocks. A modeler with knowledge of the geologic depositional environment may

impart their expertise on the model and control for extrapolation effects by selecting a different NN-model.

Table 6.1: Maximum distance and NN volumes

Max Distance (m)	Volume (blocks)
50	11001
100	27811
200	40145
300	46990
500	54068
1000	55898
1500	55898
2000	55898
3000	55898

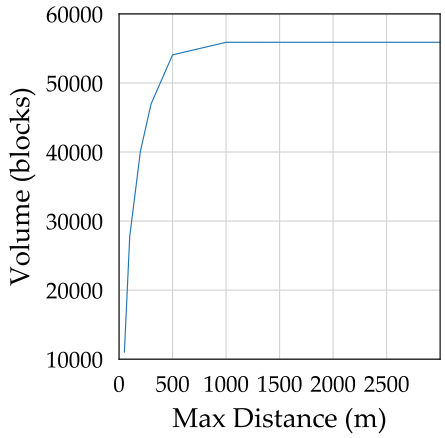


Figure 6.2: NN volumes as a function of maximum distance radii

A combination of different NN models from nine different radii distances can be seen in Figure 6.3 and global volumes in Table 6.1. The NN models in Figure 6.3 represent the unbiased geometric configuration of the conditioning data with different search radii. The model changes by 500% as distances increase from 50m to where they stabilize at 1000m.

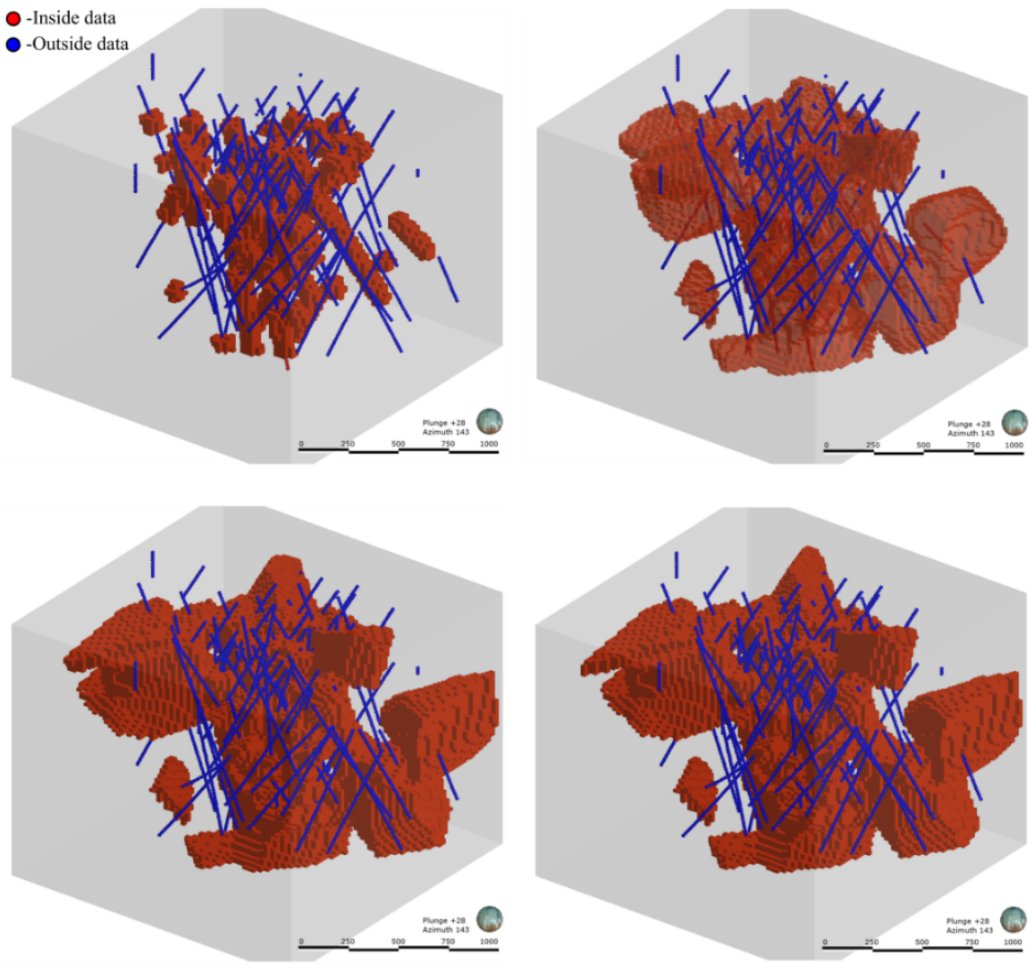


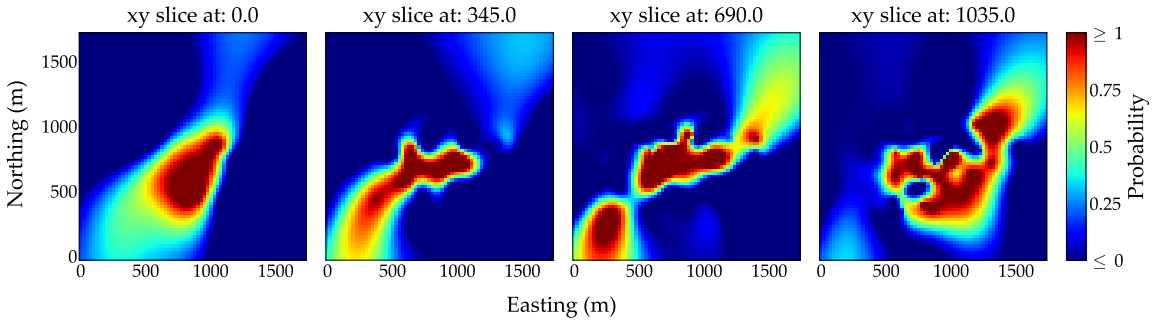
Figure 6.3: Nearest Neighbour models with maximum distance 50m (*top left*), 300m (*top right*), 500m (*bottom left*) & 1000m (*bottom right*)

.

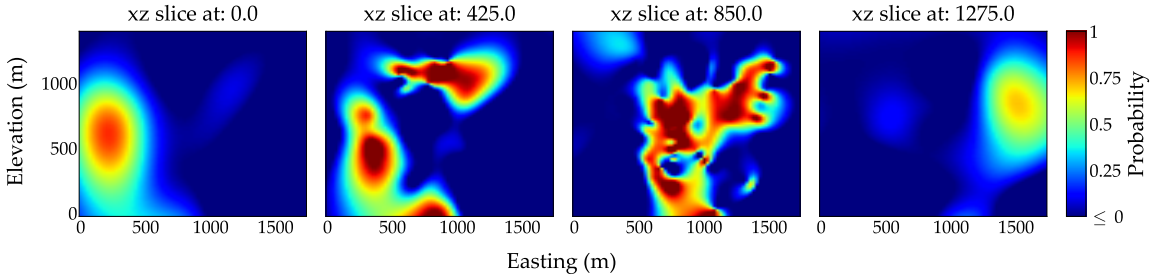
6.1.2 Indicator Estimation

Interpolation of the indicator values maps a field of probabilities from which a boundary interface can be extracted. As discussed in Section 5.1, the algorithm chosen will produce smoothly varying estimates to be thresholded. Global techniques, such as Global/Dual Kriging or RBFs, use all the conditioning data and result in smooth models that are artifact-free (Carvalho, 2018). Limitations of N<30,000 samples are a drawback of global estimators; however, the dataset has N=3276 samples. A common approach to dealing with larger datasets when using global estimators is to limit the estimator search neighbourhood in order to use less informing samples.

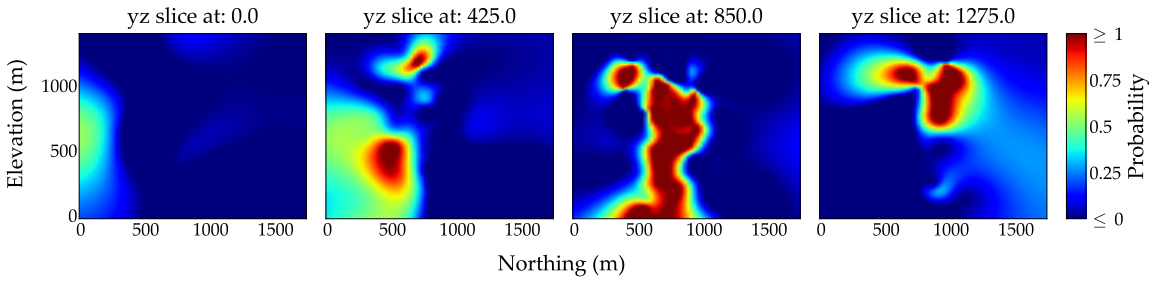
Moreover, the advantages of using an RBF over kriging is that first-order stationarity is not required, and variograms are not necessary (Martin, 2019). Therefore, an RBF utilizing a Gaussian kernel was selected to interpolate the porphyry indicator data. The Gaussian kernel is advantageous because it replicates short-scale continuity well, resulting in smooth models. The data does not have preferential directions allowing for easy comparison between SDF and NN-Threshold workflows. The resulting estimate is seen in Figure 6.4.



(a) Plan view slices



(b) East-West Section view slices



(c) North-South Section view slices

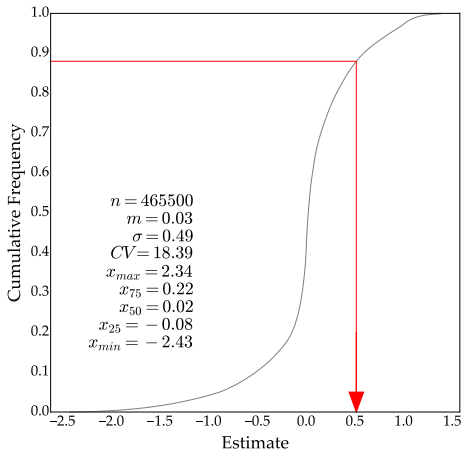
Figure 6.4: RBF interpolation of indicator values in plan and section views

Minor edge effects along strike of the North-East, South-West drilling and at depth in the model are results of extrapolation issues where no data exists. A geological

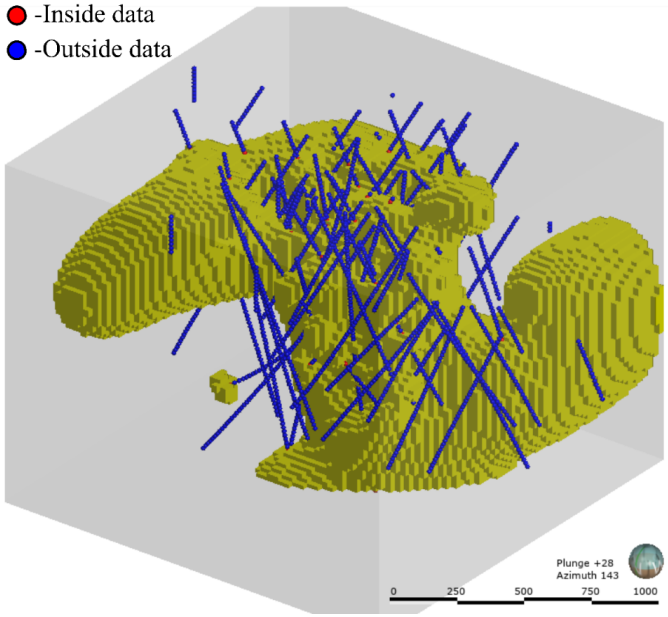
modeler would limit this extrapolation based on a site-specific geological model of deposition; however, the model is left unchanged in this demonstration.

6.1.3 NN-Thresholding

With the NN-model volume and indicator estimate complete, the thresholding step follows. As discussed in Chapter 3, the threshold value, z, is equal to one minus the ratio of the NN-model block volume to total volume (Equation 3.3) and is equal to 0.52. The CDF of the estimate in Figure 6.5a shows the threshold step whereby the volume ratio on the ordinate axes coincides with the z-value for thresholding the estimate.

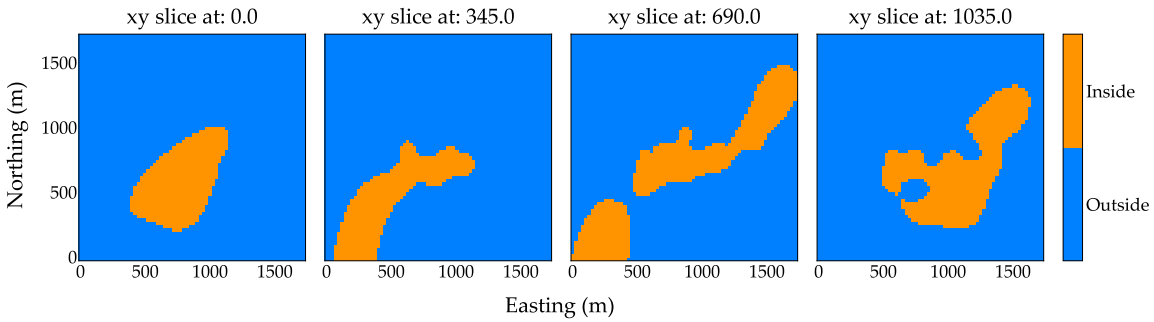


(a) Cumulative Distribution Function of RBF indicator estimate with volume ratio (0.88; ordinate axes) and threshold value (0.52; abscissa axes)

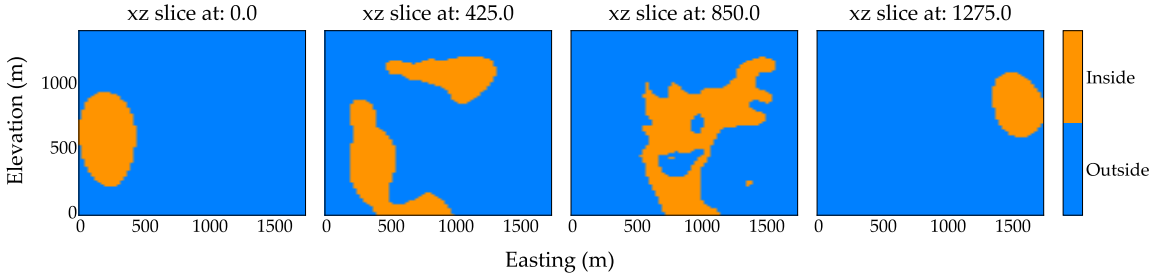


(b) Resulting 3-D thresholded indicator estimate representing unbiased global volume

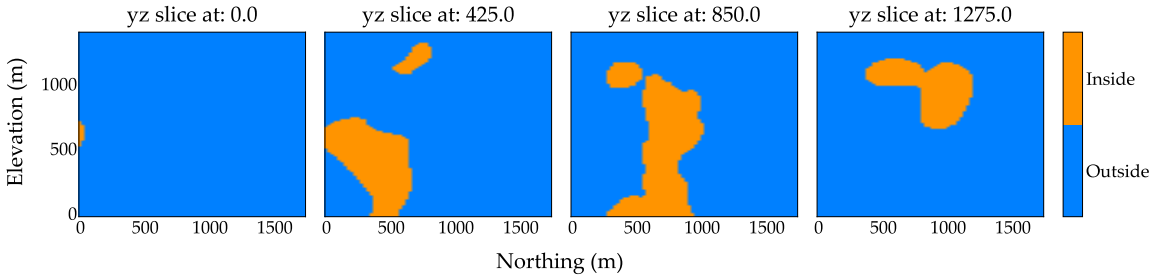
The z-value corresponds to an unbiased global volume for the indicator estimation represented in the CDF. All probabilities above the threshold are considered inside the model, while probabilities below are outside. The nearest neighbour threshold indicator estimate in Figure 6.5b & Figure 6.6 show the base-case model.



(a) Plan view slices



(b) East-West Section view slices



(c) North-South Section view slices

Figure 6.6: Categorical model from NN-threshold workflow in plan and section views)

6.2 Boundary Uncertainty

To account for uncertainty, further thresholding of the indicator estimate occurs in order to extract dilated and eroded boundaries. Volume uncertainty distributions with respect to different thresholds are illustrated in Figure 6.7. The distributions show the effect of incrementally changing thresholds and the resulting model volume ranges.

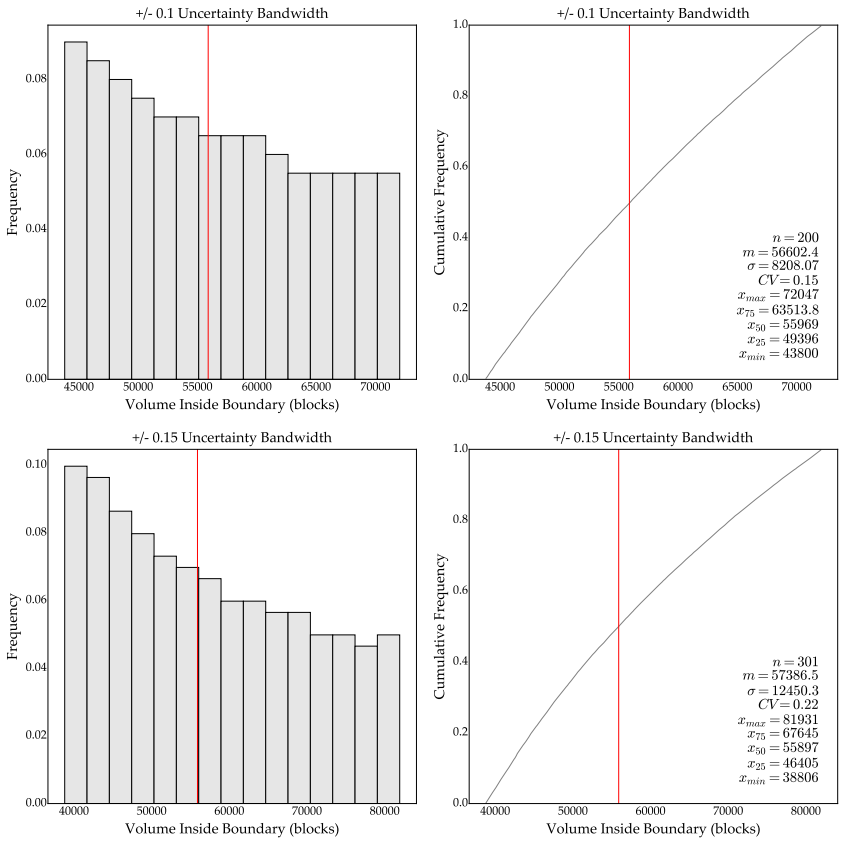
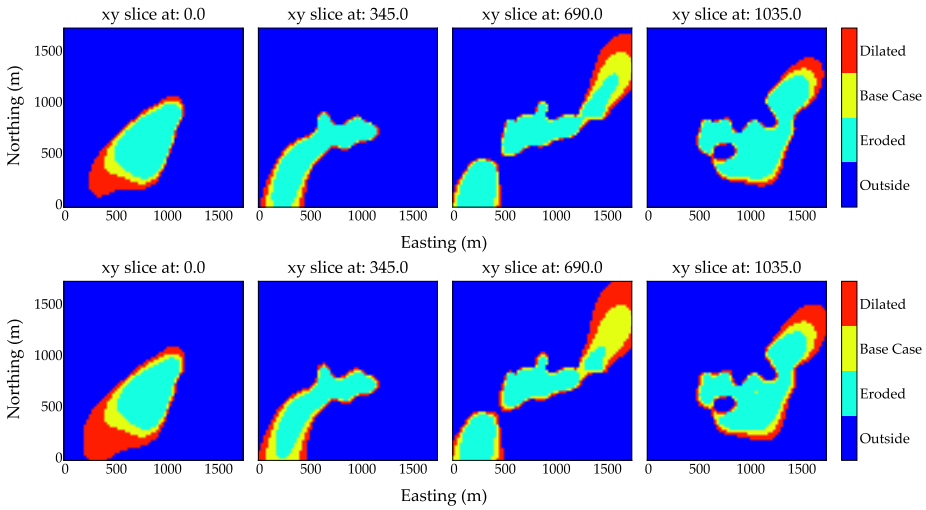


