On the Controlled Extrapolation of Geological Anomalies

J.B. Boisvert and C. V. Deutsch

Geostatistical methods such as kriging and sequential Gaussian simulation (SGS) are well suited to the problem of interpolation but often give undesirable results when used for extrapolation. Interpolation is defined as estimating or simulating at an unknown location that is surrounded by data. In contrast, extrapolation is defined as estimating or simulating at an unknown location that is outside the effective convex hull of the data. When extrapolating far from data, the kriging estimate tends towards the global mean and variance. This is inappropriate in many situations. For example, a disseminated deposit often tappers off near the boundary of the deposit. In this situation the kriging estimate should tend towards a mean of 0 rather than the global mean. However, geostatistical models are not often constrained in this manner, especially if there is no data available to delineate the extents of the deposit. In this paper, a correction is made to the kriging estimate such that when extrapolating, the global mean is smoothly reduced away the from convex hull of the data. Sampled locations outside the effective convex hull are adjusted to ensure reproduction of conditioning data after correction. There is no correction inside the convex hull and an increasing correction to the global mean and variance as the distance to the convex hull increases. The correction is smooth and introduces no abrupt discontinuity at the edge of the convex hull. An illustrative example using data from an oil sand deposit in Fort McMurray, Canada is provided.

Introduction

Geostatistics is used to generate maps of properties of interest. Often, kriging or SGS is implemented to populate these property models. It is desirable for these property maps to be as accurate as possible; however, kriging and SGS do not distinguish between locations that are being interpolated and locations that are being extrapolated. In this manuscript the convex hull of the data is used to distinguish between interpolation and extrapolation (Figure 1). Estimation or simulation locations inside the convex hull of the sample data are considered to be interpolated while locations outside the convex hull are considered to be extrapolated. It would be reasonable to consider some expansion of the convex hull by an arbitrary distance to consider the influence of the wells lying directly on the convex hull; a reasonable arbitrary distance would be a multiple of the average sample spacing, S, up to a reasonable distance, such as $1.5S$. In this paper, we consider no expansion of the convex hull and limit the discussion to the effect of extrapolating outside the convex hull defined only by the sample data.

Clearly, extrapolation is more difficult than interpolation as there are fewer nearby samples to inform the estimate. Fortunately, the areas of importance, i.e. the locations that are of the most importance to the geomodeler, is normally within the convex hull. In interpolation mode, kriging can generate reasonable results as it effectively weights nearby sample data and can consider important details such as anisotropy, data closeness, data redundancy and secondary data; this is why kriging has proven very effective in geostatistics. However, once outside the convex hull, fewer samples are available for estimation and the assumptions underlying kriging become more important. If implementing simple kriging (SK), the estimates tend to the global mean far from sample data. If implementing ordinary kriging (OK) the estimates tend to the mean defined by the sample data within the search radius. In either case, the tendency towards the mean may be inappropriate. Consider the case where the deposit is disseminated and it is known that outside the convex hull the deposit tends to a value of 0. In fact, this is often the reason for the lack of samples outside the convex hull; if an area is known to be barren, there is no reason to sample. Therefore, if the geology indicates that a deposit tends to 0 in extrapolation mode, this information should be incorporated in the modeling process.

Currently, such information would be incorporated with a trend model. Often generating a trend model for areas of the model that are deemed unimportant (extrapolation mode) is skipped and the models are only used at interpolation location. This is acceptable good practice as the models are not used in extrapolation where they are known to be incorrect; however, it does mean that the models depart from the available knowledge. This paper presents a simple correction that can be used to correct the models such that they tend towards a target mean away from the convex hull of the data.

