

Simulating Boundary Realizations

Brandon J. Wilde and Clayton V. Deutsch

The boundaries between the stationary domains must be modeled with uncertainty. Interpolating a distance function is a useful method for modeling boundaries with uncertainty. A methodology for calibrating the uncertainty has been proposed. From this calibration, bounding surfaces that are big everywhere or small everywhere can be drawn. There are cases where it is desirable to simulate realizations of the boundary. Simulating values uniformly within the calibrated limits permits this. Wherever the interpolated distance function is greater than the simulated value is outside the domain; wherever the interpolated distance function is less than the simulated value is inside the domain. The transition between these occurrences is where the boundary lies.

Introduction

The spatial extent of stationary a domain is represented by a boundary. The boundary indicates the transition from one domain to another. There is uncertainty in the boundary location away from data. Interpolating a distance function has been shown to be a useful method for estimating the boundary location (McLennan, 2008; Munroe and Deutsch, 2008 a,b; Hosseini, 2009). In its conventional formalism, distance function interpolation is a deterministic method with no access to uncertainty. There are simple modifications that can be made to the method to quantify uncertainty. One method for quantifying the uncertainty includes calibrating an additive parameter, C , using available data in a way similar to the jackknife (Wilde and Deutsch, 2011). The work presented herein extends that method to the simulation of multiple realizations of boundary location. Values are simulated uniformly between $-C$ and C . The interpolated distance function values are then compared to the simulated values. Wherever the interpolated distance function is greater than the simulated value is outside the domain; wherever the interpolated distance function is less than the simulated value is inside the domain. The transition between these occurrences is where the boundary is located. Each realization of simulated values $U(-C,C)$ can locate the boundary in a different position resulting in multiple realizations of the boundary.

Review of Distance Function and C Calibration

Using a distance function to determine the location of a domain boundary starts with data which have been coded as either inside or outside the domain:

$$i(\mathbf{u}_\alpha) = \begin{cases} 1, & \text{if domain present at } \mathbf{u}_\alpha \\ 0, & \text{otherwise} \end{cases}$$

For each sample located at \mathbf{u}_ω the nearest sample location $\mathbf{u}_{\alpha'}$ is determined such that $\text{Min}\{\mathbf{u}_\alpha - \mathbf{u}_{\alpha'}\}$, $i(\mathbf{u}_\alpha) \neq i(\mathbf{u}_{\alpha'})$. The distance between the two locations is the distance function value at location \mathbf{u}_α . If \mathbf{u}_α is within the domain, the distance is set to negative; otherwise the distance is positively signed:

$$\begin{aligned} df(\mathbf{u}_\alpha) &= +(\mathbf{u}_\alpha - \mathbf{u}_{\alpha'}), \text{ if } i(\mathbf{u}_\alpha) = 0 \\ df(\mathbf{u}_\alpha) &= -(\mathbf{u}_\alpha - \mathbf{u}_{\alpha'}), \text{ if } i(\mathbf{u}_\alpha) = 1 \end{aligned}$$

Distance function values vary smoothly from high negative values inside the domain far from the boundary to zero at the boundary location to high positive values outside the domain far from the boundary. Uncertainty in the boundary is accounted for by considering an additive parameter, C , which modifies the distance function. C is added to the distance function for locations outside the domain and is subtracted for locations inside the domain:

$$\begin{aligned} \hat{d}f(\mathbf{u}_\alpha) &= df(\mathbf{u}_\alpha) + C, \text{ if } i(\mathbf{u}_\alpha) = 0 \\ \hat{d}f(\mathbf{u}_\alpha) &= df(\mathbf{u}_\alpha) - C, \text{ if } i(\mathbf{u}_\alpha) = 1 \end{aligned}$$

The C parameter increases the difference between positive and negative distance function values. It is calibrated in a manner similar to the jackknife. A subset of the data is removed and the remaining data are used to estimate distance function at the jackknife locations. The number of jackknife data that fall on the wrong side of the

boundary is reduced as C is increased. C is increased until the number of data falling on the wrong side of the boundary is acceptable.