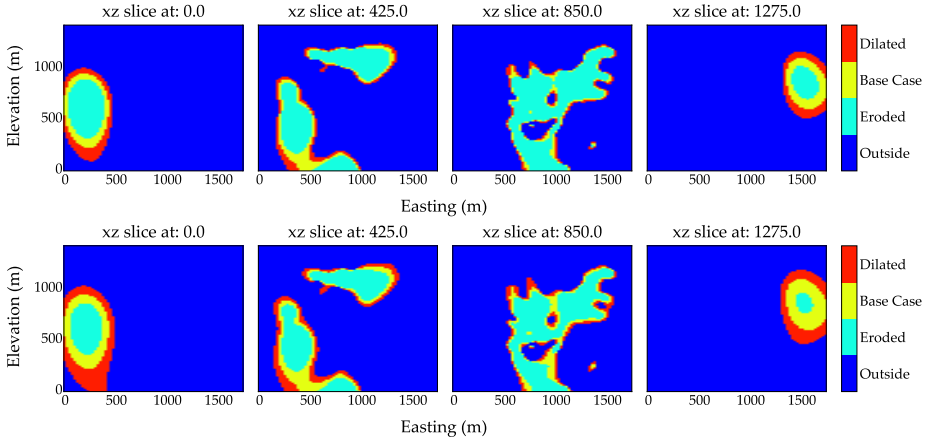
Figure 6.7: Histograms and CDFs of volume uncertainty distributions with respect to +/- 0.1 and +/-0.15 bandwidth limits. Red lines signify the volume of the NN-model volume

Chapter 4 shows that uncertainty thresholds between 0.1 and 0.15 of the NN-model base case threshold are reasonable. This conclusion was drawn from numerous experimental PTCs of varying model characteristics. Modeling at both thresholds was undertaken to assess the uncertainty. The categorical comparison results can seen below in Figure 6.8. Both uncertainty models exhibit realistic uncertainty; however, the wider 0.15 uncertainty threshold is conservative and chosen for the final model. The final 3-D model in Figure 6.9 consists of a single categorical model made of a base-case, eroded case, and dilated case. Edge effect mitigation through trimming and maximum distance NN-model parameterization can follow based on site-specific geological knowledge. The base-case volume is the exact volume of the NN-model at 55,898 blocks; the eroded and dilated boundaries are 38,806 and 81,931 blocks, respectively. Therefore, the range in uncertainty in the boundary model is 43,125 blocks

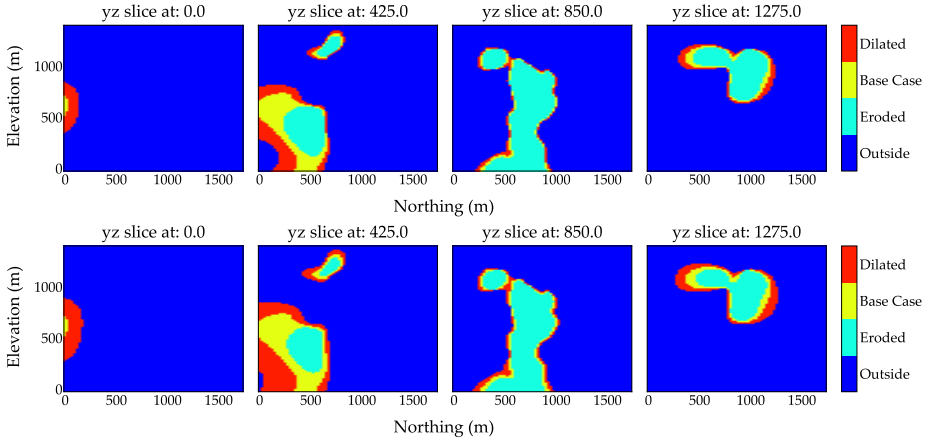
or 9.3% of the entire block model. The inside volume of uncertainty bandwidth is 77.1% of the inside NN-thresholded domain.



(a) Plan view slices of 0.1 (*above*) & 0.15 (*below*) threshold uncertainty



(b) East-West section view slices of 0.1 (*above*) & 0.15 (*below*) threshold uncertainty



(c) North-South section view slices of 0.1 (*above*) & 0.15 (*below*) threshold uncertainty

Figure 6.8: Categorical models for uncertainty based on 0.1 & 0.15 thresholds

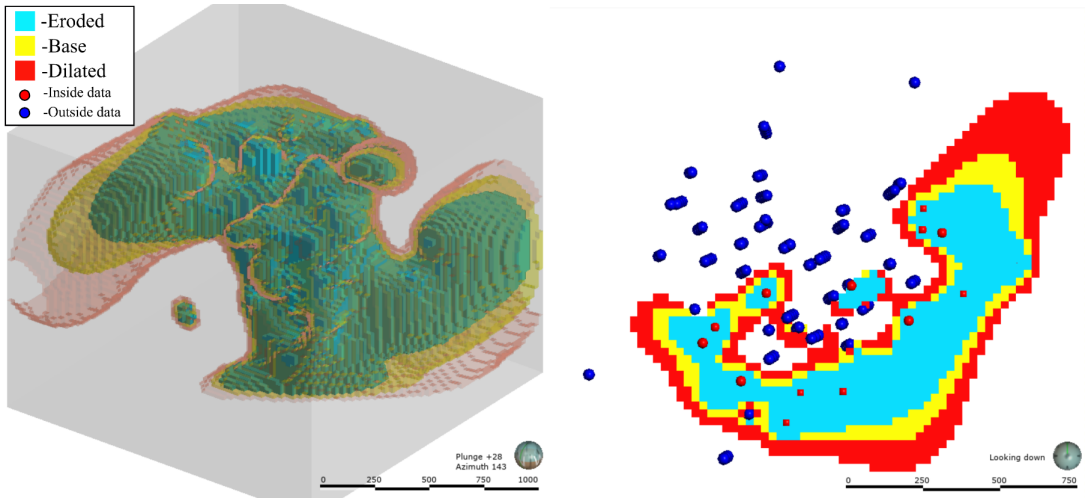


Figure 6.9: On the left, an isometric view of final boundary model with uncertainty: globally unbiased base-case (*yellow*), eroded case (*cyan*), and dilated case (*red*). On the right, a plan view slice showing drilling and uncertainty model.

6.3 Comparison to SDF Modeling

In order to assess the robustness of the indicator thresholding workflow and judge its practicality and viability, a comparison to SDF modeling is conducted. The SDF workflow is a popular implicit technique for modeling boundaries and quantifying volumetric uncertainty in the mining industry (Martin, 2019). For the SDF workflow, the methodology of D. Silva (2015) is utilized.

6.3.1 Boundary Models

A comparison to single domain SDF modeling illustrates the efficacy of the threshold indicator workflow. Both model workflows are interpolated with an RBF utilizing a Gaussian kernel. The SDF workflow establishes a boundary by extracting the isozero interface, whereby the model transitions from negative to positive values, as discussed in Section 1.2.1. The indicator threshold boundary extraction follows a NN-model volume threshold providing an unbiased global volume for the conditioning data, as discussed in Chapter 3. The models are seen in Figure 6.10. The SDF model global volume, defined as inside the isozero boundary, is 40,795 blocks, whereas the

NN and indicator threshold estimate of unbiased global proportions is 55,898 blocks. Therefore, the SDF model appears conservative and is 73% of the unbiased global volume.

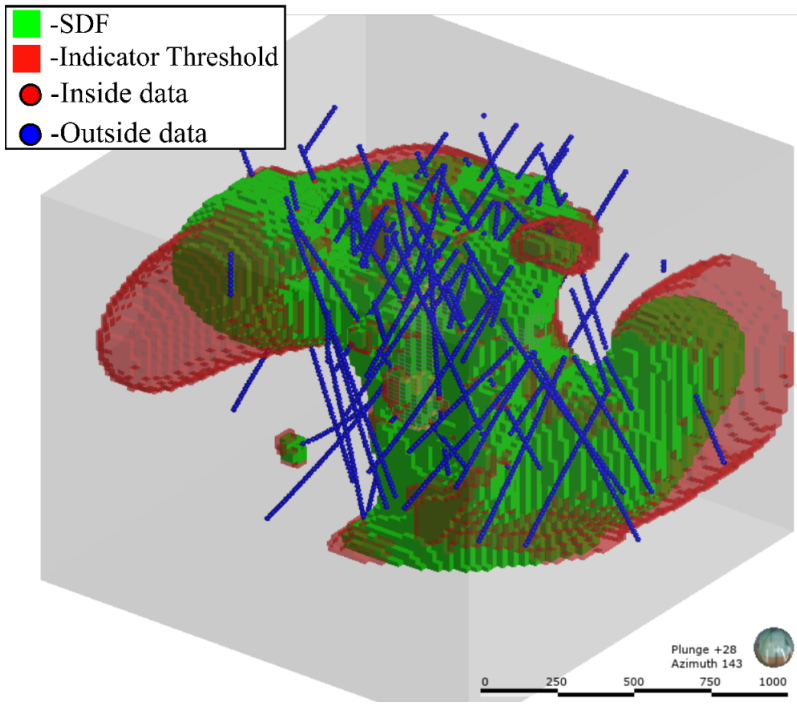


Figure 6.10: Indicator threshold model (*red*) and SDF interpolated model (*green*)

Both models have edge effects from boundary extrapolation. In these instances, domains may need constraining in order to minimize bias. The spatial structure of the conditioning data is mainly responsible for these extrapolation issues as the model extends into areas with few conditioning data(D. Silva, 2015). Figure 6.9 illustrates the extrapolation issue on the North-East boundary of the model in plan view, and two instances in the isometric 3D view (Figure 6.10). The kriging weights for the conditioning samples control the extrapolation and lead to faraway location estimates to become constant (D. Silva, 2015). Mitigation of edge effects can be approached by introducing artificial controlling points, changes to the RBF kernel or variogram range, or limiting the search in the nearest neighbour model extrapolation to produce smaller overall volume. Moreover, identification of these zones of increased uncertainty can be understood after the dilated and eroded boundaries are extracted.

Local bias attributed to the SDF algorithm only choosing the closest sample–yet not necessarily the sample in the direction of the boundary being modeled – is seen in Figure 6.11. In instance A, the SDF algorithm selects the closer data point indicated by the yellow arrow. The selection is correct for defining the boundary to the right (East); however, the SDF value, when interpolated on the left-hand side, is conservative. In contrast, the indicator threshold estimation appears more reasonable. At location B, the SDF selects the sample immediately beside it, resulting in a small conservative boundary. The indicator estimate renders a more significant boundary that closely follows the structure of the surrounding data.

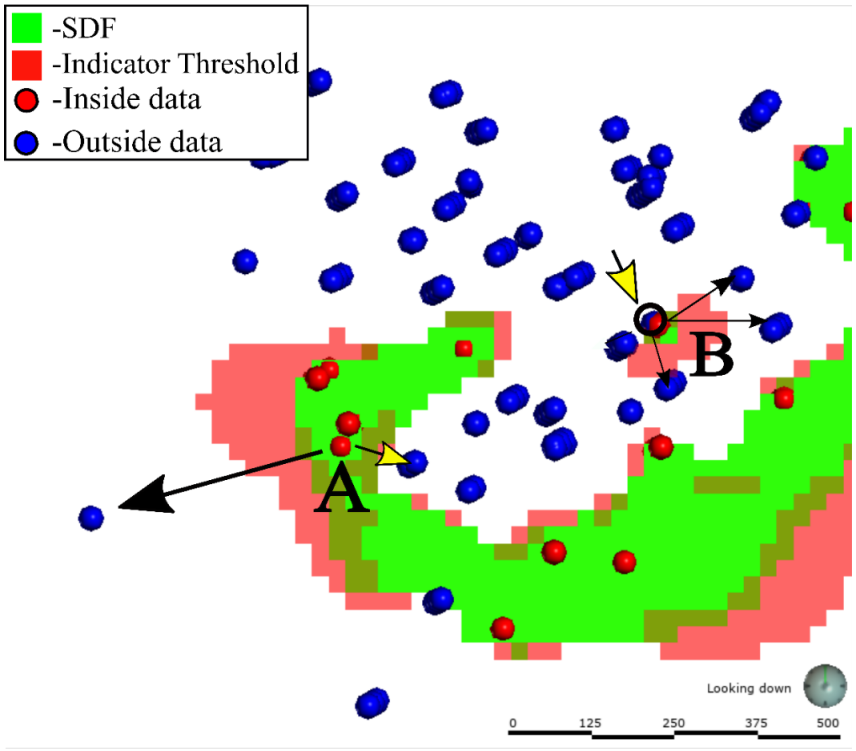


Figure 6.11: Indicator threshold base case model (*red*) and SDF interpolated model (*green*). Two instances of local conservative bias attributed to SDF algorithm (A, B). Yellow arrows indicate SDF algorithm choosing closest data point, but not necessarily most pertinent. White arrows indicate other samples not chosen from the SDF algorithm selection, yet important

6.3.2 K-Fold Analysis

K-fold analysis is a popular method for checking and validating geostatistical models. The method subsets the conditioning data into K folds. One subset of the data is

left out of the model workflow, while the $K-1$ remaining folds are used to estimate. The left out subset is considered the true value and compared to the estimate from the remaining $K-1$ folds. The process is repeated for all folds (C. Deutsch, 2018). For the case study, K-fold analysis was done partitioning into five-folds. Each fold in the 5 K-fold series contains between 651-659 data. The SDF & Indicator threshold workflows from K-fold validation are compared.

6.3.2.1 Error

Error is calculated using the validation model versus the left out data. The error is the sum of the False Positive (FP) and False Negative (FN) from a specific K-fold and its validation subset. In this context, the Negatives are being modeled as 'inside' and Positives are modeled as 'outside'. The summed values are divided by the total number of validation subset data to arrive at a percentage of misclassification. The classification of errors from K-fold analysis are in Figure 6.12.

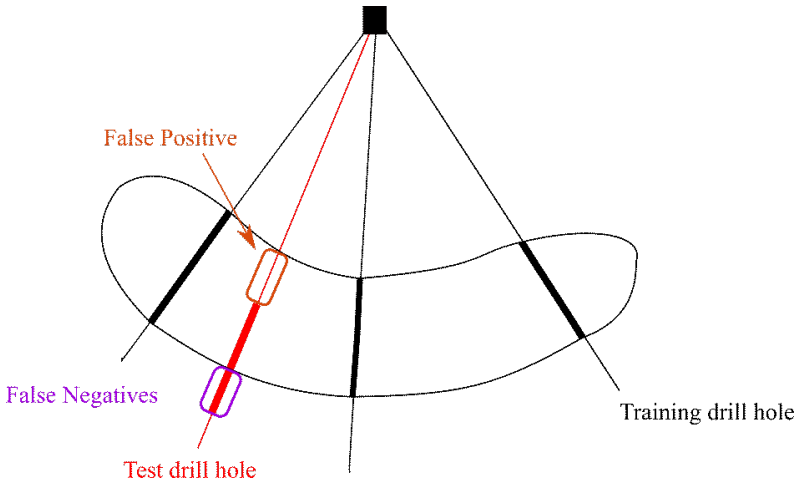


Figure 6.12: Leave-n-out classification using drill hole samples. The black lines signify drill holes. The thicker black lines indicate samples considered inside of the geology with the modeled geology seen in black. The test drill hole that is left out of the workflow is seen in red with the thicker intersection indicating where the true geology is inside. False negatives are highlighted in fuschia and false positives are seen in orange.

An error rate of zero means there were no misclassifications, while an error rate of 1 indicates complete misclassification. In the 5-fold analysis the SDF performed better 80% of the time with an average error rate of 0.155 compared to 0.169 for the indicator

workflow. In an additional 20-fold analysis, the SDF outperformed indicator workflow 55% of the time with average error rate per fold of 0.146 compared to 0.148 for the indicator workflow. The error on a per fold basis can be seen in Tables 6.2 & 6.3 and Figure 6.13.

Breaking out Type I & Type II error, the indicator threshold method misclassifies less Type I errors (False Positives) for all five folds at an average rate of 0.055 errors compared to 0.069 for the SDF models. However, Type II (False Negatives) show the SDF outcompeting the indicator threshold in 80% of instances with an average error rate of 0.086 compared to 0.114 for the indicator threshold folds. In the case of 20-folds, the indicator threshold workflow equals or outperforms the SDF workflow 60% and 75% of instances for Type I and II errors, respectively. The averages over the 20 folds for Type I errors for the indicator threshold is lower at 0.063 in contrast to 0.066 for the SDF. Moreover, the Type II average errors are similar for the SDF and indicator threshold workflows at 0.080 and 0.086, respectively. Both methodologies perform similarily in K-fold analysis, with instances of the indicator threshold workflow outperforming the SDF, and circumstances where it is worse.

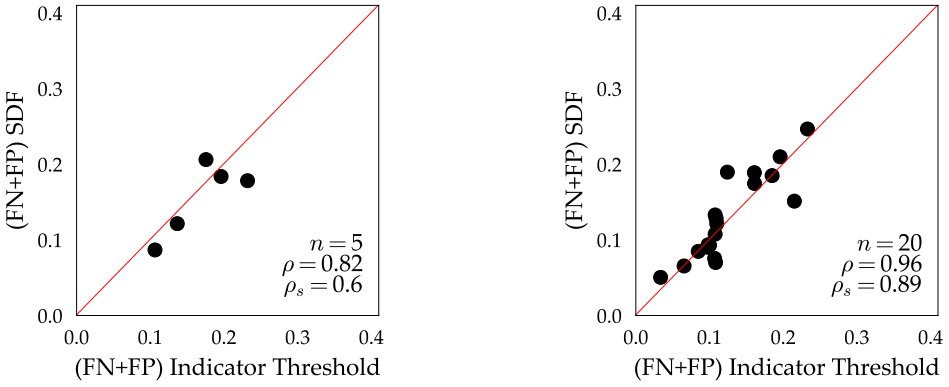
Although the SDF out-competes the indicator threshold method in certain circumstances, the minimal error difference between workflows on a per-fold basis is evidence the indicator threshold methodology is robust.