Methodology

First, the convex hull of the data is required. Within this convex hull the kriging estimate and variance are unchanged. In this paper the implementation QHULL (Barber et al, 1996) is utilized to determine the convex hull of the data. Consider the available 383 wells from an oil sands deposit near Fort McMurray (Figure 1). The variable of interest is the net continuous bitumen (NCB) thickness well data available and its convex hull:

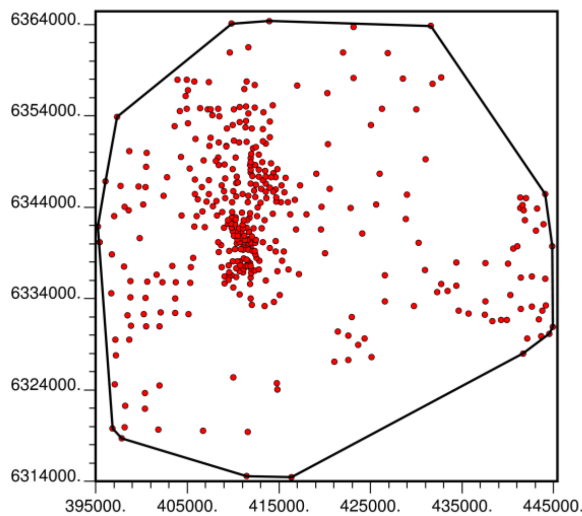

Figure 1: Convex hull of 383 well data.

A single data point far from all other available well data may have undue influence on the convex hull. Consider removing each data point and recalculating the area of the convex hull. If the area of the new convex hull changes by more than EPSILON, then the point is removed from the data set to generate the effective convex hull. If EPSILON = 2.0% then one well is removed from the well data (Figure 2 left) and the effective convex hull is recalculated with 382 wells as shown in Figure 2 right. The correction is proportional to the distance to the effective convex hull (Figure 2).

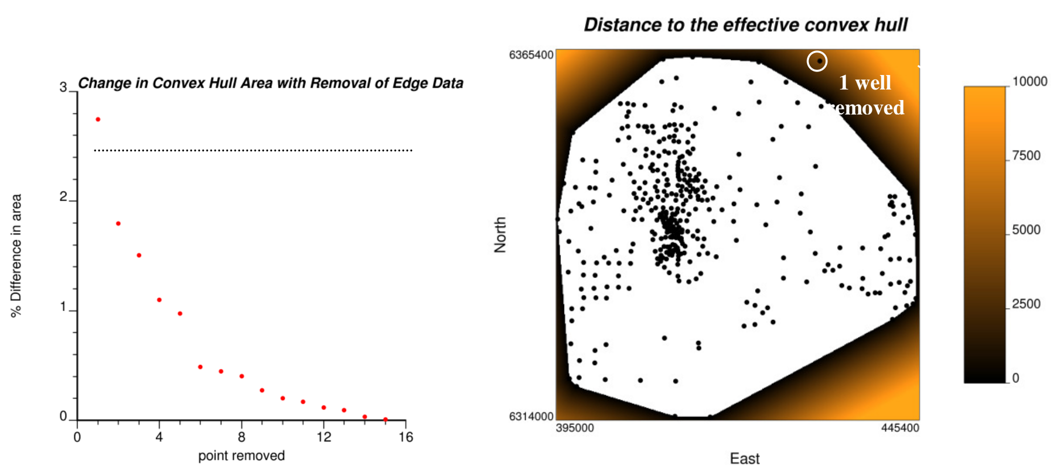

Figure 2: Left: wells that have a large influence on the area of the convex hull are removed to generate the effective convex hull. Right: effective convex hull of 383 well data with distance to the convex hull.

The proposed correction (Equation 2 below) applies to all grid locations outside the effective convex hull. If sample data are removed from the data base to calculate the effective convex hull, there are sampled grid cells that will fail to reproduce the conditioning data after correction of the model. The sample data

that fall outside the effective convex hull must be altered such that after correction the original data values are reproduced. Therefore, the following data adjustment is required:
Step 1) Determine the magnitude of the correction at the data location (Equation 2)
Step 2) Add the correction to the data (Equation 3)
Step 3) Krige as usual.
Step 4) Correct the model (Equation 2). Because the correction was included in the data used in kriging (Step 2), the original data are reproduced at all locations, including sample locations outside the effective convex hull (Figure 2).

