Updating Simulated Realizations with New Data

E.B. Niven and C. V. Deutsch

Simulated realizations are often updated with new data by resimulating the entire model. This results in a completely new set of realizations that have different features from the old ones, even at large distances from the new data. An easy and theoretically valid method for updating the existing realizations with new data is presented. The resulting realizations honor the new and old data as well as the general features of the old realizations. Furthermore, the realizations are beyond the range of correlation from the new data. A program is documented that automatically updates existing realizations with new data.

Introduction

For many mining and petroleum projects, drilling for samples occurs in defined seasons or, at least, incrementally. The data obtained from drilling is analyzed by geologists and geostatisticians, who then build numerical geological models such as simulated realizations. The numerical geological models are rebuilt with the old and new data. One disadvantage of this method is that the resulting updated simulated realizations will look very different, even at large distances away from the new data locations (i.e. beyond the variogram range) due to the implementation most simulation algorithms. This is a disadvantage if planning, design, or realization ranking has been performed. It would advantageous to update the realizations only within the range of the new data. Of course, there is no choice but to rebuild all realizations if the new data change the modeling parameters.

This research presents a simple method of updating old realizations based upon the new data collected, rather than completely rebuilding new realizations from scratch.

Proposed Methodology

Say we have *n* samples, y_α, where $\alpha = 1,...,n$, from an old drilling program and that those *n* samples are used to build a set of simulated realizations. We can also say that there are *m* new samples, y_ω, where $\omega = 1,...,m$, that we wish to use in order to update our simulated realizations (generated with sequential Gaussian simulation). At each of the new and old sample locations, we can calculate the difference between the simulated realizations and the new sample attribute value:

$$\Delta_i = y_i - y_{i,simulated} \qquad i = 1,...,n+m \tag{1}$$

Note that $\Delta_i = 0$ at each of the old sample locations since sequential Gaussian simulation honors the data. For each simulated realization, we can simple krige an estimate of Δ_i^* (using a mean of 0) at every location in our field of interest. Then we can add together the kriged estimates of Δ_i^* and the simulated realizations, y_s^*. The result is a set of updated simulated realizations that honor the new and old data while leaving the realizations unchanged at locations beyond the range of correlation from new data and preserving the general features of the old realizations. This method is similar to that presented by Barnes and Watson (1992). This method is theoretically valid as is shown in the following proof:

Proof

Figure 1 shows an arrangement of old samples (shown as red stars), a location for estimation (location 2, shown by the green square) and a new sample (location 1, shown by an orange circle). Say there are *n* old samples, y_α, where $\alpha = 1,...,n$. Now say we have one new sample at location 1, y_{new1}, (as is shown in Figure 1).

Then, the kriged estimate at location 1, using all *n* old data is (Isaaks and Srivastava, 1989):

$$y^*_{1,n} = \sum_{\alpha=1}^{n} \lambda_\alpha y_\alpha \tag{2}$$

And the kriged estimate at location 2 using all *n* old data is:

$$y^*_{2,n} = \sum_{\alpha=1}^{n} \gamma_\alpha y_\alpha \tag{3}$$

However, the kriged estimate at location 2 using all *n* old data plus 1 new data is:

$$y^*_{2,n+1} = \sum_{\alpha=1}^{n} \mu_\alpha y_\alpha + \varphi y_{new1} \tag{4}$$

Meanwhile, the estimate of Δ_i at location 2 using all *n* old data plus 1 new data is:

$$y^*_{2,n+1} = \sum_{\alpha=1}^{n} \eta_\alpha (y_\alpha - y^*_{2,n}) + \delta (y_{new1} - y^*_{1,n}) \tag{5}$$

In order to show that simulated results can be updated, it is sufficient to show that:

$$y^*_{2,n+1} = y^*_{2,n} + \Delta^*_{2,n+1} \tag{6}$$

The simple kriging equations for $y^*_{1,n}$ are:

$$\sum_{\beta=1}^{n} \lambda_\alpha C_{\alpha\beta} = C_{\alpha 1} \qquad \alpha = 1,...,n \tag{7}$$

The simple kriging equations for $y^*_{2,n}$ are:

$$\sum_{\beta=1}^{n} \gamma_\alpha C_{\alpha\beta} = C_{\alpha 2} \qquad \alpha = 1,...,n \tag{8}$$

The simple kriging equations for $y^*_{2,n+1}$ are:

$$\begin{aligned} &\sum_{\beta=1}^{n} \gamma_\alpha C_{\alpha\beta} + \varphi C_{\alpha 1} = C_{\alpha 2} \qquad \alpha = 1,...,n \\ &\sum_{\beta=1}^{n} \gamma_\alpha C_{1\beta} + \varphi C_{11} = C_{12} \qquad (\textit{the } n+1 \textit{ equation}) \end{aligned} \tag{9}$$