Table 6.2: Error rates for 5-fold analysis

SDF	Indicator Threshold
0.121	0.137
0.178	0.232
0.206	0.176
0.087	0.106
0.184	0.196

Table 6.3: Error rates for 20-fold analysis

SDF	Indicator Threshold
0.085	0.085
0.108	0.108
0.093	0.1
0.189	0.161
0.151	0.215
0.05	0.034
0.19	0.124
0.174	0.161
0.185	0.185
0.128	0.109
0.093	0.098
0.065	0.065
0.467	0.56
0.121	0.11
0.247	0.233
0.09	0.097
0.07	0.108
0.21	0.196
0.133	0.108
0.075	0.107



(a) 5 K-folds **(b)** 20 K-folds

Figure 6.13: Error comparison for SDF and Indicator Threshold workflows

6.3.2.2 Matthews Correlation Coefficient

The Matthews Correlation Coefficient (MCC) measures the quality of binary classifications. In this context, the classification of being correctly inside or outside of the domain. Matthews (1975) introduced the coefficient, which is widely used in bioinformatics and machine learning and comes from the confusion matrix seen in Figure 6.14. The classification frequencies are input into Equation 6.1. The return values range from -1 to 1 with 1 being perfectly predicted, 0 indicating that the values are no better than random predictions, and -1 signifies that there is a definitive disagreement between prediction and observation.

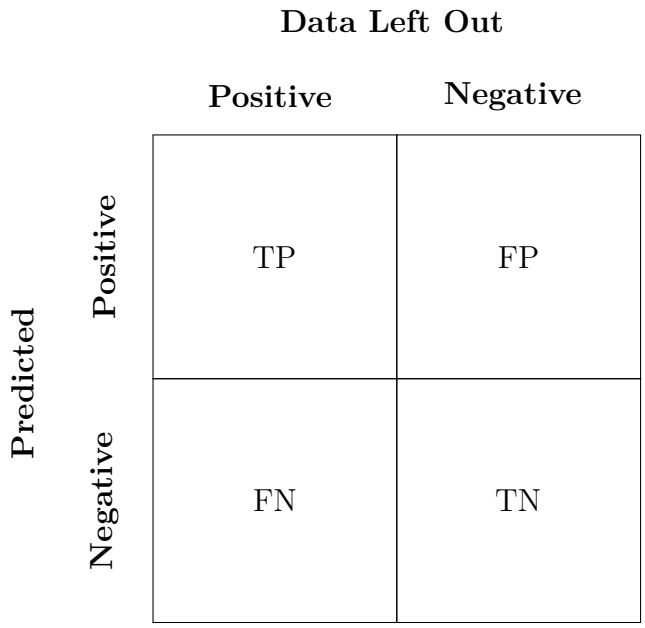


Figure 6.14: Confusion Matrix

The True Positive (TP) are predicted in and are actually in; True Negative (TN) are predicted out and are actually out. This is in contrast to False Positives (FP) that are predicted in but are actually out, and False Negatives (FN) that are predicted out, but are actually in (see Figure 6.12). Once the data is classified, the MCC is as follows:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FP)(TN + FN)}} \tag{6.1}$$

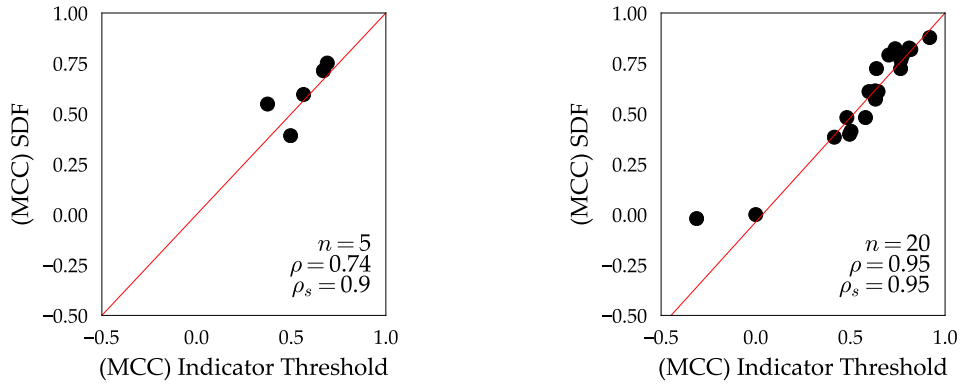
The MCC results comparing SDF and indicator theshold workflows over K folds are in Figure 6.15. Higher values indicate better performance. For the 5-fold analysis (Figure 6.15a) the SDF outperforms the indicator workflow in 80 % of the instances with average MCC values of 0.60 & 0.56, respectively. However, the 20-fold analysis (Figure 6.15b) shows the indicator thresholding workflow equalling or outperforming the SDF in 55% of the instances; however, with average MCC values of 0.584 for the SDF and 0.580 for the indicator thresholding. Therefore, the workflows compete closely for K-fold analysis and misclassification metrics.

Table 6.4: Matthews Correlation Coefficient for 5-fold analysis

SDF	**Indicator Threshold**
0.714	0.67
0.548	0.375
0.391	0.497
0.752	0.691
0.596	0.565

Table 6.5: Matthews Correlation Coefficient for 20-fold analysis

SDF	**Indicator Threshold**
0.819	0.819
0.767	0.767
0.788	0.774
0.481	0.58
0.724	0.638
0.878	0.919
0.573	0.632
0.610	0.646
0.481	0.481
0.724	0.765
0.610	0.598
0.000	0.000
-0.020	-0.313
0.399	0.496
0.384	0.415
0.826	0.811
0.791	0.703
0.614	0.630
0.413	0.503
0.823	0.736



(a) 5 K-folds

(b) 20 K-folds

Figure 6.15: MCC comparison for SDF and Indicator Threshold workflows

6.3.2.3 Outliers

The outlier in both Figure 6.13b & 6.15b corresponds to Fold 13. This particular fold subsets a validation drillhole comprised of inside data surrounded by multiple training holes of outside data leading to numerous False-Negative misclassifications in both the SDF & indicator threshold workflows. Moreover, the fold also subsets a dominantly outside drillhole in the center of the training model. The result is many False-Positive misclassifications in both methods. Figure 6.16 illustrates the misclassification issues in section view.

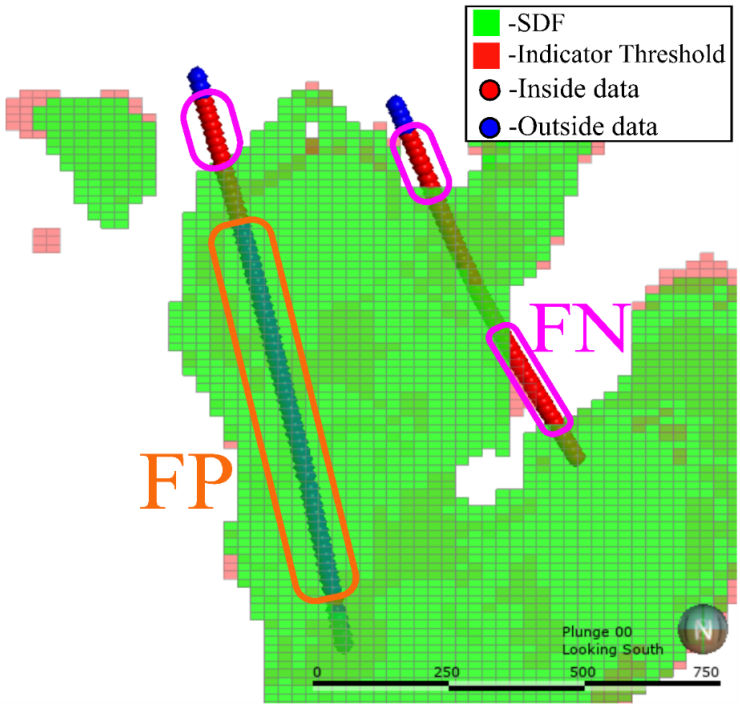


Figure 6.16: Training Indicator threshold model (*red*) and SDF interpolated model (*green*) along with validation data for fold 13. False negative and False Positive zones are outlined in fuschia and orange, respectively. The numerous misclassifications lead to high error rates and low MCC coefficients for both the SDF & Indicator threshold models

6.3.3 Boundary Uncertainty

Boundary uncertainty for the modeling methods and their respective eroded, and dilated cases are seen in Figure 6.19. For the SDF boundary uncertainty, the C-parameter calibration uses a jacknife workflow, as discussed in Chapter 1. SDF models for uncertainty with increasing C-values are seen in Figure 6.17. The uncertainty volumes are summarized in Table 6.6.

Table 6.6: C-values and corresponding volumes for SDF uncertainty

C-value (m)	Uncertainty Volumes (blocks)			
	Eroded	**Dilated**	**Difference**	**% Difference**
25	35,602	48,884	13,282	27.2
50	32,243	55,744	23,501	42.2
75	29,789	61,842	32,053	51.8
100	27, 971	67,232	39,261	58.4
150	23,573	84,104	51,209	72
200	22,280	90,870	60,531	75.5
250	21,263	96,642	68,590	78

The final SDF uncertainty model execution was with a C-value of 50m. The C-value coincides roughly with the drill hole spacing– which is often a C-value chosen for reasonable SDF model uncertainty in the absence of jacknife analysis (Martin, 2019).

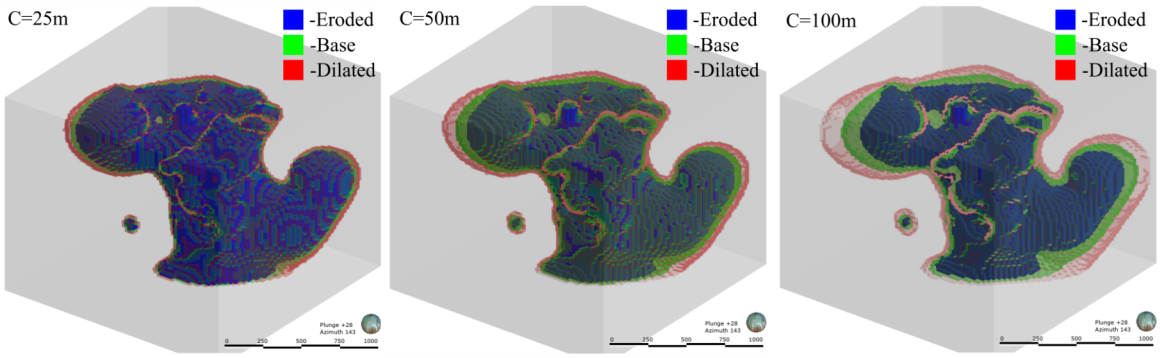


Figure 6.17: SDF models with uncertainty for 25, 50, and 100m C-parameter values

The indicator threshold uncertainty follows by taking 0.15 thresholds above and below the NN-model unbiased threshold, as discussed in Chapter 4. Plan and section views of the respective dilated and eroded cases for modeling workflows are in Figure 6.19. The uncertainty bandwidths for the indicator threshold models expand in areas with fewer informing data. The bandwidth extensions are pronounced in plan views where data configurations are less concentrated with higher variation across the drilling grid. This phenomenon is also visible in North-South section views in areas with fewer conditioning data nearby. Both uncertainty models behave comparatively well

in zones with plenty of nearby informing data and tight spacing; however, a noticeable conservative bias for the SDF is evident in zones where data spacing increases or areas where no data is present.

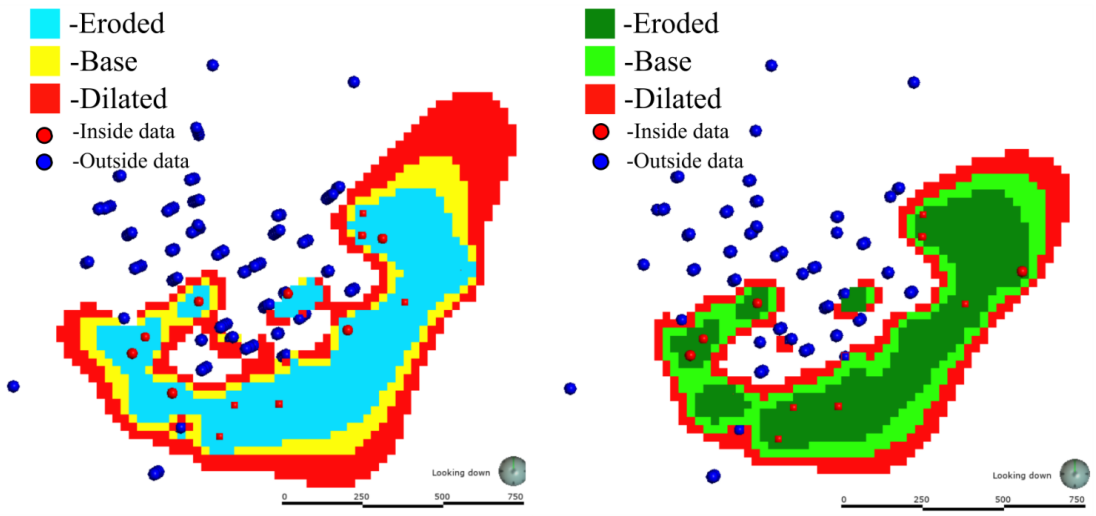
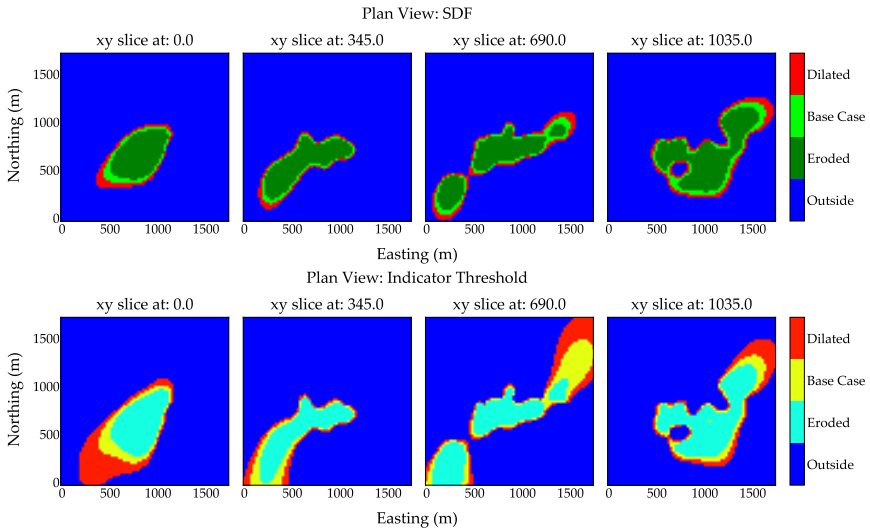


Figure 6.18: Plan view of Indicator threshold model for uncertainty (*left*) and SDF model for uncertainty (*right*) showing conservative bias in SDF modeling

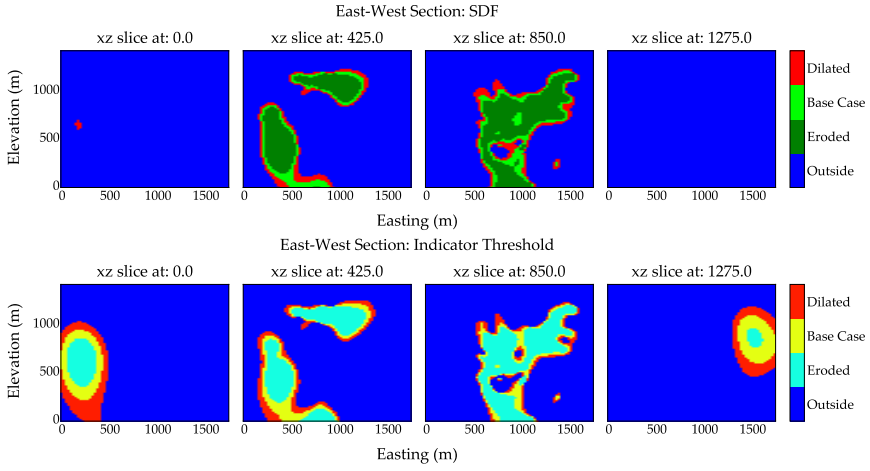
Figure 6.18 illustrates the SDF conservative bias related to the C-parameter for bandwidth. On the left side of the domains, where widely spaced data exists, the indicator threshold model shows variation through its uncertainty bandwidth. In contrast, the SDF consistent bandwidth is not following the structure of the local conditioning data. The upper and lower right of Figure 6.18, where there are few conditioning data, exhibits the same bias and the less realistic consistency in the bandwidth attributed to the C-parameter is pronounced. Globally, the indicator threshold dilated boundary is 81,932 blocks compared to the 55,744 blocks comprising the SDF dilated case.

The eroded cases for uncertainty vary between workflows globally with the eroded SDF comprising 32,243 blocks compared to the indicator threshold model with 38,807 blocks. The consistency in the indicator threshold model and adherence to local data results in continuity throughout the model, whereas the SDF model is disjointed.

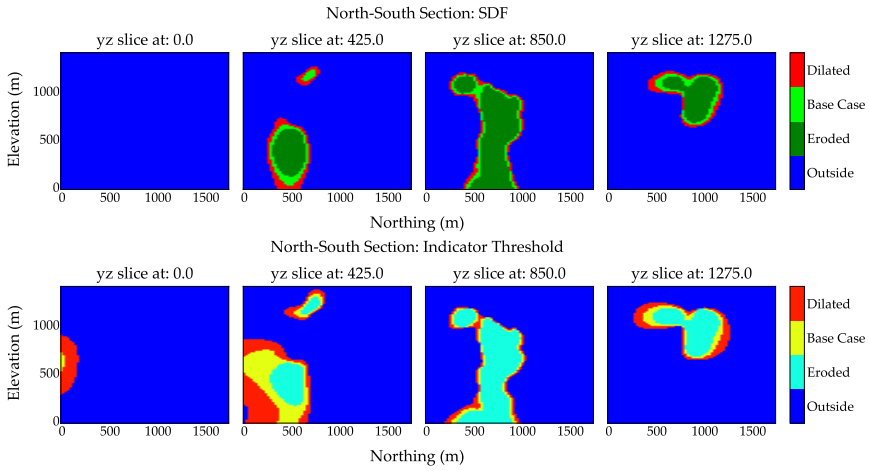
The uncertainty from indicator thresholding gives more substantial bandwidth uncertainty and closely follows the structure of the data, resulting in a realistic model for uncertainty.



(a) Plan view slices of SDF uncertainty (*above*) & Indicator Threshold uncertainty (*below*)



(b) East-West section view slices of SDF uncertainty (*above*) & Indicator Threshold uncertainty (*below*)



(c) North-South section view slices of SDF uncertainty (*above*) & Indicator Threshold uncertainty (*below*)

Figure 6.19: SDF & Indicator Threshold categorical models for uncertainty

6.4 Results and Considerations

Data from a porphyry deposit was used to construct a boundary model for an intrusion. Firstly, the nearest neighbour model, with a reasonable maximum distance parameter, yields a globally unbiased volume for the inside domain. Indicator interpolation using an RBF with Gaussian kernel, maps a field of probabilities throughout the domain. The resulting interpolation is an indicator estimate which is thresholded to the unbiased global volume ratio derived from the nearest neighbour model giving a base case boundary model. Figure 6.20 shows the volume uncertainty from competing workflows. The global and local uncertainty studied through K-fold analysis for competing workflows also illustrates how the SDF methodology performs adequately locally, yet still results in globally biased model volumes.