In order for the correction to be smooth, there must be no alteration of the mean near the convex hull and an increasing correction with increasing distance, *d*, from the convex hull. A Gaussian shaped function is used to apply the correction to the mean (m). The two input parameters in this equation are the *extrapolation range* parameter, a, and a *target mean*, TM.

$$m(d) = \left(1 - e^{-\frac{9d^2}{a^2}}\right) \cdot TM \quad (1)$$

The global mean (Figure 3, right) outside the effective convex hull is reduced as per the correction determined in Figure 3 (left) with an extrapolation range a = 10000m and a target mean of -1.0. The global mean within the convex hull remains 0.0. The reduction in the mean necessitates a similar reduction in the variance. Without a reduction in variance, very low estimated values are possible. The variance is also smoothly reduced based on maintaining a constant probability to be below an input threshold. To maintain a constant probability to be below -1.5 (Figure 4 left) a reduction in the local variance is required. The proposed methodology is to select a lower threshold and calculate the distribution variance such that the probability below the threshold is unchanged (Figure 4 left). Setting the probability below -1.5 to be 0.067, as for a N(0,1) distribution, the locally varying variance can be calculated (Figure 4 right).

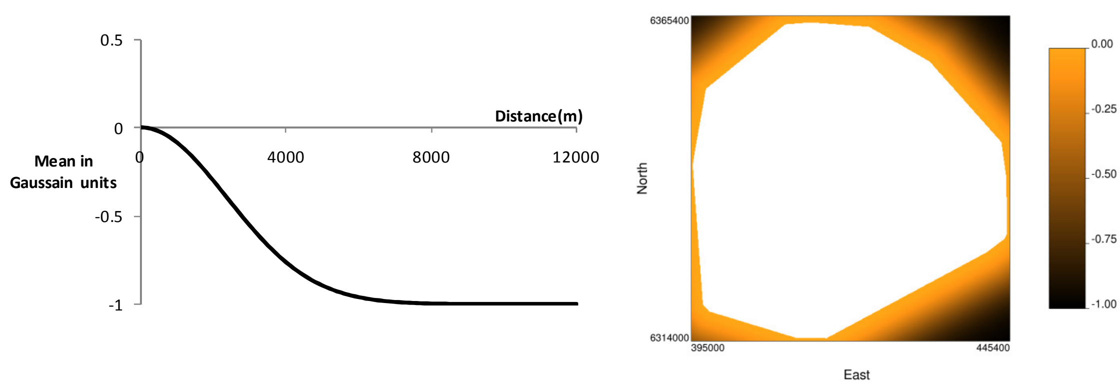

Figure 3: Left: Mean function with an extrapolation range a=10000m and a target mean = -1.0. Right: The global mean, inside the convex hull the mean is 0.0.

