Updating Simulated Realizations with New Data

Eric B. Niven, David Garner and Clayton V. Deutsch

Currently, the only way to update simulated realizations with new data is to re-simulate the entire model. This results in a completely new set of realizations that may look very different from the old ones, even at large distances from the new data. This paper presents an easy and theoretically valid method to update the existing realizations with new data. The resulting realizations honor the new and old data as well as the general 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 GSLIB-style computer program is presented that automatically updates existing realizations with new data.

Introduction

For many mining and petroleum projects, drilling for samples occurs in defined seasons. For instance, many mining projects in the pre-feasibility stages may conduct drilling campaigns in the summer when the weather is good. The data obtained from drilling is analyzed by geologists and geostatisticians, who then build numerical geological models, which may result in a set of simulated realizations. After successive drilling programs, there is new data that needs to be incorporated into the geological models. Usually, this requires completely rebuilding the numerical geological models with the old data and the new data combined. 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 of sequential Gaussian simulation.

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 $\omega = n+1,...,m$, 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:

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

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 Proof 1.

Proof 1 – Show that Updating Simulated Realizations is Valid

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:

$$\Delta^*_{2,n+1} = \sum_{\alpha=1}^{n} \eta_\alpha \Delta_\alpha + \delta \Delta_{new1} \tag{5}$$

Remember that $\Delta_\alpha = y_\alpha - y^*_\alpha$. Now, $\Delta_\alpha = 0$ for all α since previously kriged or simulated surface honors the old data. So, we have:

$$\Delta^*_{2,n+1} = \delta(y_{new1} - y^*_{1,n}) \tag{6}$$

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

$$y^*_{2,n+1} = y^*_{2,n} + \Delta^*_{2,n+1} \tag{7}$$

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

$$\sum_{\beta=1}^{n} \lambda_\alpha C_{\alpha\beta} = C_{\alpha 1} \qquad \alpha = 1,\ldots,n \tag{8}$$

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

$$\sum_{\beta=1}^{n} \gamma_\alpha C_{\alpha\beta} = C_{\alpha 2} \qquad \alpha = 1,\ldots,n \tag{9}$$

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

$$\begin{aligned} &\sum_{\beta=1}^{n} \gamma_\alpha C_{\alpha\beta} + \varphi C_{\alpha 1} = C_{\alpha 2} \qquad \alpha = 1,\ldots,n \\ &\sum_{\beta=1}^{n} \gamma_\alpha C_{1\beta} + \varphi C_{11} = C_{12} \qquad (the\ n+1\ equation) \end{aligned} \tag{10}$$

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

$$\begin{aligned} &\sum_{\beta=1}^{n} \eta_\alpha C_{\alpha\beta} + \delta C_{\alpha 1} = C_{\alpha 2} \qquad \alpha = 1,\ldots,n \\ &\sum_{\beta=1}^{n} \eta_\alpha C_{1\beta} + \delta C_{11} = C_{12} \qquad (the\ n+1\ equation) \end{aligned} \tag{11}$$

Note that:

$$\eta_\alpha = \mu_\alpha \tag{12}$$

$$\delta = \varphi \tag{13}$$

Now, if we substitute equation (8) and (9) into equation (11), we have:

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

Simplifying:

$$\sum_{\beta=1}^{n} [\eta_\alpha + \delta\lambda_\alpha] C_{\alpha\beta} = \sum_{\beta=1}^{n} \gamma_\alpha C_{\alpha\beta} \qquad \alpha = 1,\ldots,n \tag{15}$$

Simplifying we have:

$$\eta_\alpha + \delta\lambda_\alpha = \gamma_\alpha \qquad \alpha = 1,...,n \tag{16}$$

Since this is *a* solution, it must be *the* solution since kriging is unique. Now, we can substitute equations (3) and (6) into (7) to get:

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

If we substitute equation (16) into equation (17) we get:

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

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

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

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

$$= \sum_{\alpha=1}^{n} \eta_\alpha y_\alpha + \delta y_{new1} \qquad \alpha = 1,...,n \tag{21}$$

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. Next, we must show that the conditional mean and variance are unaffected by updating and that the condicional covariance between unsampled locations is preserved.