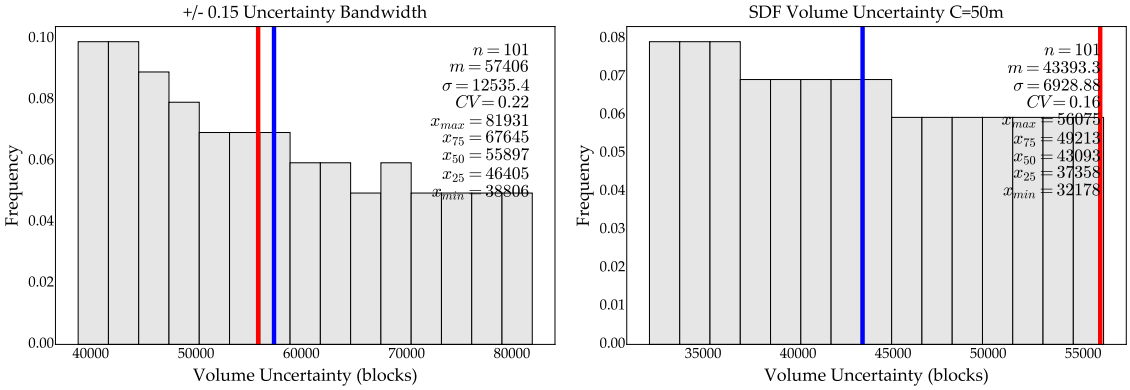


Figure 6.20: Comparison of volume uncertainty for Indicator Threshold (*left*) and SDF (*right*) workflows with their respective mean volumes seen as blue lines. The NN model volume is 55,898 blocks and shown by red line. The Indicator volume uncertainty mean is 57,406 blocks compared to the SDF uncertainty mean of 43,393 blocks. The distribution of volume uncertainty and the abrupt transitions at higher and lower thresholds are an area for future research. The frequency of model volumes in the tails of the distribution indicate that both workflows result in models with significant uncertainty.

Uncertainty is accessed via further thresholds at +/-0.15 of the base case threshold. The final model consists of a base-case, eroded case, and dilated case categorical boundary model. The model is robust, as shown through comparison to SDF modeling over numerous K-fold testing. The indicator thresholding has distinct advantages over the SDF, which is exemplified in areas of data asymmetry. These asymmet-

ric data configurations lead to conservative biases in the SDF models, but are not as pronounced in the indicator threshold models. Moreover, uncertainty assessment provides a continuous and realistic bandwidth of uncertainty that strictly honours local conditioning data structure in the indicator threshold models. The final model checking and validation using K-fold analysis and assessing error and the MCC is carried out on the competing methods.

Assessment and considerations for the algorithm to be reliable have a foundational basis on six key concepts: simplicity, speed, objectivity, data integration, access to uncertainty, and geologic realism (McLennan, 2007). For simplicity, the algorithm is easy to understand and easy to implement. For speed, the indicator estimation using RBFs is fast and reliable; large datasets (N>30,000) can slow the process down. Objectivity: The purpose of implicit modeling is not only to speed the modeling process, but also control the subjectivity involved in explicit approaches. With estimation parameters equal, the reproducibility of models follows, as seen in the K-fold analysis. Data Integration pertains to further indicator information easily being integrated into the model software for up to date boundary assessment. Orientation data, such as strike and dip of structural data, can also be integrated into RBF workflows (Martin, 2019). Access to uncertainty is straightforward. Indicator estimates map a field of local uncertainty in their probabilities. Thresholding these probabilities, above and below the base case gives easy, realistic access to uncertainty. The algorithm creates geologically realistic models. A model is realistic if it agrees with the geological model of interpretation and evidence gathered from the field (D. Silva, 2015). RBFs honor arbitrary shapes and can be integrated with field measurements leading to realistic models to be checked by the geo-modeler. Robustness is also a key principle to check for new algorithms (D. Silva, 2015). Through K-fold analysis, the model was shown to compete with the popular, industry-favoured, SDF methodology. The known issues arising from SDF modeling –conservative bias in presence of data asymmetry and C-parameter simplicity– are mitigated by the indicator estimation workflow for single domain modeling. The result is a globally unbiased volume model with fair

eroded and dilated boundaries for bandwidths of uncertainty.

Chapter 7
Multi-category Case Study

A modified indicator threshold workflow for multi-categorical boundary modeling is demonstrated. Five lithological domains are modeled separately and combined to form a single multi-categorical model. The methodology follows the same principles as the single domain example with additional steps to determine the prevalent category at a given node. There are three intrusions, an oxide domain, and a sulphide domain. The data has been previously modeled by D. Silva (2015) using the SDF multicategory methodology. K-fold analysis, and a comparison to equivalent SDF modeling, validates and highlights the multi-indicator threshold method efficiencies that produce realistic boundary models with access to uncertainty.

7.1 Multi-category Domain Boundary Modeling

An oxide domain, a sulphide domain and three intrusions from the porphyry drill hole dataset used in Chapter 6, are used for a multi-category boundary modeling case study. The coded data in Figure 7.1 shows the lithological data. The data is within a domain 1750mN by 1750mE and 1425mZ with a block dimension of 25mX25mX15m for a total of 465,500 blocks. The categories are modeled separately and combined to form a single final model.

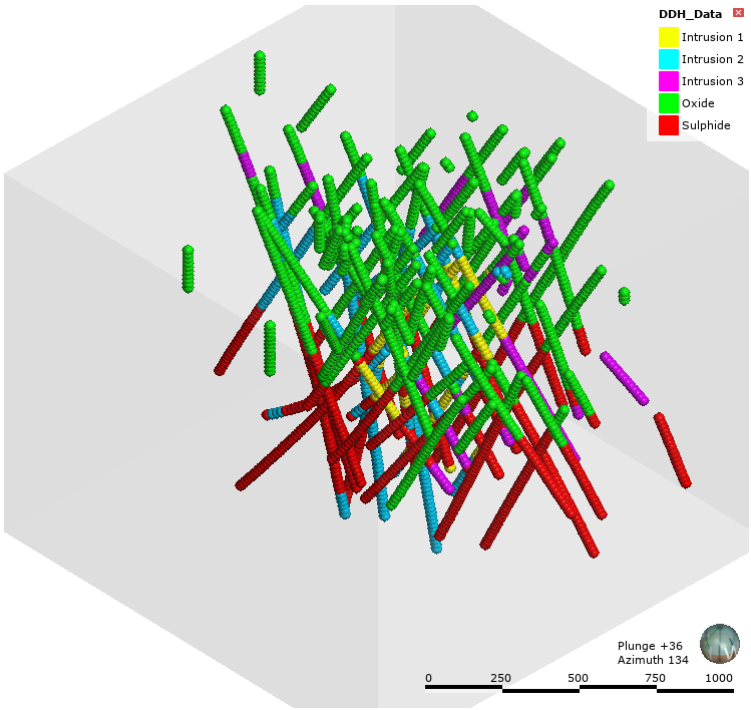


Figure 7.1: Drill data from porphyry dataset with multiple categories. Codes 4,5,6 are intrusives. Codes 2,3 are oxide and sulphide domains.

.

7.1.1 Nearest Neighbour Model

Nearest neighbour models, as discussed in Chapters 3 and 6, give a spatially unbiased configuration of the data. NN model volumes of all individual cells, return an unbiased global volume for that category. Nearest neighbour models with varying maximum distances can be executed in order to understand and control edge effects created from extrapolation. For this case study the maximum distance chosen was the distance whereby the models stabilize at a maximum inside volume. Table 7.1 shows the domains and their corresponding NN-model volumes.

Table 7.1: NN-Model global volumes for categories

Domain	Volume (blocks)
Intrusive 1	5,579
Intrusive 2	20,063
Intrusive 3	30,256
Oxide	230,518
Sulphide	179,084

The resulting NN-models are seen below in Figure 7.2. The intrusions are opaque, while the oxide-sulphide domains are translucent for viewing purposes. Minor edge effects due to extrapolation issues exist for intrusion 3. Intrusion 1 is the smallest domain and forms the core of the porphyry, partially obscured by the other intrusions in Figure 7.2.

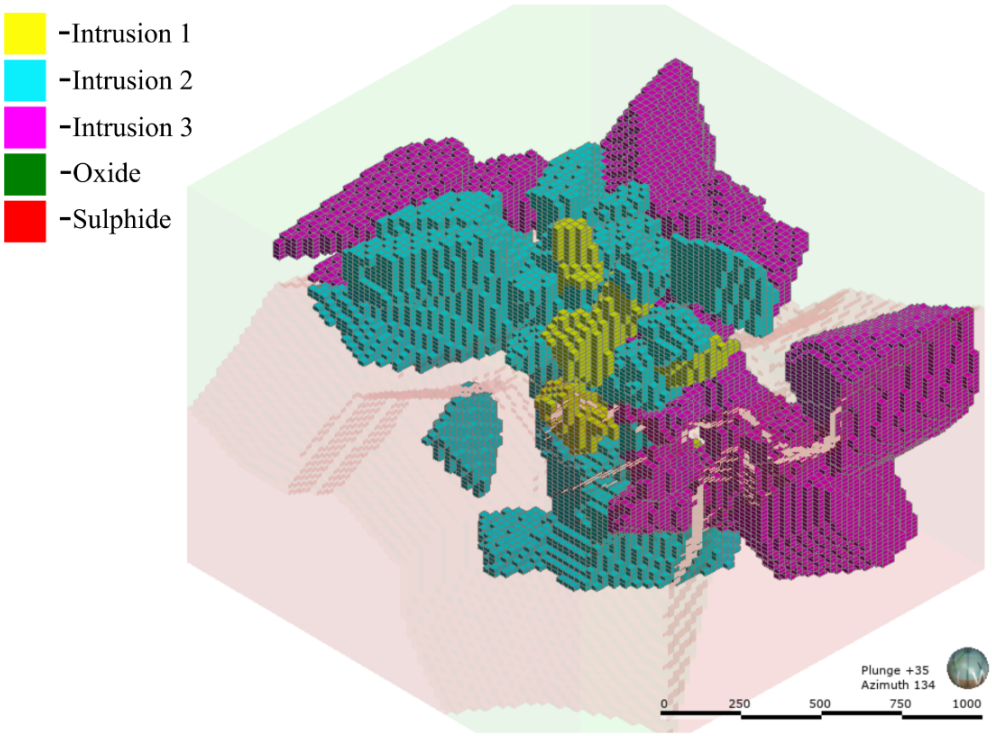


Figure 7.2: Nearest Neighbour models with opaque intrusions and translucent oxide & sulphide domains

.

7.1.2 Indicator Estimation

Multiple indicator estimation for interpolating the indicator data results in a field of probabilities for each category. As discussed in Section 5.1, the algorithm chosen will ideally produce smoothly varying estimates to be thresholded. Radial Basis Functions (RBFs) are a global estimation interpolator that fits a unique function to the conditioning data. The result is an artifact-free, smooth model that honours arbitrary shapes. The RBF utilizes an isotropic Gaussian kernel allowing for easy comparison between SDF and NN-Threshold workflows. The resulting indicator estimates are seen in Figure 7.3.

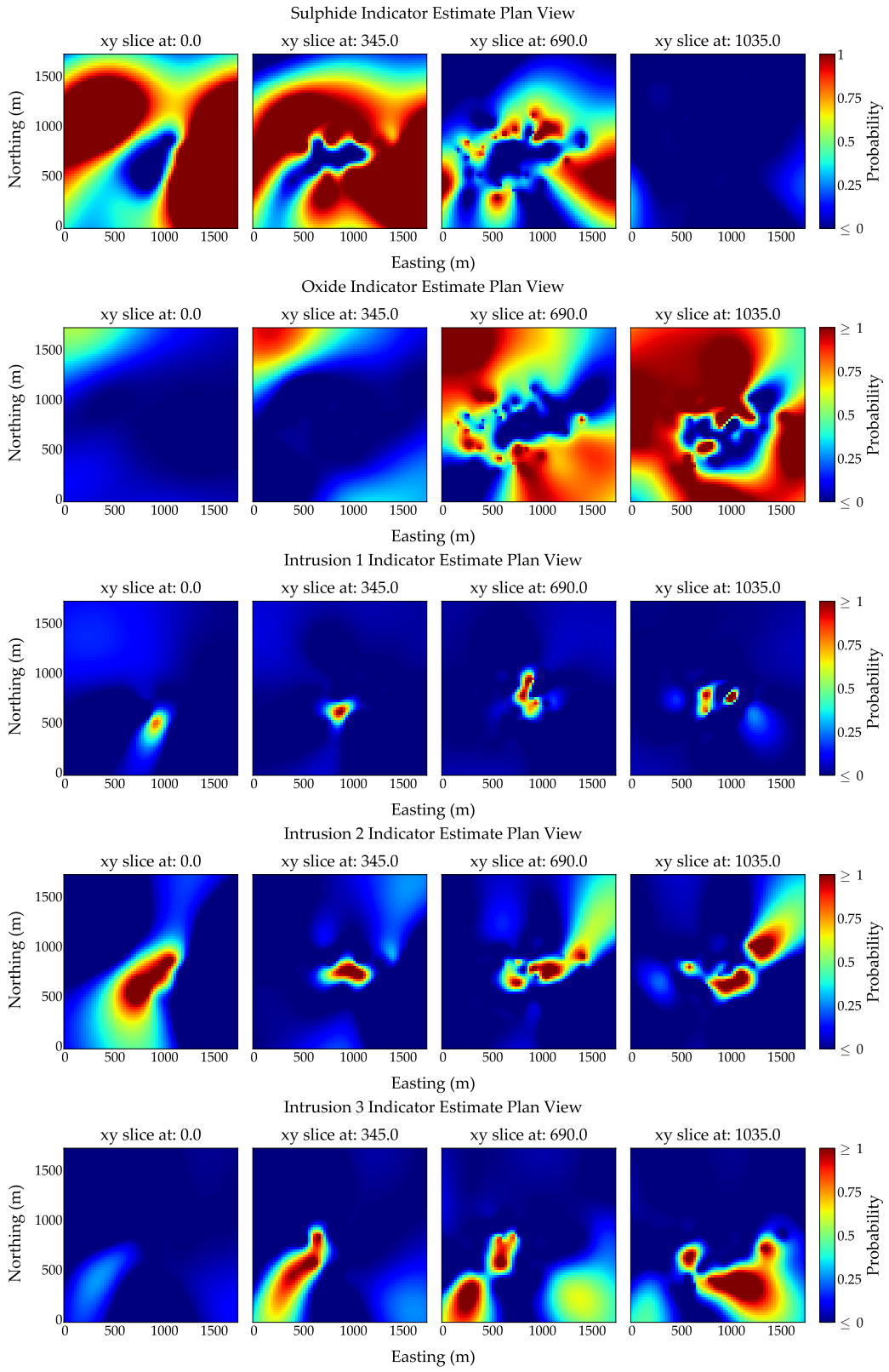


Figure 7.3: Plan view sections of indicator estimates for the five domains

The oxide and sulphide domains comprise the majority of the domain area under study. Intrusion 1, the smallest, is confined to the center of the domain with no evident edge effects. In contrast, minor edge effects are evident in Intrusions 1 and 2 where there are few conditioning outside data.

7.1.3 NN-Thresholding

With the NN-model volume calculated and the indicator estimate complete, the thresholding step follows. As discussed in Chapter 3, the threshold is 1 minus the NN-model block volume divided by the total model volume (Equation 3.3). The CDFs of the estimate in Figure 7.4 show the thresholding step whereby the volume ratio on the ordinate axes determines the z-value used for the thresholding of each indicator estimate.

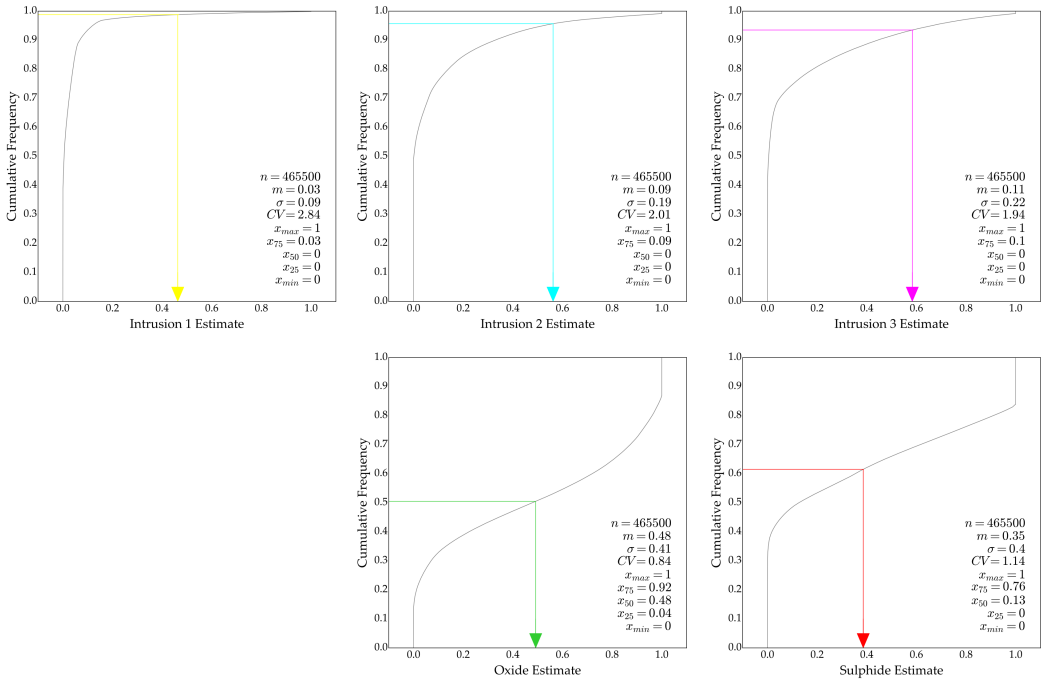


Figure 7.4: Cumulative Distribution Functions of RBF indicator estimates with volume ratios on ordinate axes and resulting threshold values on abscissa axes identified by coloured arrows

.

The unbiased global volume for a particular estimate corresponds to a z-value thresh-

old extracted from its CDF. Probabilities above the threshold are inside the domain, while probabilities below are outside of the boundary. Each indicator estimate is thresholded separately resulting in five distinct boundary models. The plan view sections of the five categorical models are seen in Figure 7.6.