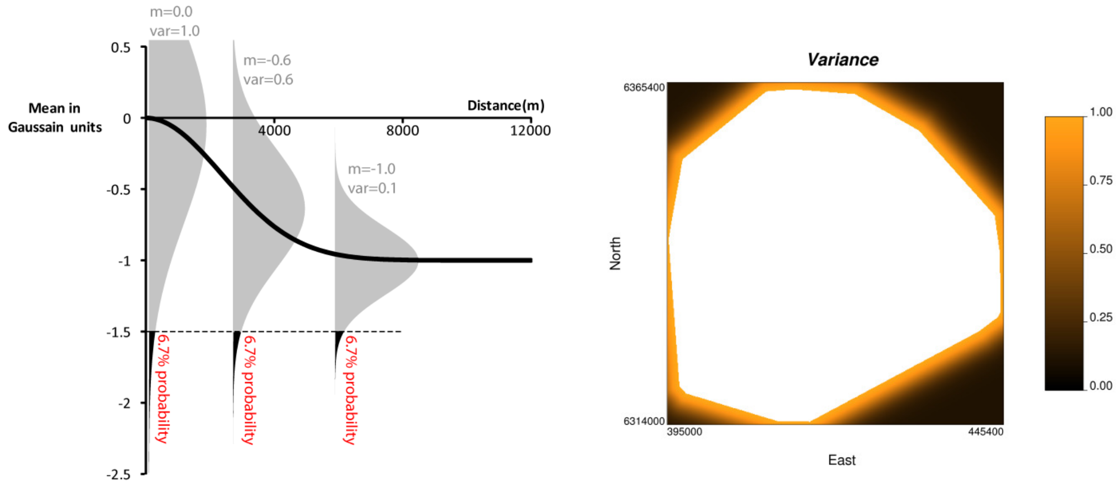

Figure 4: Left: The black line is identical to Figure 3 and shows the decreasing mean with increasing extrapolation distance from the convex hull. Depending on the mean, the global variance of the Gaussian distribution must be reduced to maintain a 0.067 probability to be below -1.5. Right: variance for the NCB data, inside the convex hull the variance is 1.0.

With the global mean (Equation 1) and global variance (Figure 4) known it is possible to correct a given model. In this example a model from Bayesian updating was generated for the available data and secondary data. It is known that at the extents of this area the net bitumen thickness tends to 0m; however, no samples are available to control the geostatistical models. In order to enforce this constraint the following correction (Equation 2) is made to every estimated value based on the new global mean (Figure 3) and global variance (Figure 4). Within the convex hull there is no change to the maps while the correction is smoothly applied to locations outside the convex hull, with increasing magnitude as the extrapolation distance, *d*, increases (Figure 6).

$$y_{corrected}(\mathbf{u}) = y(\mathbf{u})\sigma(d) + m(d) \tag{2}$$

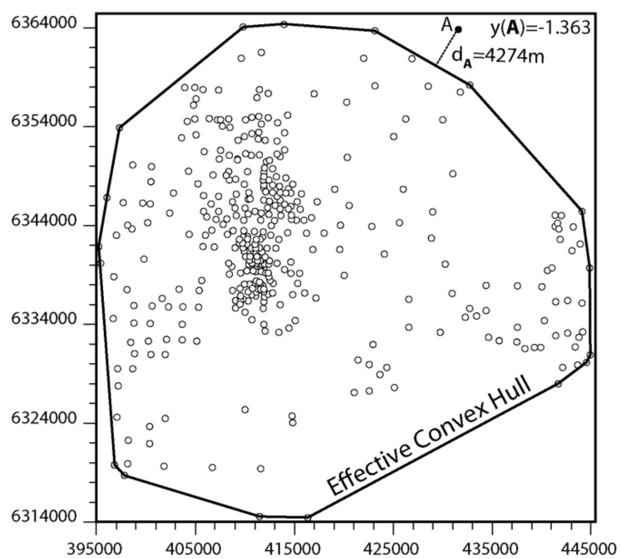

Figure 5: Sample location A is outside the effective convex hull.

Recall that data locations that are outside the effective convex hull must be altered before kriging. y(**A**) must be adjusted so that after applying the correction in Equation 2 the sample data values are reproduced in the final corrected model. If this adjustment is not considered, the correction would be applied as follows, resulting in the incorrect value at location **A**:

From Equation 1: $m(4274) = -0.807$
From Figure 4: $\sigma(4274) = 0.463$
From Equation 2: $y_{corrected}(\mathbf{A}) = -1.363 \cdot \sqrt{0.214} - 0.807 = -1.44$ (incorrect)

To ensure the reproduction of the conditioning data, all sample data outside the convex hull must be adjusted prior to kriging (Equation 3). Kriging is performed with the adjusted value and following the correction (Equation 2) conditioning data are honored.