This process is illustrated in Figure 1. Distance function values are calculated at the sample locations for various values of C . These modified distance function values are then interpolated. The jackknife data are observed whether they fall in the correct domain. For a value of $C=0$ there are two white values which fall in the black domain and four black values which fall in the white domain. Increasing C to 4 reduces the number of white values in the black domain to one and the number of black values in the white domain to 3. Further increasing C to 6 further reduces the number of black values in the white domain to two. C can be further increased until none of the jackknife data fall on the wrong side of the boundary, but this would likely lead to too large uncertainty. Choosing some acceptable quantity of misclassified jackknife data gives more reasonable uncertainty.

Once C has been determined it is used to calculate the modified distance function at all sample locations. The C parameter is applied as shown previously; it is added to positive distance function values and subtracted from negative distance function values. The resulting modified distance function is then interpolated using global kriging. The boundary is considered to fall inside the distance function transition from $-C$ to C . Different boundaries can be extracted by applying a threshold between $-C$ and C . Consider the 1D schematic shown in Figure 2 where the distance function starts positive, transitions to negative, and then increases to positive again. The location of the boundary depends on the distance function threshold used. A threshold of zero gives the base case domain boundary. Choosing a threshold near C yields a dilated domain boundary and choosing a threshold near $-C$ yields an eroded domain boundary.

Simulating Boundary Realizations

There are cases where we want to simulate realizations of the boundary such that it is neither big everywhere nor small everywhere, yet uncertainty in its location is accounted for. This can be done by simulating distance function values uniformly between $-C$ and C . Wherever the interpolated distance function is less than the simulated distance function is inside the domain and wherever the interpolated distance function is greater than the simulated distance function is outside the domain. This is illustrated in **Figure 3** where the same distance function that transitions from positive to negative to positive that was shown in **Figure 2** is shown. Distance function values are simulated uniformly between $-C$ and C . Where the interpolated distance function is less than the simulated distance function is inside the domain; where the interpolated distance function is greater than the simulated distance function is outside the domain. The point where the interpolated distance function is equal to the simulated distance function is the boundary location.

To simulate distance function uniformly between $-C$ and C , a Gaussian deviate, $y'(\mathbf{u})$, can be simulated and transformed according to:

$$df'(\mathbf{u}) = 2C[G^{-1}(y'(\mathbf{u})) - C] \quad \forall \mathbf{u}, \forall l$$

where $df'(\mathbf{u})$ is the simulated distance function value, $y'(\mathbf{u})$ is an unconditionally simulated standard normal value and G^{-1} represents the determination of the standard normal CDF value corresponding to $y'(\mathbf{u})$. Multiplying by $2C$ and subtracting C ensures that the values are between $-C$ and C . It is only necessary to simulate at locations where the interpolated distance function is between $-C$ and C . Interpolated values that are greater than C are surely outside the domain; interpolated values that are less than $-C$ are surely inside the domain. This can reduce the time required to simulate distance function realizations.

Determining whether a location, \mathbf{u} , is inside or outside the domain requires comparing the interpolated distance function with the simulated distance function. Wherever the interpolated distance function is less than the simulated distance function is inside the domain; wherever the interpolated distance function is greater than the simulated distance function is outside the domain:

$$i'(\mathbf{u}) = \begin{cases} \text{inside, if } df'(\mathbf{u}) < \hat{d}f^*(\mathbf{u}) \\ \text{outside, if } df'(\mathbf{u}) \geq \hat{d}f^*(\mathbf{u}) \end{cases}$$

Since a boundary is being simulated, it is recommended that a Gaussian variogram be used in the simulation of the distance function as this model maintains high short-range continuity. A very small positive nugget is recommended for mathematical stability. There is no clear method for determining the range of the variogram model. A short range leads to a more rapidly changing boundary; a longer range creates a smoother boundary. This variogram range could come from geologic calibration from data or observations, visualization and expert judgment, or an analogue deposit. The variogram range can be too small leading to unrealistic transitions between domains as illustrated in **Figure 4**. The effect of the variogram range can be seen in **Figure 5**. The boundary falls within the grey area shown in **Figure 1** for $C=6$. A short range variogram causes the boundary to change rapidly (**Figure 5 left**) while a longer range leads to a smoother boundary (**Figure 5 right**). A too-long variogram range is similar to using a single distance function threshold to locate the boundary; the domain will be big or small everywhere.