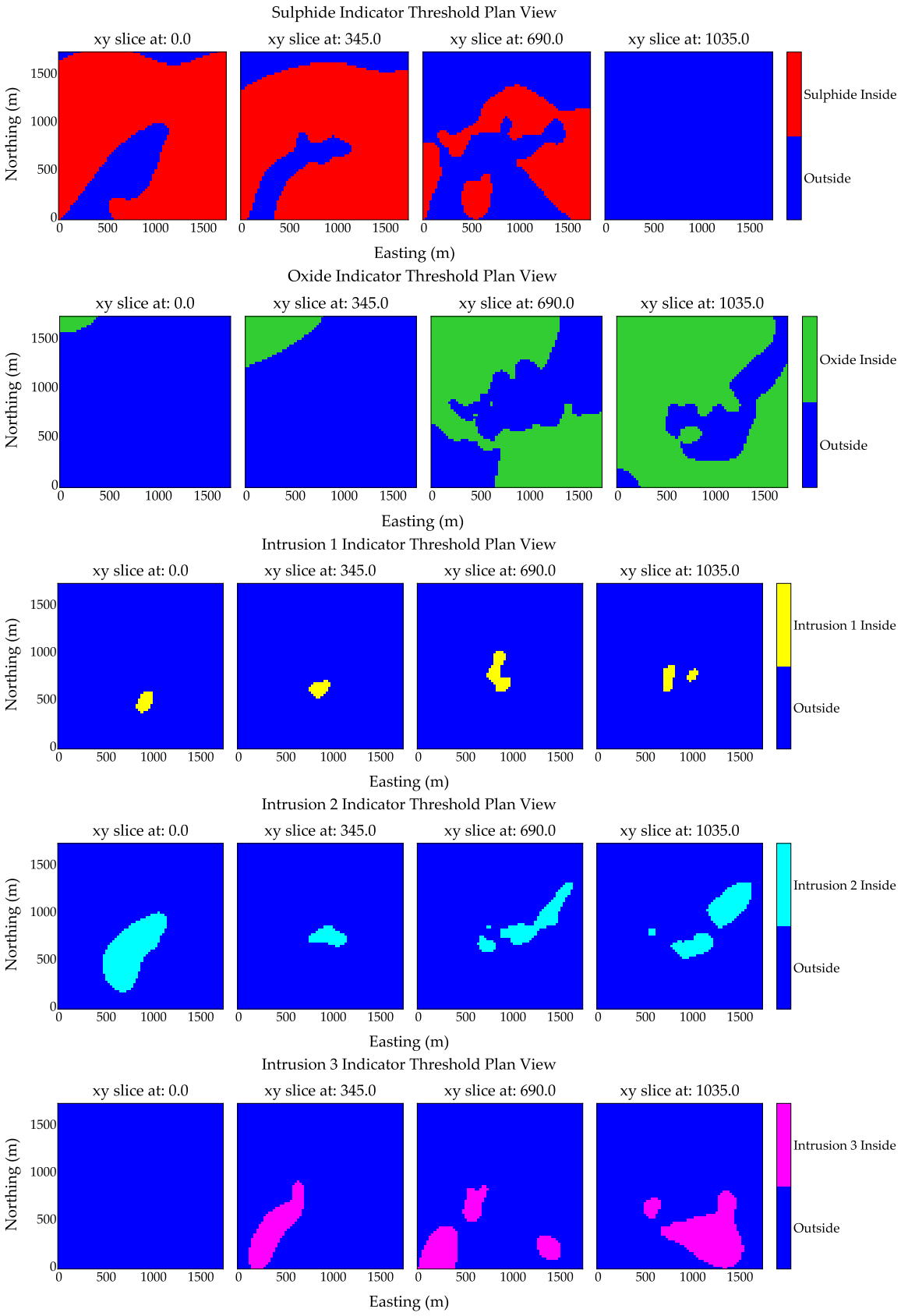


Figure 7.5: Plan view sections of categorical models for five domains

7.1.4 Post Processing

With five indicator threshold models – one for each category– the issue of how to combine the models into one unique multi-categorical model arises. Each model matches its NN-model in block volume; however, there is overlap in the blocks between the estimates. Two methods for determining the lithology at a given node are tested. The first, Probability-Threshold Selection (PTh), selects the lithology with the largest difference between the estimate and its respective NN-threshold z-value. The results are in Figure 7.6.

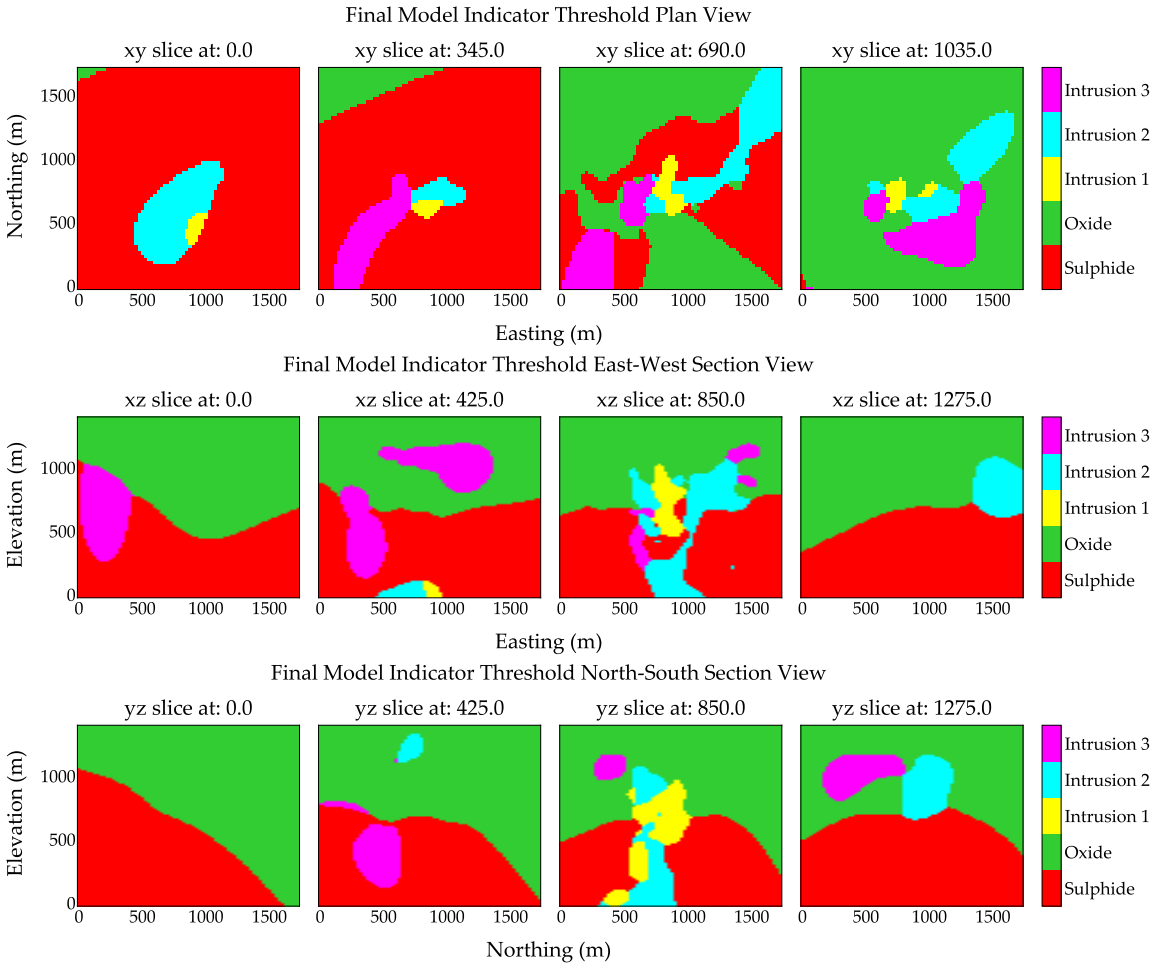


Figure 7.6: Plan and Section view slices of categorical models for five domains using Probability-Threshold selection algorithm

The second method, Probability-High Selection (PH), selects the highest indicator

estimate value to define the category (Figure 7.7). The methods do not vary significantly across the oxide and sulphide domains; however, noticeable differences exist within the intrusives. Intrusion 2 is significantly overestimated by the PH method with accentuated edge effects. The volumes from each method and NN model volumes for the domains are in Table 7.2. In every domain, the methods differ from their respective NN-models, confirming the overlap of nodes.

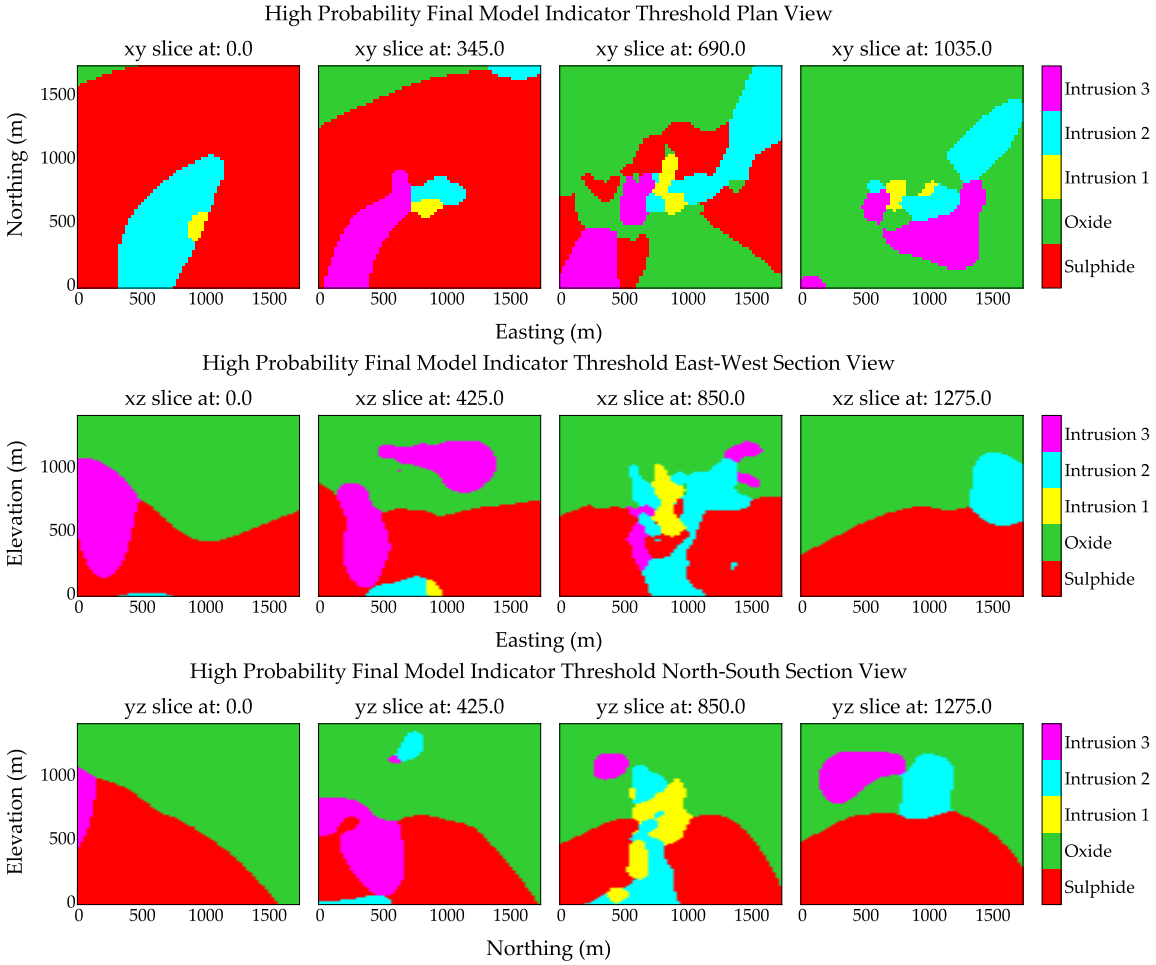


Figure 7.7: Plan and Section view slices of categorical models for five domains using Probability-High selection algorithm

Table 7.2: Volumes for multi-category domains using Probability-Threshold, Probability-High, and NN-models

Domain	Volumes (blocks)				
	NN-model	**Prob-Thresh**	**% NN**	**Prob-High**	**% NN**
Intrusion 1	5,579	5,585	100.1	5,179	92.8
Intrusion 2	20,063	26,652	132.8	33,517	167.1
Intrusion 3	30,256	24,387	80.6	30,615	101.2
Oxide	230,518	225,476	97.8	227,120	98.5
Sulphide	179,084	183,400	102.4	169,051	94.4

The summation of the intrusion volumes for the PTh algorithm is 56,624 blocks compared to 55,898 blocks for the NN model volume. In contrast, the PH method returns a much larger volume of 69,329 blocks. Although there is variation between the intrusive domains in the PTh method, on the whole, the intrusion volumes are near (101.3% of NN volume) compared to the PH method (124.0% of NN volume). The volumes of each method and NN-model volumes are in Table 7.2 and the respective 3-D models are in Figure 7.8.

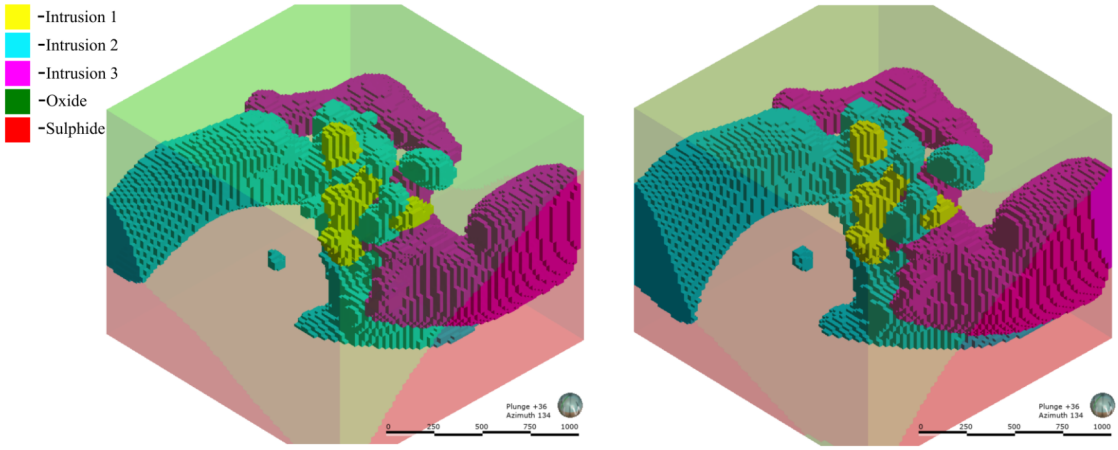


Figure 7.8: Final multicategory models from Probablity-Threshold selection method (*left*) and Probability-High selection method (*right*)

The selection methods in Figure 7.8 are similar, but the PH selection of Intrusion 2 is expansive in comparison with more pronounced edge effects. The PTh selection method is chosen for the final boundary model to undergo uncertainty assessment

because of the proximity of the global volume to the NN-model volumes for the porphyritic intrusions.

The combining of multiple binary models into a single model is an important step in the multi-categorical boundary modeling process. Cross-cutting relationships and geochronology could be incorporated into a site-specific selection method. The imparting of geological interpretations, in concert with the modifying selection methods described above, will result in a geologically realistic boundary model for a given deposit.

7.2 Boundary Uncertainty

Access to uncertainty in a multi-categorical model is not possible for all categories because of the overlap between the uncertainty bandwidths. Therefore, for uncertainty assessment, Intrusion 1 is selected as the mineralized domain of interest. Further thresholding of the Intrusion 1 domain's indicator estimate occurs to extract dilated and eroded interfaces. Volume uncertainty with respect to a +/-0.15 uncertainty threshold is illustrated in Figure 7.9. The smallest model volume is 3,693 blocks, while the maximum volume model is 8,381 blocks. The histogram shows the effect of uncertainty thresholds between z+0.15 and z-0.15 on the distribution of model volumes.

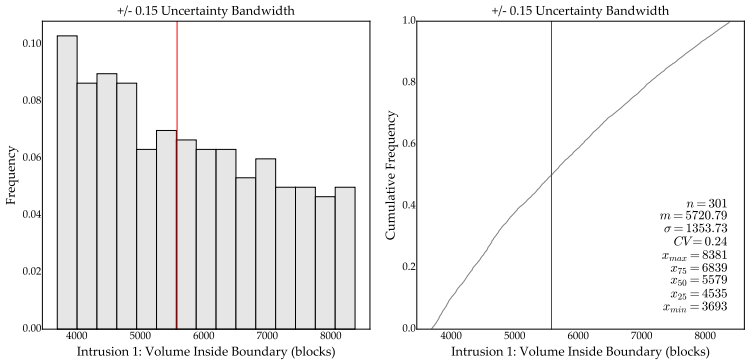


Figure 7.9: Histogram and CDF of Intrusion 1 volume uncertainty with red line signifying the volume of the NN-model volume

Plan and section views of the final model for Intrusion 1 uncertainty are in Figure 7.10. Intrusion 1 forms the core of the deposit. Tight drilling in plan view results in smaller bandwidth uncertainty; however, section views highlight zones with varying bandwidth. The categorical 3-D model consisting of a base-case, eroded case, and dilated case for Intrusion 1 along with the other domains' base cases is in Figure 7.11. The oxide and sulphide domains are faint in order for the structure of the intrusions to be viewed. Knowledge of the geological model of deposition for Intrusions 2 and 3 could control over extrapolation by trimming the models or changing the maximum distance parameter in the NN modeling. The Intrusion 1 base-case volume is 5,599 blocks, inclusive of the eroded model volume. The eroded boundary volume is 3,693 while the dilated boundary volume is inclusive of the base-case and eroded volumes at 8,402 blocks. The range of uncertainty from the outside edge of the eroded boundary to the outside edge of the dilated boundary is 4,709 blocks or 56% of the Intrusion 1 uncertainty block model.

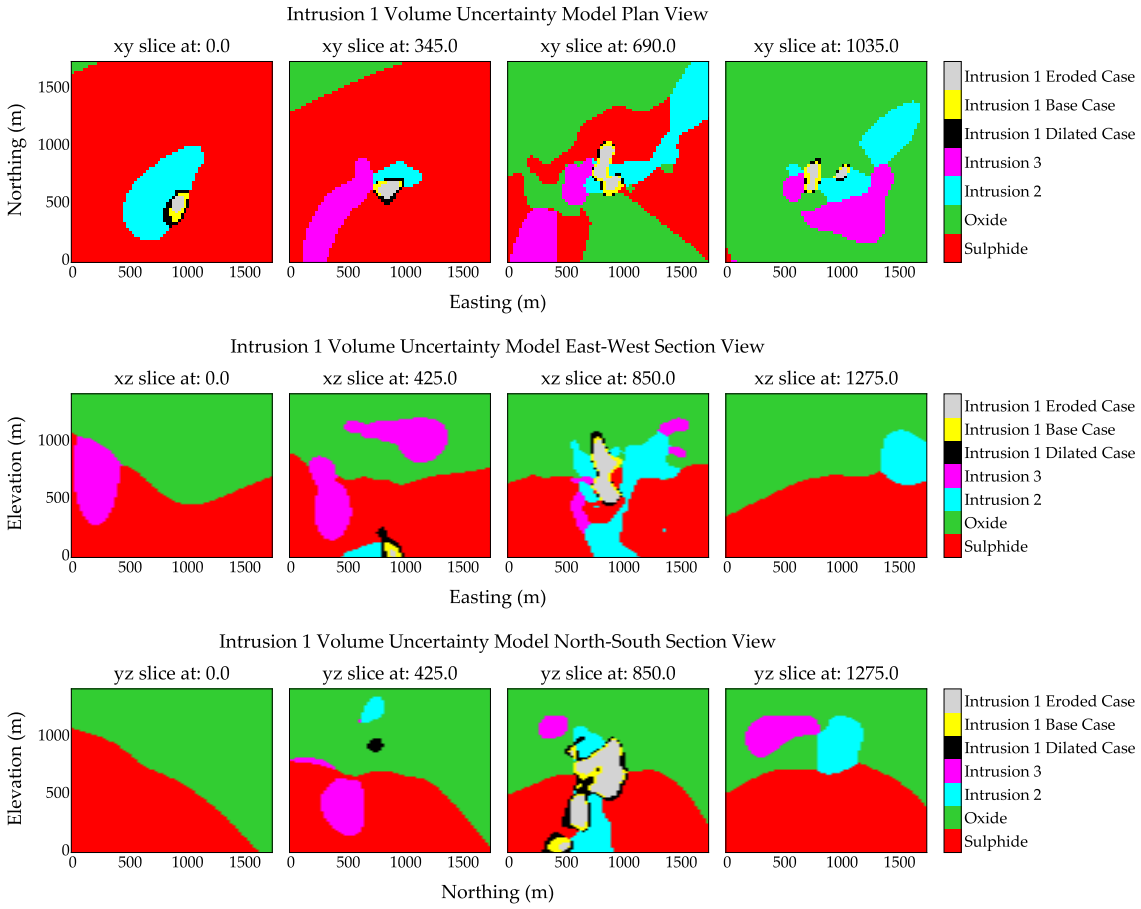


Figure 7.10: Plan view and Sections of Categorical model with volume uncertainty for mineralized domain Intrusion 1

The model for the dilated case expands from the base case of Intrusion 1 in the multi-categorical model without uncertainty. Therefore, some grid nodes, previously categorized as one of the other four domains, are redefined. A total of 2,817 blocks recategorize in the uncertainty model; the majority of which (36.7%) are from Intrusion 2. The Sulphide domain accounts for 33.5% of the changes. The Oxide and Intrusion 3 change at a rate of 21.5% and 8.2%, respectively.