The simple kriging equations for $\Delta^*_{2,n+1}$ are:

$$\sum_{\beta=1}^{n}\eta_\alpha C_{\alpha\beta} + \delta C_{\alpha 1} = C_{\alpha 2} \qquad \alpha = 1,\ldots,n$$

$$\sum_{\beta=1}^{n}\eta_\alpha C_{1\beta} + \delta C_{11} = C_{12} \qquad (\textit{the } n+1 \textit{ equation}) \tag{10}$$

Note that:

$$\eta_\alpha = \mu_\alpha \tag{11}$$

$$\delta = \varphi \tag{12}$$

Now, if we substitute equation (7) and (8) into equation (10), we have:

$$\sum_{\beta=1}^{n}\eta_\alpha C_{\alpha\beta} + \delta\sum_{\beta=1}^{n}\lambda_\alpha C_{\alpha\beta} = \sum_{\beta=1}^{n}\gamma_\alpha C_{\alpha\beta} \qquad \alpha = 1,\ldots,n \tag{13}$$

Simplifying:

$$\sum_{\beta=1}^{n}[\eta_\alpha + \delta\lambda_\alpha]C_{\alpha\beta} = \sum_{\beta=1}^{n}\gamma_\alpha C_{\alpha\beta} \qquad \alpha = 1,\ldots,n \tag{14}$$

Simplifying we have:

$$\eta_\alpha + \delta\lambda_\alpha = \gamma_\alpha \qquad \alpha = 1,\ldots,n \tag{15}$$

Since this is *a* solution, it must be *the* solution since kriging is unique. Now, we can substitute equations (3) and (5) into (6) to get:

$$y^*_{2,n+1} = \sum_{\alpha=1}^{n}\gamma_\alpha y_\alpha + \sum_{\alpha=1}^{n}\eta_\alpha(y_\alpha - y^*_{2,n}) + \delta(y_{new1} - y^*_{1,n}) \qquad \alpha = 1,\ldots,n \tag{16}$$

Note that $(y_\alpha - y^*_{2,n}) = 0$ for all α since kriging (and sequential Gaussian simulation) honor the data points. Thus we have:

$$y^*_{2,n+1} = \sum_{\alpha=1}^{n}\gamma_\alpha y_\alpha + \delta(y_{new1} - y^*_{1,n}) \qquad \alpha = 1,\ldots,n \tag{17}$$

If we substitute equation (15) into equation (17) we get:

$$y^*_{2,n+1} = \sum_{\alpha=1}^{n}(\eta_\alpha + \delta\lambda_\alpha)y_\alpha + \delta(y_{new1} - y^*_{1,n}) \qquad \alpha = 1,\ldots,n \tag{18}$$

And if we substitute equation (2) into equation 18 and simplify, we get:

$$y^*_{2,n+1} = \sum_{\alpha=1}^{n}\eta_\alpha y_\alpha + \delta\sum_{\alpha=1}^{n}\lambda_\alpha y_\alpha + \delta(y_{new1} - \sum_{\alpha=1}^{n}\lambda_\alpha y_\alpha) \qquad \alpha = 1,\ldots,n \tag{19}$$

Which equals:

$$y^*_{2,n+1} = \sum_{\alpha=1}^{n} \eta_\alpha y_\alpha + \delta \sum_{\alpha=1}^{n} \lambda_\alpha y_\alpha + \delta y_{new1} - \delta \sum_{\alpha=1}^{n} \lambda_\alpha y_\alpha \qquad \alpha = 1,...,n \tag{20}$$

Which simplifies to:

$$y^*_{2,n+1} = \sum_{\alpha=1}^{n} \eta_\alpha y_\alpha + \delta_\alpha y_{new1} \qquad \alpha = 1,...,n \tag{21}$$

Note that since $\eta_\alpha = \mu_\alpha$ and $\delta = \varphi$, we are left with:

$$y^*_{2,n+1} = \sum_{\alpha=1}^{n} \mu_\alpha y_\alpha + \varphi y_{new1} \qquad \alpha = 1,...,n \tag{22}$$

Which is correct and the same as our original definition for $y^*_{2,n+1}$ in equation (4). Of course, the result is the same if *m* new data are used instead of just one new data. Therefore, it is proved that we can add together the kriged estimates of Δ^*_i, based upon *m* new data and the kriged estimate of the variable based upon *n* old data at each location in the domain of interest.

Implementation

A GSLIB-style program called UPDATE_SIM was created that calculates Δ_i at each data location. Then, kriged estimates of Δ^*_i are calculated at each location. Finally, the simulated realizations based on the old data and the estimate of Δ^*_i are added together to arrive at the updated realizations.