Conclusions

Boundaries are used in geostatistical modeling to separate domains which contain different stationary random functions. There is often uncertainty in the location of the boundary between domains. The magnitude of this uncertainty can be calibrated using the existing data. The uncertainty is summarized by the C parameter. Boundary realizations can be created by simulating a uniform deviate between $-C$ and C . A given boundary realization is determined by considering all locations where the interpolated distance function is greater than the simulated distance function as outside the domain. All locations where the interpolated distance function is less than the simulated distance function are inside the domain. The boundary is located at the transition between locations outside and locations inside the domain.

References

Hosseini, A.H., 2009, Probabilistic Modeling of Natural Attenuation of Petroleum Hydrocarbons. *Ph.D. Thesis*, University of Alberta, 359p.
 McLennan, J., 2008. The Decision of Stationarity, *Ph.D. Thesis*, University of Alberta, 191p.
 Munroe, M.J. and Deutsch, C.V., 2008a. A methodology for modeling vein-type deposit tonnage uncertainty. Center for Computational Geostatistics Annual Report 10. University of Alberta. 10p.
 Munroe, M.J. and Deutsch, C.V., 2008b. Full calibration of C and β in the framework of vein-type deposit tonnage uncertainty. Center for Computational Geostatistics Annual Report 10. University of Alberta. 16p.
 Wilde, B.J. and Deutsch, C.V., 2011. A new way to calibrate distance function uncertainty. Center for Computational Geostatistics Annual Report 13. University of Alberta. 9p.

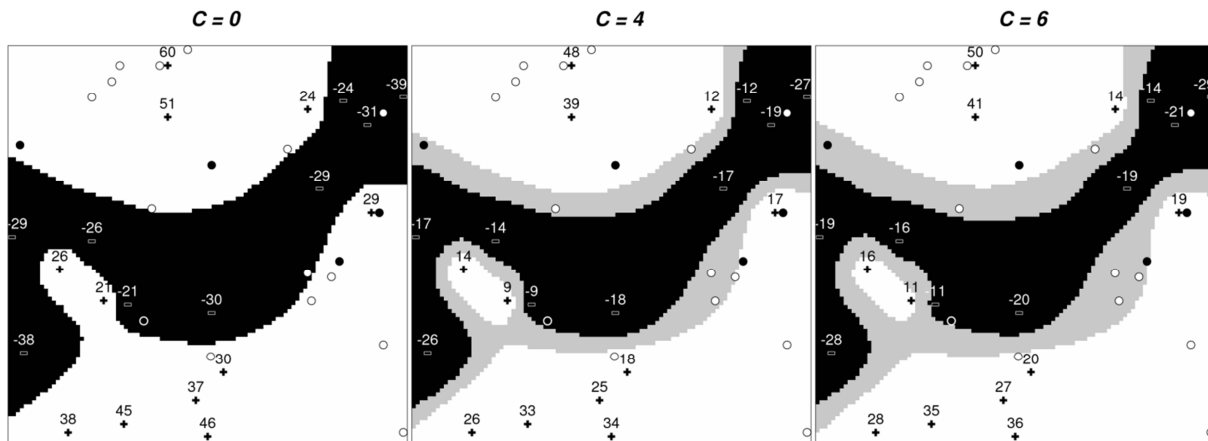


Figure 1: Distance function values are calculated at sample locations for different values of C . Distance function is interpolated. Jackknife data (black and white circles) are used to calibrate C .