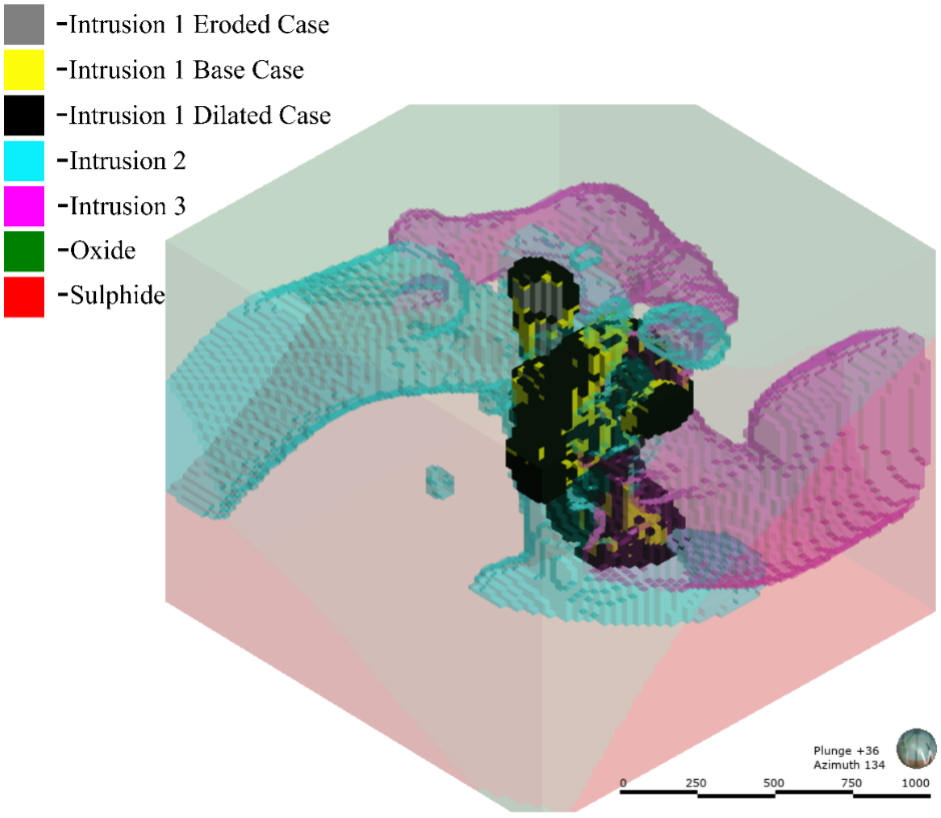


Figure 7.11: Isometric view of final boundary model with uncertainty for Intrusion 1: globally unbiased base-case (*yellow*), eroded case (*light grey*), and dilated case (*black*). Intrusion 1 and 2 are highlighted with the Oxide and Sulphide domains being translucent for viewing purposes.

7.3 Comparison to Multicategorical SDF Modeling

A comparison of the multiple indicator threshold workflow to an equivalent multicategory SDF model illustrates the efficacy of the proposed methodology. Uncertainty bandwidths are an essential aspect of boundary models for calculating the tonnage uncertainty. K-fold analysis and misclassification metrics measure the performance of the proposed multiple indicator threshold technique.

7.3.1 Boundary Models

Multicategory estimation for both workflows is interpolated with RBFs using isotropic Gaussian kernels. The multicategory SDF workflow follows the methodology of D. Silva (2015). The SDF algorithm establishes a boundary by extracting the isozero interfaces of the categories where the model transitions from one category's negative values to another's negative values. For grid nodes with multiple negative SDF values, the category with the most negative value prevails. The multiple indicator boundary extractions follow the probability-threshold workflow outlined in the previous sections (7.1). The multicategory models are seen in Figure 7.12. The SDF global volumes for specific domains are compared to the multiple indicator threshold model volumes in Table 7.3.

Table 7.3: Volumes for multi-category domains using NN, Indicator Thresholds, and SDF methods

Domain	**Volumes (blocks)**				
	NN-model	**Indicator Threshold**	**% NN**	**SDF**	**% NN**
Intrusion 1	5,579	5,585	100.1	4,192	75.1
Intrusion 2	20,063	26,652	132.8	16,863	84.1
Intrusion 3	30,256	24,387	80.6	18,651	61.6
Oxide	230,518	225,476	97.8	244,961	106.3
Sulphide	179,084	183,400	102.4	180,833	101.0

The SDF intrusion models show significant conservative bias in terms of global volumes. Moreover, discrepancies exist between the NN model volumes and multiple indicator threshold models, with Intrusion 2 being overestimated and Intrusion 3 underestimated. The summation of the SDF intrusion volumes is 39,706 blocks. In contrast, the NN-model cumulative volume is 55,898 blocks and indicator threshold volume is 56,624 blocks. Therefore, although differences arise in the indicator threshold model volumes for each intrusion, on the whole, the global volume of the porphyritic domains is near the unbiased global volume determined from NN-modeling.

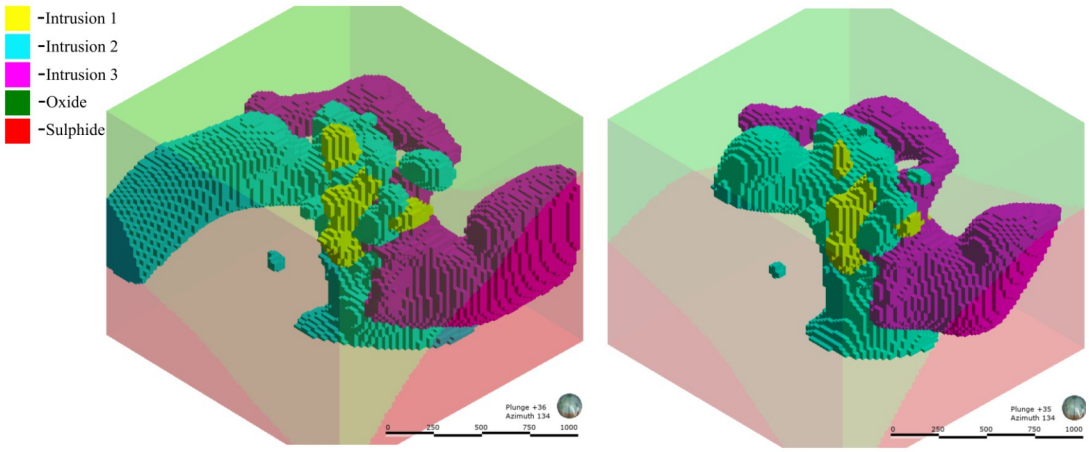


Figure 7.12: Indicator threshold model (*left*) and SDF interpolated model (*right*)

Figure 7.12 shows the final models for the two workflows. The indicator threshold model shows more extrapolation effects than the SDF workflow. Most noticeably, Intrusion 2 (*cyan*) is markedly different between the models. The indicator threshold model extrapolates out to the left edge in contrast to the SDF model that is reserved. A geo-modeler with site-specific knowledge could trim the extrapolation of Intrusion 2 in order to have the model conform to the conceptual geological interpretations.

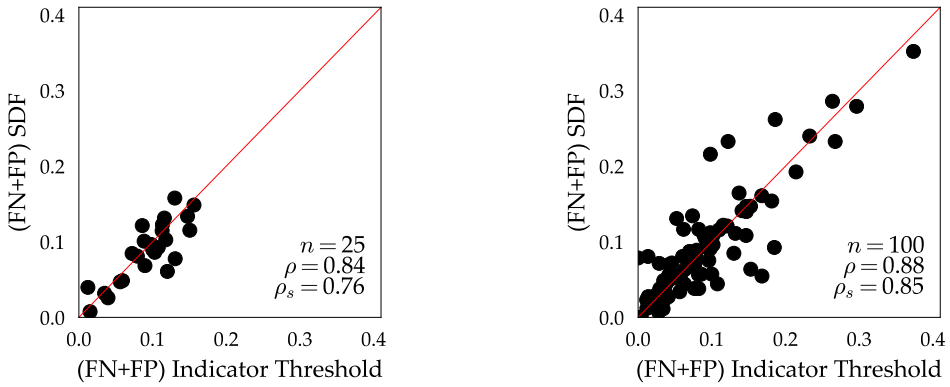
Intrusion 1, the mineralized domain, is the smallest domain and forms the core of the porphyry deposit with the highest density of drilling. Therefore, it has the highest density drilling per unit volume in combination with being the most tightly constrained by 'outside' data. The indicator threshold model for Intrusion 1 is 100.1% of the NN-model volume, whereas the SDF model is 75.1% highlighting the conservative bias inherent to the SDF.

7.3.2 K-Fold Analysis

For the multi-categorical case study, K-fold analysis is undertaken for five folds and 20 folds. The process follows the same procedure explained in the single domain modeling case study in Chapter 6. Each fold in the 5 K-fold series contains between 651-659 data, whereas the 20 K-fold contains between 140-238 data in its subsets. The two workflows for multi-categorical boundary modeling are compared and contrasted.

7.3.2.1 Error

Error is calculated by testing the validation model versus the left-out data. The error is the sum of the False Positive and False Negatives misclassifications from a specific K-fold and its validation subset. The summed values are standardized by dividing by the total number of validation data in a subset, with a value of zero signifying no misclassification. In contrast, a value of 1 indicates a universal disagreement between the model and the left-out data. The product is a measure of misclassification rates and efficiencies. The classification of errors from K-fold analysis in drillhole data is in Figure 6.12.



(a) 5 K-folds for 5 categories

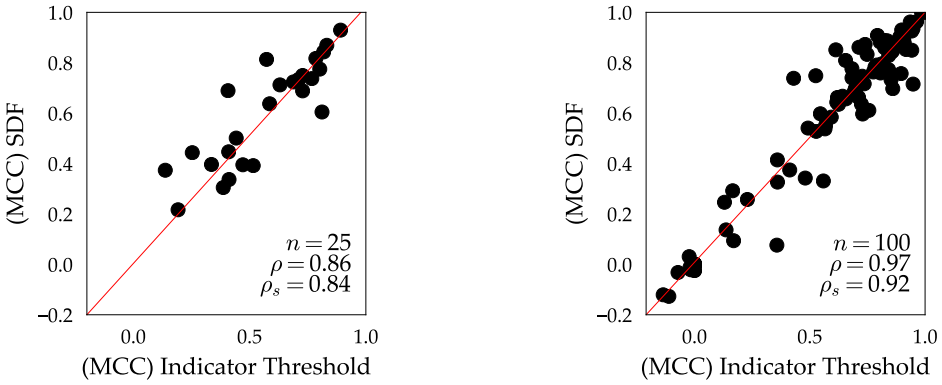
(b) 20 K-folds for 5 categories

Figure 7.13: Error comparison for multicategorical modeling using SDF and Indicator Threshold workflows

In the 5-fold analysis the SDF performed better 64% of the time with an average error rate of 0.087 compared to 0.093 for the indicator workflow. For the 20-fold analysis, the indicator threshold outperformed the SDF 64% of the time, with an average error rate per fold of 0.081 compared to 0.082. Error rate comparisons on a per fold and per-category basis are seen in Figure 7.13. The small difference in error between workflows is evidence that the indicator threshold methodology is robust.

7.3.2.2 Matthews Correlation Coefficient

Another metric for classifying the quality of error rates between models is the Matthews correlation coefficient (MCC). For multi-categorical modeling, the binary classification is for correctly predicting subset data as inside or outside of the appropriate domain. Figure 6.14 shows the confusion matrix that defines the correct and incorrect classifications that are input to Equation 6.1. An MCC value of 1 indicates that the models are in total agreement, whereas a value of -1 signifies that the models are in total disagreement. The measure of how related the performance of the models varies within [-1,1] with zero meaning they are no better than randomly predicted. The outperforming method consists of higher MCC values.



(a) 5 K-folds **(b)** 20 K-folds

Figure 7.14: MCC comparison for multicategorical SDF and multiple Indicator Thresholding techniques

The MCC results comparing the competing workflows over K-folds are in Figure 7.14. The multicategory SDF outperforms the multiple indicator workflow in the 5-fold analysis 68% of the time, with an average MCC value of 0.61 compared to 0.57 (Figure 7.14a). The multiple indicator thresholding workflow equals or bests the SDF in 64% of the instances in the 20-fold analysis (Figure 7.14b). The 20-fold average of the MCC values for the multiple indicator workflow is 0.531 in comparison to the SDF average of 0.532. The results between the fold analyses indicate the workflows compete performance-wise with an increase in informing data (20 K-folds) favouring

the multiple indicator methodolgy.

7.3.2.3 Outliers

The Oxide domain and Intrusion 1 have error values from Fold-13 that are comparatively high. The outlier is in the upper right of Figure 7.13b and coincides with an error value for the Oxide domain of 0.373 and 0.352 for the multiple indicator and multi-categorical SDF, respectively. Moreover, the error values for Intrusion 1 are also high at 0.264 (multiple indicator) and 0.286 (multi-categorical SDF). This particular fold is the same problematic subset from the single domain boundary modeling 20-fold analysis (Section 6.3). The fold subsets a validation drillhole comprised primarily of the Oxide, Sulphide, and Intrusion 2 domains on the periphery of Intrusion 1, the core of the porphyry. Multiple training holes of Intrusion 1 data in the proximity lead to numerous misclassifications in both the workflows. Figure 7.15 illustrates the issues in section view. The False-Positive misclassifications (*white*) occur in two areas: (1) where Intrusion 1 is predicted instead of the correct Sulphide domain, and (2) where the Oxide domain is predicted instead of the correct Intrusion 2. Another misclassification exists (*black*) where Intrusion 2 is predicted and where the validation data is Oxide. The validation drillhole, at the deposit center and adjacent to abundant Intrusion 1 data, leads to high error in both of the boundary models.

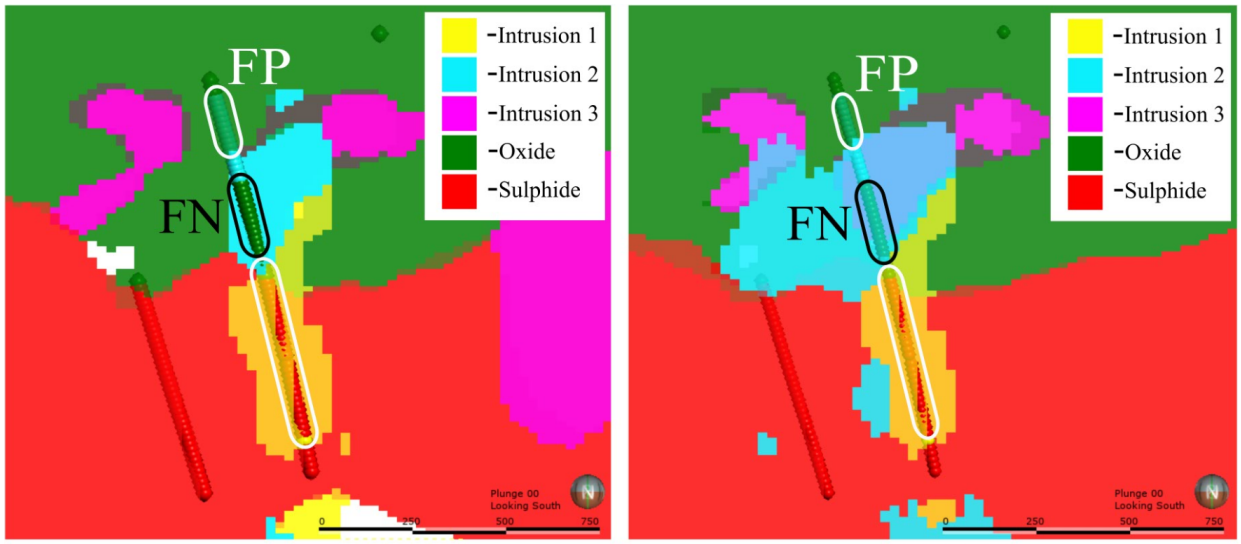


Figure 7.15: Fold 13 training model for multiple indicator (*left*) and multicategorical SDF (*right*) methods along with validation drill hole data. False-Negative and False-Positive zones are outlined in black and white, respectively. The numerous misclassifications lead to high error rates and low MCC coefficients for both the SDF & Indicator threshold models

7.3.3 Boundary Uncertainty

Boundary uncertainty for the multicategory modeling methods is difficult due to overlapping zones of uncertainty between the domains. Therefore, the mineralized domain, Intrusion 1, is used as the domain for comparing uncertainty methodology. The two uncertainty models for Intrusion 1 are seen in Figure 7.19. For the SDF boundary uncertainty for Intrusion 1, the C-parameter calibration uses a jacknife workflow, as discussed in Chapter 1. SDF models for uncertainty with increasing C-values are seen in plan view Figure 7.16. The uncertainty volumes are summarized in Table 7.4.