$$y_{adjusted}(\mathbf{A}) = \frac{y(\mathbf{A}) - m(d)}{\sigma(d)} \tag{3}$$

$$y_{adjusted}(\mathbf{A}) = \frac{-1.363 + 0.807}{0.463} = -1.20$$

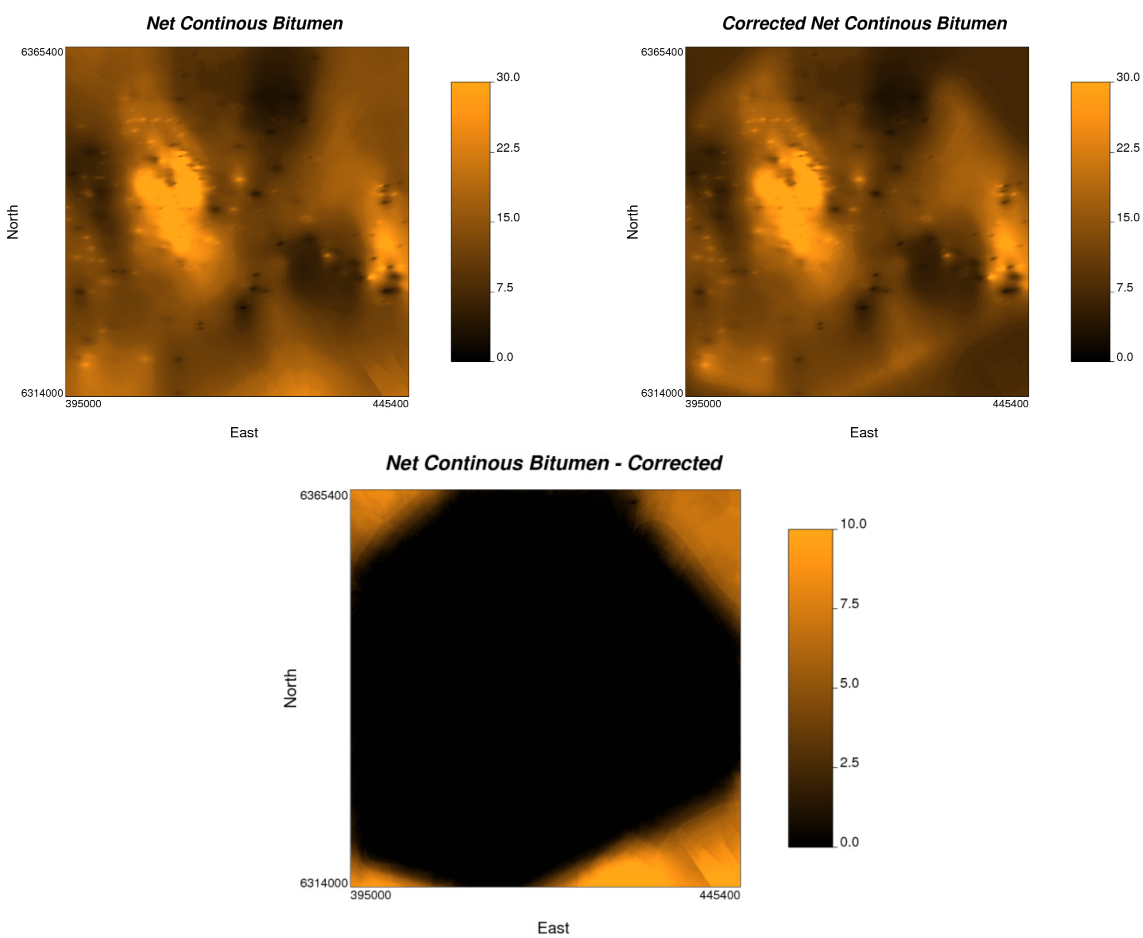

Figure 6: Above: Pre and post corrected net continuous bitumen. Below: The difference between the models.

Conclusions

The proposed correction is fast and simple. It only requires a single mathematical operation for every grid cell location (Equation 2). The correction can be used in place of a trend model when the deposit is known to gradually trend in extrapolation to a target mean other than the global. The implementation of the correction is made slightly more difficult with the use of the effective convex hull. In this situation, the practitioner must adjust their input data such that after correction and back transformation to original units, the data are honored outside the convex hull. If all sample data are used for determining the convex hull, this adjustment is not required. The example provided in this paper was for a kriged map. The same methodology and correction can be directly applied without change to a set of realizations.

References

Barber, C. B., D.P. Dobkin, and H.T. Huhdanpaa, 1996, "The Quickhull Algorithm for Convex Hulls," ACM Transactions on Mathematical Software, Vol. 22, No. 4, Dec. 1996, p. 469-483.