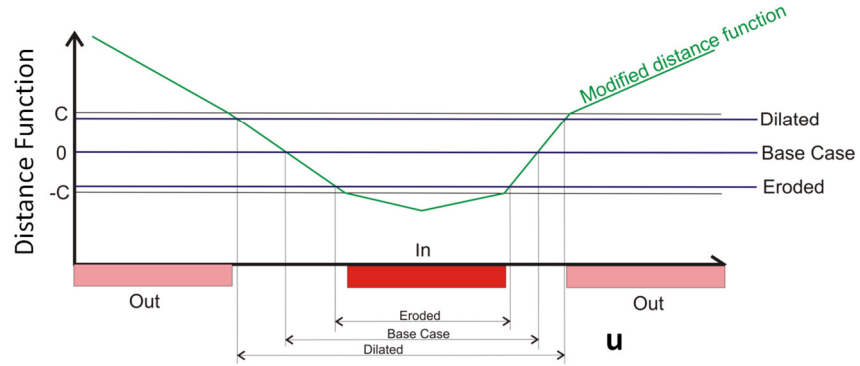


Figure 2: 1D schematic of distance function thresholds applied to arrive at different boundary locations.

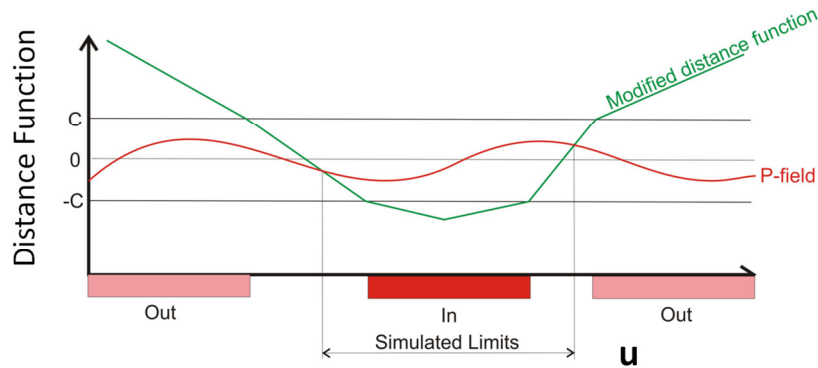


Figure 3: 1D schematic of simulated values determining one realization of the boundary location.

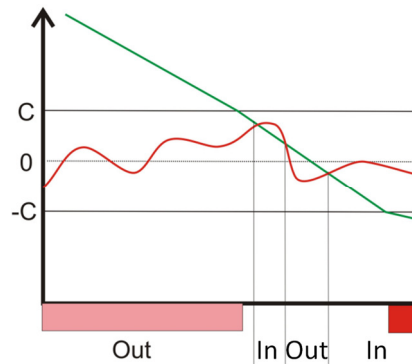


Figure 4: Simulating with a too-short variogram range can lead to unrealistic 'islands' of in or out.

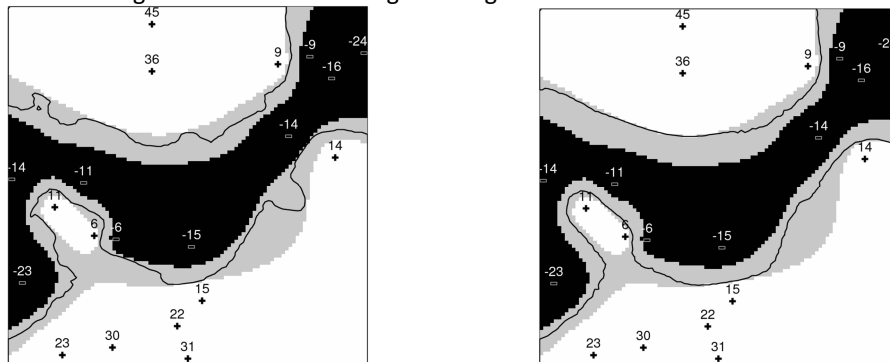


Figure 5: Simulated boundaries for short (left) and long (right) range simulation variogram.