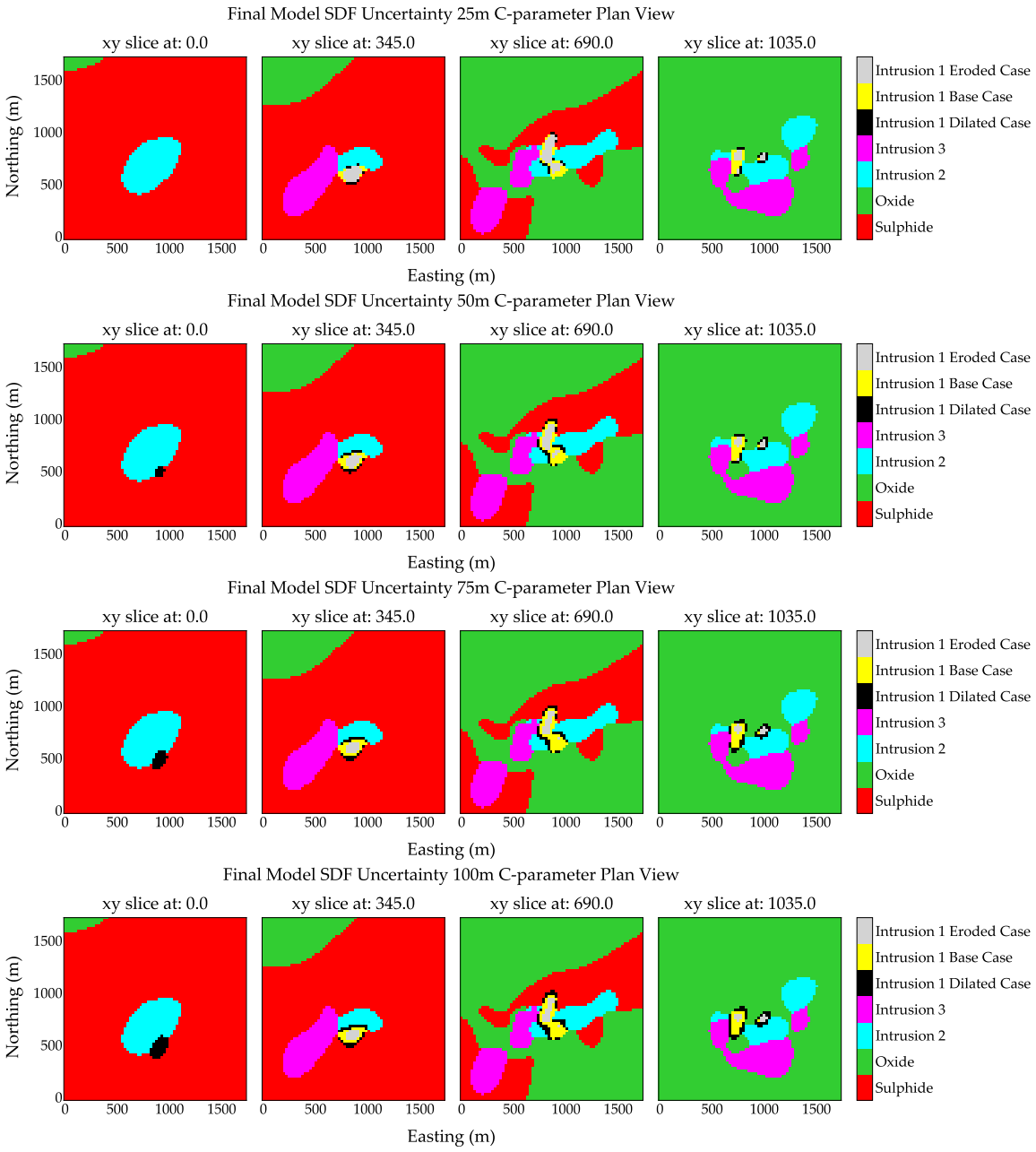


Figure 7.16: Plan view sections of multicategorical SDF models with varying C-parameter values for uncertainty assessment for Intrusion 1

Table 7.4: C-values and corresponding Intrusion 1 volumes for SDF uncertainty

C-value (m)	Uncertainty Volumes (blocks)			
	Eroded	**Dilated**	**Difference**	**% Difference**
25	2,687	5,444	2,757	72.8
50	2210	6,874	4,664	57.8
75	1,921	8,044	6,123	48.2
100	1724	9,026	7,302	41.6

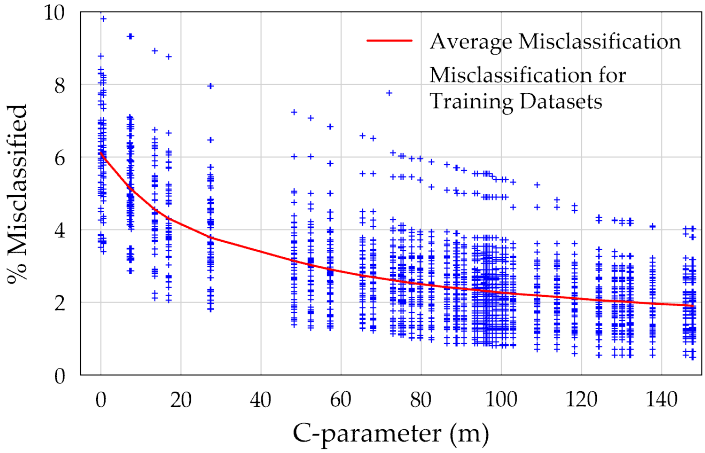


Figure 7.17: C-parameter calibration for Intrusion 1

Figure 7.17 shows the amount of misclassified data for varying C-parameters. The methodology follows the jacknife calibration from Wilde and Deutsch (2012). The final SDF uncertainty model execution is with a C-value of 50m whereby the misclassified data is around 3%, the dilated and eroded global volumes are reasonable and comparable to the indicator threshold volumes, and model inspection identifies realistic uncertainty (Figure 7.16). The C-parameter of 50m also corresponds to the single domain boundary uncertainty value used in Chapter 6.

Uncertainty for the multi-indicator threshold model follows 0.15 thresholds taken above and below the NN-model unbiased threshold, as discussed in Section 7.2. Plan and section views of the respective dilated and eroded cases for modeling workflows are in Figure 7.19. The indicator threshold model's uncertainty bandwidth expands in areas with less informing outside data, notably at depth in plan view. The bandwidth

extensions in plan views occur where data configurations are less concentrated with higher variation across the drill hole spacing, as seen in Figure 7.18. Both uncertainty models behave comparatively well in zones with plenty of nearby informing data and tight spacing; however, a noticeable conservative bias for the SDF is evident in zones where data spacing increases or areas where no data is present.

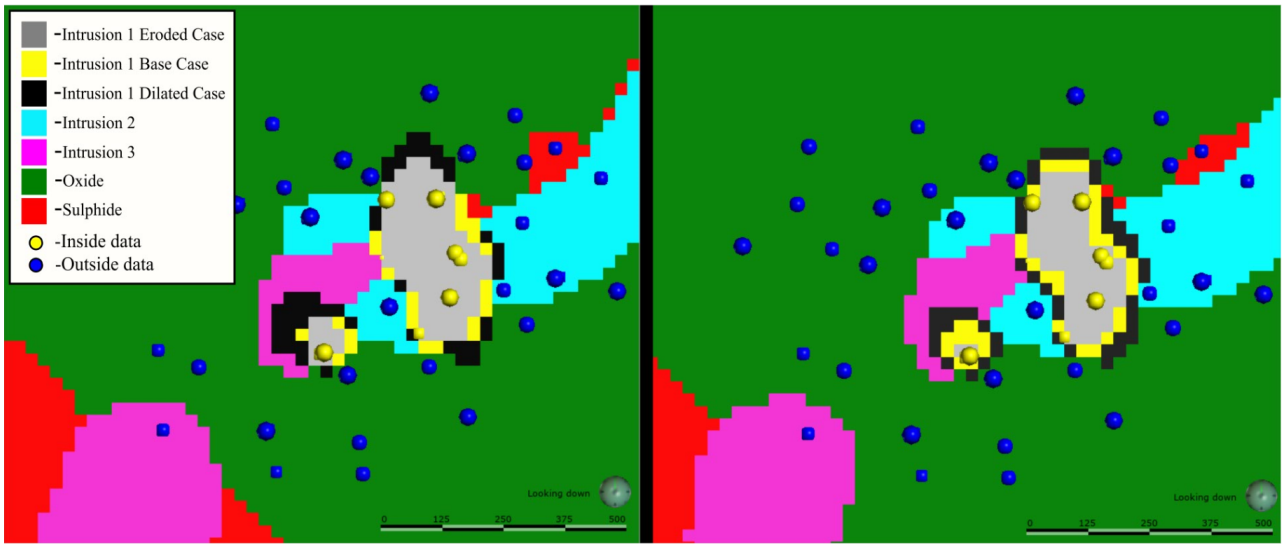
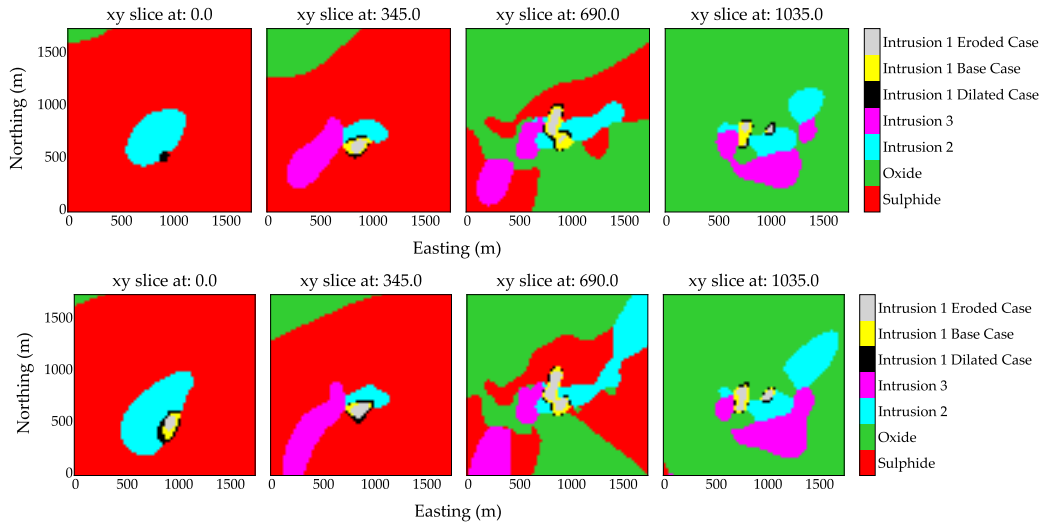


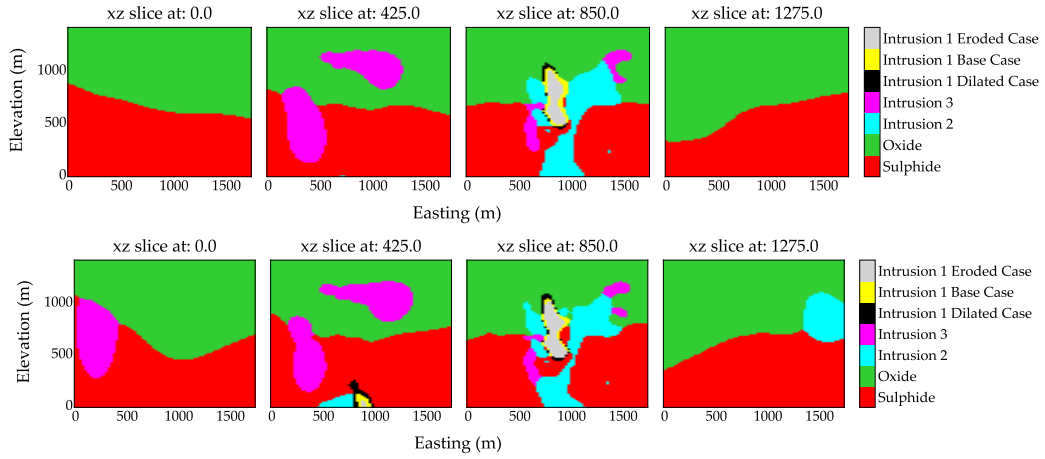
Figure 7.18: Plan view of Indicator threshold model for uncertainty (*left*) and SDF model for uncertainty (*right*) for Intrusion1 showing constant bandwidth uncertainty in SDF modeling

Interpolation of indicators results in local uncertainty, which manifests through probabilities of a given category at a specific node. These probabilities in the indicator threshold workflow are thresholded for boundary extraction. The resulting bandwidth uncertainty closely follows the structure of the conditioning data. In contrast, the SDF workflow follows the additive C-parameter constant for interpolation of uncertainty bandwidths from dilated and eroded boundary models. Figure 7.18 illustrates the SDF issue with the constant C-parameter bandwidths. On the left side of the domain, where widely spaced data exists, the indicator threshold model shows larger variation through its uncertainty bandwidth. In contrast, the SDF consistent bandwidth is not following the structure of the local conditioning data and is more uniform in width. The upper part of Figure 7.18, where the data spacing widens, the SDF model exhibits a comparatively constant bandwidth with less realistic uncertainty.

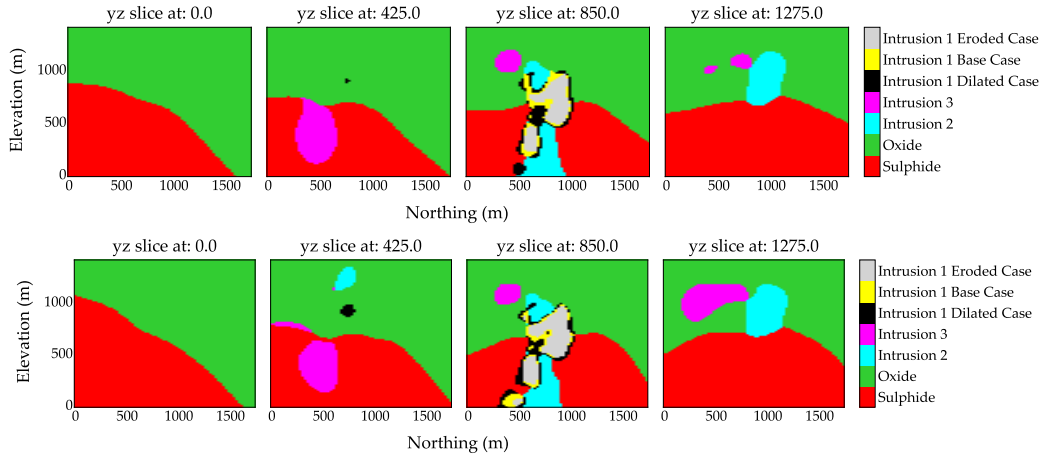
Globally, the indicator threshold dilated boundary is 8,402 blocks compared to the 6,945 blocks comprising the SDF dilated case. Furthermore, the eroded boundary global volumes for the indicator threshold and SDF methods are 2,210 and 3,693 blocks, respectively. For the multi-categorical case, the uncertainty from indicator thresholding gives more substantial bandwidth uncertainty and closely follows the structure of the data, resulting in a realistic model for uncertainty.



(a) Plan view slices of multicategorical SDF uncertainty (*above*) & Indicator Threshold uncertainty (*below*) for Intrusion 1



(b) East-West section view slices of multicategorical SDF uncertainty (*above*) & Indicator Threshold uncertainty (*below*) for Intrusion 1



(c) North-South section view slices of multicategorical SDF uncertainty (*above*) & Indicator Threshold uncertainty (*below*) for Intrusion 1

Figure 7.19: SDF & Indicator Threshold multicategorical model for Intrusion 1 uncertainty

7.4 Results and Considerations

Lithological data from a multi-phase porphyry deposit is used to construct a multi-categorical boundary model. Nearest Neighbour models yield globally unbiased volumes for their respective domains. Probabilities are mapped throughout the domain by global methods of indicator interpolation using an RBF with a Gaussian kernel. The multiple indicator estimates are thresholded to their respective NN model volumes ratios resulting in five base case boundary models. The base case models are combined into a single categorical model. The overlap between categories in the grid is mitigated by selecting the lithology at a specific node that has the highest Probability-Threshold differential. Figure 7.20 shows the volume uncertainty for competing workflows.

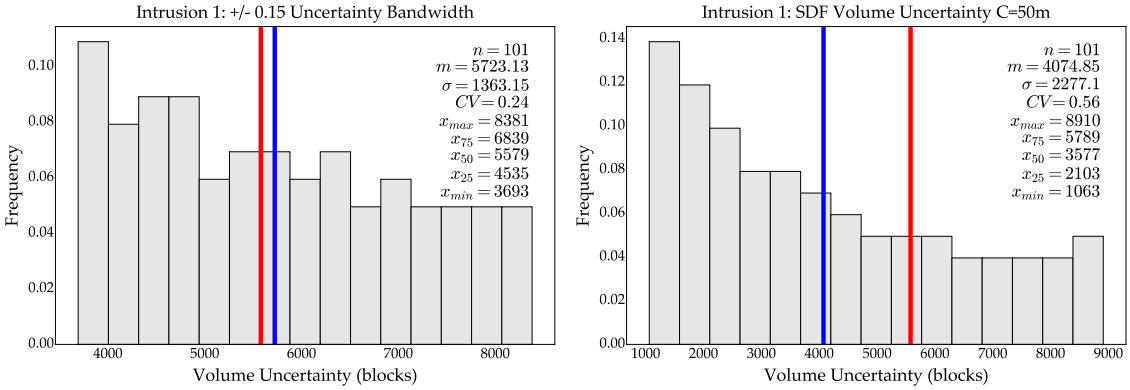


Figure 7.20: Comparison of Intrusion 1 volume uncertainty for Indicator Threshold (*left*) and SDF (*right*) workflows for Intrusion 1 with their respective mean volumes seen as blue lines. The NN model volume of 5,579 blocks is shown by red line and is close to the Indicator workflow volume uncertainty mean of 5,723 blocks. The SDF volume uncertainty mean is 4,075 blocks. The distribution of volume uncertainty and the abrupt transitions at higher and lower thresholds are an area for future research. The frequency of model volumes in the tails of the distribution indicate that both workflows result in models with significant uncertainty.

A mineralized domain is chosen for uncertainty assessment. Uncertainty is accessed by thresholding the indicator estimate 0.15 above and below the NN threshold. The final multi-categorical model consists of an uncertainty model for the mineralized intrusion and boundaries for the surrounding lithologies. Comparing the workflow to SDF mod-

eling over numerous K-fold testing shows that the methodology is robust. Error and MCC comparisons show the advantages and disadvantages of competing workflows. For the 20-fold models with more conditioning data, the indicator workflow bests the SDF modeling; however, in the 5-fold training models with less conditioning data, the SDF prevails. Extrapolation issues in both models can be mitigated by a geomodeler with site-specific knowledge of the local geology. In zones of data asymmetry or sparse drilling, the indicator thresholding workflow shows distinct advantages. The SDF approach, in the presence of data asymmetry, leads to conservative bias in the boundary placement. These biases are introduced as the initial SDF values are calculated for the conditioning data and pass through the entire workflow. By comparison, the indicator model boundaries closely adhere to the structure of the conditioning data and appear more realistic. Uncertainty assessment for a chosen lithology results in a bandwidth of uncertainty that strictly honours local conditioning data structure. The bandwidth expands in low-density drilling areas and contracts in zones of concentrated drilling. The final model checking and validation using K-fold analysis assess error and MCC values.

The efficacy of the multi-categorical algorithm is assessed based on the concepts of simplicity, speed, objectivity, data integration, access to uncertainty, and geologic realism (McLennan, 2007). The simple and easy to understand methodology in concert with the speed of the interpolation thresholding make the algorithm appealing. The implicit technique mitigates subjectivity issues inherent to explicit modeling. Model reproduction is natural when parameters are held constant. With additional data, the integration of new information into the model is fast and efficient. The indicator estimation maps a field of probabilities with easy local uncertainty assessment. The thresholding of these estimates gives access to a realistic bandwidth of uncertainty. For an implicit model to be geologically reasonable, a modeler must be able to confirm observations and impart interpretations made in the field onto the model. The integration of field measurements and orientation data is possible with RBFs (Martin & Boisvert, 2017). The final models are robust. Testing through K-fold analysis

shows the multiple indicator thresholding methodology competes with the implicit multi-categorical SDF workflow. Distinct advantages in the indicator approach arise in areas of sparse data, and data asymmetry. The final model consists of multiple domains with approximate globally unbiased volumes and a core mineralized domain with uncertainty.