Example

Figure 2 shows six old samples (shown as blue circles) and 1 new sample (shown as a red square) of some attribute along a line. The distance between the samples is noted. Figure 3 shows three sequential Gaussian simulated realizations of the attribute based upon only the old samples. The values are shown simulated at a 1 distance unit spacing. A spherical variogram with no nugget effect and a range of 20 was assumed. Naturally, the realizations converge at the old data locations since sequential Gaussian simulation honors the samples. At each of the old and new sample locations Δ_i can be calculated for each of the three simulated realizations ($\Delta_i = 0$ at each of the old data locations and $\Delta_i \neq 0$ at the new data location). Figure 4 shows the kriged estimates of Δ^*_i at each location and for all three realizations. Figure 5 shows the simulated realizations after they have been updated to incorporate the new data point. Note that the realizations are unchanged at distances of greater than 20 units from the new data point.

Discussion

Usually when new data is obtained, the geostatistician must rebuild the entire numerical geological model. This will result in new simulated realizations that look quite different from the old ones, even at great distances from the new data due to the implementation of sequential Gaussian simulation. Thus, it is somewhat difficult to examine the impact of new data on simulated realizations. This short note presents an easy and theoretically valid method for updating simulated realizations with new data. The updated realizations honor both the new and old data as well as the features of the old realizations. Furthermore, the realizations are unchanged at large distances away from the new data (i.e. beyond the variogram range). A computer program was developed which will automatically update a set of simulated realizations of a continuous variable.

It should be noted that this method assumes that the new data is from the same population as the old data and that they can be statistically grouped together under a decision of stationarity. Furthermore, this method assumes that the variogram is unchanged by the new data.

References

Barnes, R.J. and Watson, A.G., 1992, Efficient Updating of Kriging Estimates and Variances, Mathematical Geology, Vol. 24, No. 1, pages 129 – 133.

Isaaks, E.H. and Srivastava, R.M., 1989, An Introduction to Applied Geostatistics, Oxford University Press, New York, 561 pages.

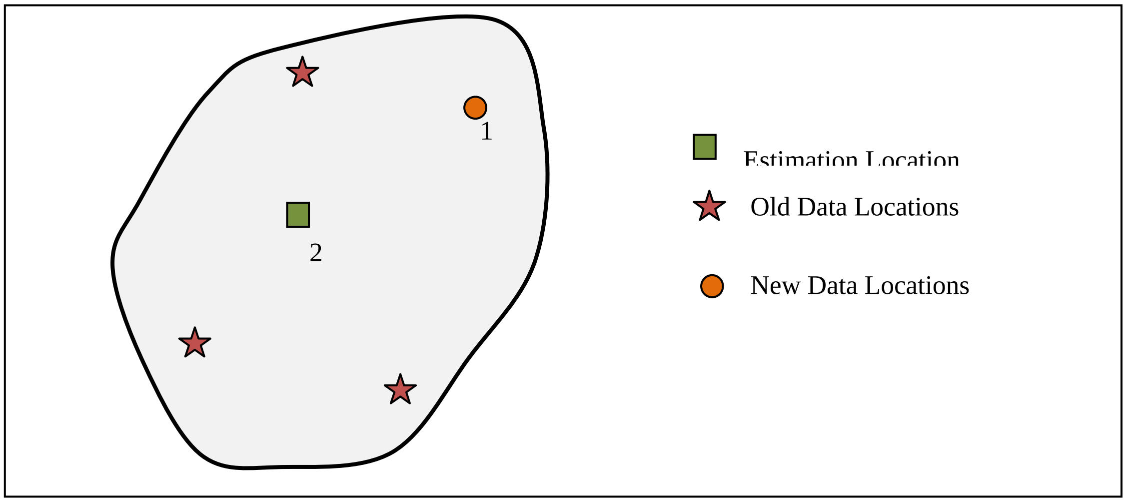

Figure 1: Three old samples, one new sample and the location for estimation within a domain of interest

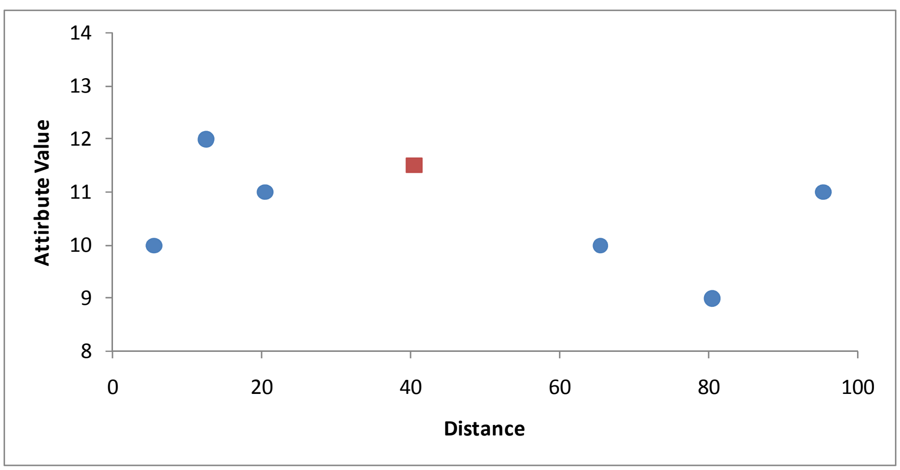

Figure 2: Six old samples (shown as blue circles) and one new sample (shown as a red square) along a line