Chapter 8
Conclusions

A novel approach to boundary modeling using thresholded indicator estimates is presented. The workflow is explained and illustrated through multiple case studies. The relationship between probability and thresholds provides insight into uncertainty. The contribution of this thesis and issues with boundary modeling are reviewed in this chapter. Boundary modeling in mining and mineral exploration presents many challenges; These limitations are expressed, and future research avenues are suggested.

8.1 Review and Contributions

Boundaries represent an interface between geological domains. In mining, the proper placement of boundaries is essential across an array of boundary types. Grade shells, lithologies, alteration, structures are all necessary features that require accurate and precise measurements in space. The sparse data involved in mining leads to uncertainty in the location of boundaries. The quantification of uncertainty is imperative for the calculation of resources and subsequent decision making. Current practices often use the SDF for boundary definition. However, the SDF falters in the presence of asymmetries in the conditioning data. These asymmetries are prevalent in drill hole data and may lead to the introduction of conservative bias to the model. Moreover, the SDF methodology gives a simplistic bandwidth for uncertainty. Often the uncertainty bandwidth appears constant in areas where data are variably spaced. A new indicator threshold approach is proposed and demonstrated.

Deterministic and stochastic approaches to boundary modeling are reviewed. The uncertainty quantification, if applicable, for the different methods, are further explored. Explicit approaches consist of the manual digitization of 2D geological sec-

tions into 3D wireframes generating a volumetric solid. The models are subjective, time-consuming, and hard to update. Implicit boundary definition is faster, being data-driven and automatic. The technique models wireframes using distance functions that use the conditioning data structure, anisotropy and category present. Both deterministic methods do not give direct access to uncertainty and do not replicate the proper small scale variability inherent to earth sciences. Kriging is a popular interpolant; however, it does not give access to joint uncertainty. Stochastic methods have more realistic variability on smaller scales yet lack larger scale variability. Process-based boundary modeling uses a physical model of deposition to help define boundary structure. Object-based modeling simulates geological features by using primitive objects from a library that are placed in a matrix of geology based code. Sequential Indicator Simulation (SIS) uses the conditioning data to estimate the probability of a category being present at an unsampled location. Truncated Plurigaussian modeling sets out geological relationships between categories and assigns a range on the Gaussian distribution from which values are simulated and extracted. Multi-Point Statistics (MPS) uses training images. An image is chosen that represents the targeted structure of the geology and reproduces these features in the model.

An in-depth analysis of the SDF methodology is conducted. Issues arising from SDF modeling are outlined and explained. In the presence of asymmetries in the data configuration, the SDF will introduce a conservative bias. The algorithm selects the closest sample of opposite indicator to calculate the distance function value. Often in earth sciences, a particular contact will be selectively drilled, leading to asymmetries in the collected data. The result is a boundary that is interpolated from distance function values that may not represent the values pertinent to the boundary with wider data spacing. This error is introduced in the initial step of the SDF and passed through the entire workflow. Access to uncertainty in the SDF methodology is through the C-parameter. The C-value selection follows a jackknife procedure where an acceptable misclassification rate is reached. The C-value is added to the distance function positive values and subtracted from the negative values. The modified dis-

tance function interpolation yields an eroded boundary and an dilated boundary with bandwidth of uncertainty. The global additive nature of the C-parameter leads unrealistic boundary uncertainty. Often the bandwidth is constant and uniform in areas of variable data configurations.

The proposed indicator thresholding workflow offers an alternative boundary modeling methodology. The workflow consists of three main steps: indicator estimation, NN-thresholding, and uncertainty extraction. The indicator estimation with a global interpolator guarantees an artifact-free model. Radial Basis Functions are advantageous as no variogram analysis is needed, the result is artifact-free, and the algorithm is fast and reliable. The Gaussian kernel replicates short-scale continuity well, resulting in smooth models. The interpolator maps a field of probabilities through the domain providing straightforward assessment of uncertainty. The generation of the unbiased spatial configuration of conditioning data comes from nearest neighbour modeling. Each unsampled location is assigned the indicator value of the nearest sample. The mean of the distribution of indicator values yields a globally unbiased volume for the data configuration. The NN model volume is used to threshold the indicator estimate CDF. The extracted z-value from the CDF corresponds to an equivalent volume from the indicator estimate that is globally unbiased. All indicator estimate values equal to the z-value form the boundary. The probabilities above the z-value are inside of the boundary. The indicator interpolation and thresholding mitigate the conservative bias attributed to the SDF as the global interpolant considers all conditioning indicator data when informing the estimate. The result is a boundary model that closely follows the structure of the data. For uncertainty, eroded and dilated cases are established. The z-value from the base case scenario is modified +0.15 to capture an eroded boundary, and -0.15 to establish a dilated boundary. Careful inspection of the uncertainty bandwidth by the geomodeler should follow, and if the uncertainty appears unrealistic, the modification value can be altered to reflect site-specific interpretations. The final model is compared and contrasted to an equivalent SDF boundary model through K-fold analysis and shown to outperform

in some instances and underperform in others. The algorithm is shown to be robust.

The indicator threshold workflow is modified for multiple category datasets. Independent indicator interpolation for each category gives k-category estimates. The k categories respective NN model volumes are calculated, and the indicator estimates are thresholded accordingly. For instances where two or more categories exist at a node, the category with the largest discrepancy between indicator probability and associated threshold prevails. The resulting multicategorical boundary model is unique. Uncertainty assessment follows the same general methodology; however, due to dilated boundaries overlapping, the assessment must be done on a single category or categories that do not come in contact with one another. The chosen category for uncertainty assessment is further thresholded at +/-0.15 of the NN threshold. The uncertainty bandwidth extends into the surrounding domains, whereby the initial models are modified. The final result is a multicategorical boundary model with uncertainty assessment. The boundary model is compared and contrasted to an equivalent SDF workflow through K-fold validation. The methodology replaces the SDF workflow and mitigates issues arising from SDF biases.

Research into the threshold values for uncertainty bandwidths is conducted. The methodology for constructing Probability-Threshold Curves (PTC) outlines a novel approach to understanding the relationship between the indicator estimates and the threshold values used for boundary extraction. The simulation of true scenarios that are sampled, interpolated, and studied presents an avenue to understand boundary model uncertainty. The N truths are sampled at varying spacings in plan and section view. The indicators are interpolated using global kriging and RBFs. The resulting models are thresholded from p100 to p0 in increments of 0.05. The thresholded model volumes are compared to their respective truths resulting in the probability of a threshold model to be larger than its truth. The probabilities are plotted against their thresholds to arrive at an experimental PTC. By changing structural attributes of the truths, the process is repeated over dozens of scenarios. The standardization of PTCs across hundreds of scenarios drilled at varying spacings is accomplished by

translating the PTCs so that their midpoints are zero. The shape of the curves are near-linear and the distance between their midpoints and where they transition to zero or one spans a zone of uncertainty thresholds predominantly between 0.1 to 0.15.

The main contribution of this thesis is the establishment of a sound and straightforward boundary modeling methodology with uncertainty. The application of proven geostatistical techniques, like indicator estimation, in concert with NN modeling, leads to a reasonable and viable boundary model. Global unbiasedness is important for the estimation of resources and reserves. The easy access to uncertainty greatly aids subsequent decision making. Research into the relationship between indicator estimates and thresholds for boundary extraction is novel. How the uncertainty changes with varying thresholds gives insight into the proper application of the thresholding technique. The methodology is tested using K-fold validation, and boundary models from real geological data are generated. The proposed indicator threshold approach is compared to an equivalent SDF workflow. The algorithm is fast and straightforward. The outputs are not subjective to the geomodeler and are reproducible with the same parameterization. The integration of new data is straightforward. Access to uncertainty is obtained directly. The reproduction of large scale structures is evident, and the models are geologically realistic. Finally, the algorithm is robust. K-fold analysis and comparisons to SDF models show the algorithm produces boundary models that mitigate the conservative bias introduced by the SDF and the unrealistic constant nature of the SDF uncertainty bandwidth. The result is a globally unbiased volume model with fair eroded and dilated boundaries for bandwidths of uncertainty completed in a simple and fast manner.

8.2 Future Work

The proposed methodology for boundary modeling gives rise to certain limitations and areas for future research. The NN modeling is susceptible to edge effects in areas with sparse drilling. Setting maximum distances for extrapolation helps mitigate these issues; however, this introduces subjectivity. In-depth analysis of the effects

on the base case boundary models following the procedure laid out for PTCs could yield promising results, giving insight into the workflow's NN modeling aspect. Furthermore, research into the changing of domain size with respect to the conditioning data and its behaviour and effect on the indicator thresholding would be pertinent.

For indicator estimation, the research in this thesis dealt with homogenous bodies with little anisotropy and omni-horizontal variograms. Further research on the effects of anisotropy on the behaviour of PTCs is important. The PTC is a global measure for uncertainty, and future work into the stationarity of PTCs should be explored. The shape of PTCs was assumed linear after exhaustive studies over numerous geological scenarios. Research into the shape of non-linear PTCs could yield functions used for uncertainty extraction instead of the simplistic +/-0.15 of NN z-value. These characteristics were discussed in Chapter 4, and a database built of parameters for PTCs could be extended. The database could train machine learning algorithms to identify unique functions for site-specific uncertainty quantification. Measures of model characteristics such as tortuousity, volume, and drill spacing to size ratios, could be considered to train machine learning algorithms to extract PTCs. Further research into multicategorical modeling for other types of soft boundaries, such as the geostatistical grouping of common populations could be useful. Moreover, uncertainty in boundary modeling arises from the sparseness of geological data and the quality of the data. This uncertainty is called aleatory and originates from error in the data themselves (Manchuk & Deutsch, 2019). Studies into the effects of aleatory uncertainty on the boundary modeling algorithm is another avenue for future research. Finally, a natural extension of this research is to increase the complexity of the modeled geology. Isotropic porphyry deposits are realistically modeled; however, the modeling of complex hydrothermal vein systems with bifurcating arrays of veinlets and splay structures would be more challenging. Complex systems, in general, require more attentive care from the geomodeler. However, the application of this research ideally would make for easier integration of other geological data sources resulting in more automatic modeling. The distribution of volume uncertainty and

the abrupt transitions at higher and lower thresholds are an area for future research. The frequency of model volumes in the tails of the distribution indicate that the workflow results in models with significant uncertainty. The global and local uncertainty studied through K-fold analysis for competing workflows also illustrates how the SDF methodology performs adequately locally, yet still results in globally biased model volumes. The research contained in this thesis forms a foundational basis from which new and exciting modifications for boundary modeling can be explored. The method is novel, and the application straightforward with promising results thus far.

References

Bridge, J., & Leeder, M. (1979, 10). Boundary modeling with moving least squares. *Sedimentology*, *26*, 617-644.

Caers, J. (2011). *Modeling uncertainty in the earth sciences* (3rd ed., Vol. 1). 111 River Street, Hoboken, NJ, USA: TWiley-Blackwell.

Carr, J. C., Beatson, R. K., Cherrie, J. B., Mitchell, T. J., Fright, W. R., McCallum, B. C., & Evans, T. R. (2001). Reconstruction and representation of 3d objects with radial basis functions. In *Proceedings of the 28th annual conference on computer graphics and interactive techniques* (p. 67–76). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/383259.383266` doi: 10.1145/383259.383266

Carvalho, D. (2018). *Probabilistic resource modeling of vein deposits* (Unpublished master's thesis). University of Alberta, Edmonton.

Cowan, J., Beatson, R., Ross, H., Fright, W., McLennan, T., Evans, T., … Titley, M. (2003, 11). Practical implicit geological modelling..

Deutsch, C. (2018). *Partitioning Drill Hole Data into K Folds* (CCG Annual Report 20). Edmonton AB: University of Alberta. Retrieved from `http://www.ccgalberta.com`

Deutsch, C., & Journel, A. (1998). *Gslib: Geostatistical software library and user's guide* (2nd ed.). New York: Oxford University Press.

Deutsch, C., & Wilde, B. (2013, 06). Modeling multiple coal seams using signed distance functions and global kriging. *International Journal of Coal Geology*, *112*, 87–93. doi: 10.1016/j.coal.2012.11.013

Deutsch, J., & Deutsch, C. (2012). *Accuracy Plots for Categorical Variables* (CCG Annual Report 14). Edmonton AB: University of Alberta. Retrieved from `http://www.ccgalberta.com`

Dimitrakopoulos, R. (1998, 01). Conditional simulation algorithms for modelling orebody uncertainty in open pit optimisation. *International Journal*

of Surface Mining, Reclamation and Environment, 173-179. doi: 10.1080/09208118908944041

Fasshauer, G. E. (2007). Meshfree approximation methods with matlab. In *Interdisciplinary mathematical sciences.*

Fazio, V. S., & Roisenberg, M. (2013). Spatial interpolation: an analytical comparison between kriging and rbf networks. In *Sac '13.*

Galli, A., Beucher, H., Le Loch, G., Doligez, B., & Group, H. (1994). The pros and conds of the truncated gaussian method. *Geostatistical Simulations*, 217-233.

Hadavand, M., & Deutsch, C. (2016). *Updated blocksis for proportion reproduction* (CCG Annual Report 18). Edmonton AB: University of Alberta. Retrieved from `http://www.ccgalberta.com`

Hillier, M., Schetselaar, E., Dekemp, E., & Perron, G. (2014, 07). Three-dimensional modelling of geological surfaces using generalized interpolation with radial basis functions. *Mathematical geosciences*, *46*, 931-953. doi: 10.1007/s11004-014-9540-3

Hosseini, A. (2007). *A Distance Function Based Algorithm to Quantify Uncertainty in Areal Limits* (CCG Annual Report 09). Edmonton AB: University of Alberta. Retrieved from `http://www.ccgalberta.com`

Hosseini, A. (2009). *Probabilistic modeling of natural attenuation of petroleum hydrocarbons* (Unpublished doctoral dissertation). University of Alberta.

Iles, K. (2009, 02). "nearest-tree" estimations. *International Journal of Mathematical and Computational Forestry & Natural-Resource Sciences*, *1*, 47-51.

Journel, A. (1983a). Fundamentals of geostatistics in five lessons. In *Short course presented at the 28th international geological conference.*

Journel, A. (1983b, 06). Nonparametric estimation of spatial distributions. *International Association for Mathematical Geology*, *15*, 445-468.

Knight, R., Lane, R., Ross, H., Abraham, P., & Cowan, J. (2007). Implicit ore delineation. In *In proceedings of exploration 07: Fifth decennial international conference of mineral exploration* (p. 1165-1169).

Li, C., Xu, C., Konwar, K. M., & Fox, M. D. (2006). Fast distance preserving level set

evolution for medical image segmentation. In *2006 9th international conference on control, automation, robotics and vision* (p. 1-7).

Lillah, M., & Boisvert, J. (2012, 08). Stochastic distance based geological boundary modeling with curvilinear features. *Mathematical Geosciences*, *45*. doi: 10.1007/s11004-012-9426-1

Lindsay, M., Ailleres, L., Jessell, M., Dekemp, E., & Betts, P. (2012, 06). Locating and quantifying geological uncertainty in three-dimensional models: Analysis of the gippsland basin, southeastern australia. *Tectonophysics*, *546-547*, 10-27. doi: 10.1016/j.tecto.2012.04.007

Mancell, S., & Deutsch, C. (2019). *The Problem with the Signed Distance Function* (CCG Annual Report 21). Edmonton AB: University of Alberta. Retrieved from `http://www.ccgalberta.com`

Manchuk, J., & Deutsch, C. (2019, 02). Boundary modeling with moving least squares. *Computers & Geosciences*, *126*. doi: 10.1016/j.cageo.2019.02.006

Martin, R. (2019). *Data driven decisions of stationarity for improved numerical modeling in geological environments* (Unpublished doctoral dissertation). University of Alberta.

Martin, R., & Boisvert, J. (2017). *Implicit Modeling* (CCG Guidbook Volume 22). Edmonton AB: University of Alberta. Retrieved from `http://www.ccgalberta.com`

Matthews, B. (1975). Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, *405*(2), 442 - 451. Retrieved from `http://www.sciencedirect.com/science/article/pii/0005279575901099` doi: https://doi.org/10.1016/0005-2795(75)90109-9

McLennan, J. (2007). *The decision of stationarity* (Unpublished doctoral dissertation). University of Alberta.

Michael, H., Li, H., Boucher, A., Sun, T., Caers, J., & Gorelick, S. (2010). Combining geologic-process models and geostatistics for conditional simulation of 3-d subsurface heterogeneity. *Water Resources Research*, *46*, 20.

Munroe, M. (2012). *A methodology for calculating tonnage uncertainty in vein-type deposits* (Unpublished master's thesis). University of Alberta, Edmonton.

Munroe, M., & Deutsch, C. (2008a). *Full calibration of C and B in the framework of vein-type deposit tonnage uncertainty* (CCG Annual Report 10). Edmonton AB: University of Alberta. Retrieved from http://www.ccgalberta.com

Munroe, M., & Deutsch, C. (2008b). *A Methodology for Modeling Vein0Type Deposit Tonnage Uncertainty* (CCG Annual Report 10). Edmonton AB: University of Alberta. Retrieved from http://www.ccgalberta.com

Naus, T. (2008). Unbiased lidar data measurement (draft)..

Pinto, F., & Deutsch, C. (2015). *Software useful for data spacing analysis* (CCG Annual Report 17). Edmonton AB: University of Alberta. Retrieved from http://www.ccgalberta.com

Rossi, M., & Deutsch, C. (2014). *Mineral resource estimation.* doi: 10.1007/978-1-4020-5717-5

Silva, D. (2015). *Enhanced geologic modeling with data-driven training images for improved resources and recoverable reserves* (Unpublished doctoral dissertation). University of Alberta.

Silva, D. A., & Deutsch, C. (2015). *Application and Practical Considerations of Implicit Geologic Signed Distance Function Modeling* (CCG Annual Report 17). Edmonton AB: University of Alberta. Retrieved from http://www.ccgalberta.com

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. Retrieved from https://doi.org/10.1007/978-3-319-78999-6_30 doi: 10.1007/978-3-319-78999-6_30

Wilde, B. (2011). *Programs to Aid the Decision of Stationarity* (CCG Guidbebook Series Volume 14). Edmonton AB: University of Alberta. Retrieved from http://www.ccgalberta.com

Wilde, B., & Deutsch, C. (2011). *A New Way to Calibrate Distance Function Un-*

certainty (CCG Annual Report 13). Edmonton AB: University of Alberta. Retrieved from `http://www.ccgalberta.com`

Wilde, B., & Deutsch, C. (2012). Kriging and simulation in presence of staionary domains: Developments in boundary modeling. *Quantitative Geology and Geostatistics 17*, *112*, 12. doi: 10.1007/978-94-007-4153-9_23