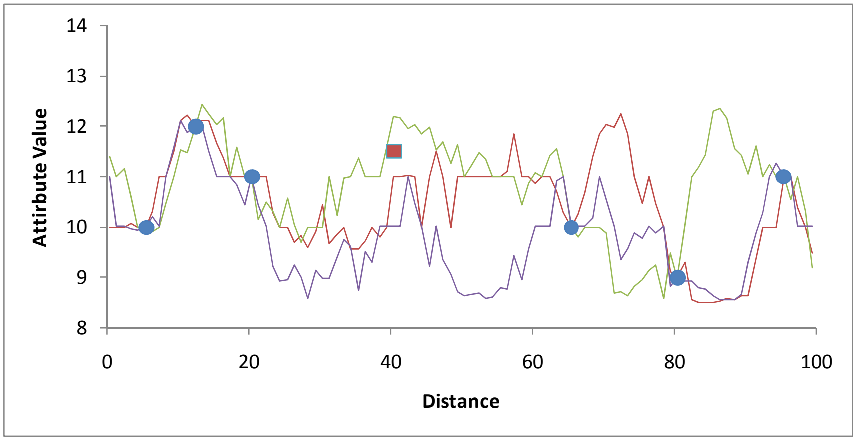

Figure 3: Three simulated realizations, based upon the old data points (the new data point was excluded from the sequential Gaussian simulation).

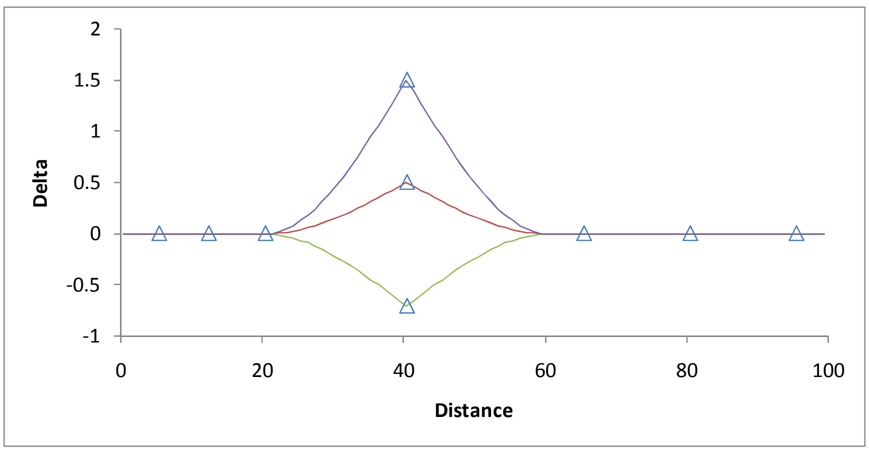

Figure 4: Three kriged estimates of the difference between new data and the simulated realization (Δ_i^*).

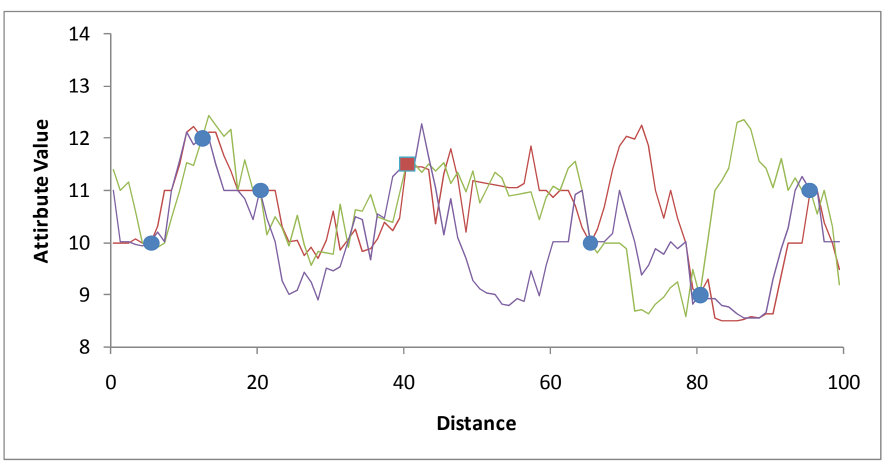

Figure 5: The updated simulated realizations, created by adding the kriged Δ_i^* and the old simulated realizations.