Proof 2 – Check Mean Value Reproduction

We start with the definition of the updated simulated value at location 2:

$$\begin{aligned} y^{new}_{sim}(u) &= y^{old}_{sim}(u) + \Delta_{sk} \\ &= y_{sk}(u) + y^{old}_{sim}(u) - y_{sk,old}(u) \end{aligned} \tag{23}$$

Where $y^{new}_{sim}(u)$ is a simulated value from the new (updated) realization at a location *u*, $y^{old}_{sim}(u)$ is the simulated value from the old realization at location *u*, $\Delta_{sk} = y_{sk}(u) - y_{sk,old}(u)$ is the kriged difference between the data values and the old simulated realization at the data value locations (i.e. $\Delta_{sk} = \sum_{\alpha=1}^{n} \lambda_\alpha [y(u_\alpha) - y_{old}(u_\alpha)]$).

$$\begin{aligned} E\left\{y^{sim}_{new}(u)\right\} &= E\left\{y_{sk}(u)\right\} + E\left\{y^{sim}_{old}(u)\right\} + E\left\{y_{sk,old}(u)\right\} \\ &= E\left\{y_{sk}(u)\right\} + 0 + \sum_{\alpha=1}^{k} \lambda_\alpha(u) E\left\{y_{old}(u_\alpha)\right\} \\ &= E\left\{y_{sk}(u)\right\} + 0 + 0 \\ &= E\left\{y_{sk}(u)\right\} \end{aligned} \tag{24}$$

So, the mean is unchanged. Now, we must check the variance reproduction.

Proof 3 – Check Variance Reproduction
Since $y_{sk}(u)$ is a constant:

$$Var\left\{y_{new}^{sim}(u)\right\} = Var\left\{y_{sk}(u) + y_{old}^{sim}(u) + y_{sk,old}(u)\right\} = Var\left\{y_{old}^{sim}(u) + y_{sk,old}(u)\right\} \tag{25}$$

Recalling that $Var\{X\} = E\{X^2\} - m_x^2$, where m_x is the mean of *X*, then:

$$Var\left\{y_{new}^{sim}(u)\right\} = E\left\{\left[y_{old}^{sim}(u) + y_{sk,old}(u)\right]^2\right\} - 0^2 \tag{26}$$

$$= E\left\{y_{old}^{sim}(u)^2 - 2y_{old}^{sim}(u)y_{sk,old}(u) + y_{sk,old}(u)^2\right\} \tag{27}$$

$$= E\left\{y_{old}^{sim}(u)^2\right\} - E\left\{2y_{old}^{sim}(u)y_{sk,old}(u)\right\} + E\left\{y_{sk,old}(u)^2\right\} \tag{28}$$

Since $E\left\{y_{old}^{sim}(u)^2\right\} = 1$ and $E\left\{y_{sk,old}(u)^2\right\}$ is the variance of the kriged estimate (i.e. $E\left\{y_{sk,old}(u)^2\right\} = 1 - \sigma_{sk}^2(u)$), then we have:

$$Var\left\{y_{new}^{sim}(u)\right\} = 1 - 2\sum \lambda_\alpha E\left\{y_{old}^{sim}(u)y_{sk,old}(u)\right\} + 1 - \sigma_{sk}^2(u) \tag{29}$$

Recalling that $\sigma_{sk}^2(u) = 1 - \sum_{\alpha=1}^{n} \lambda_\alpha C(u - u_\alpha)$, the variance of the updated realization is:

$$Var\left\{y_{new}^{sim}(u)\right\} = 1 - 2\left(1 - \sigma_{sk}^2(u)\right) + 1 - \sigma_{sk}^2(u) \tag{30}$$

$$Var\left\{y_{new}^{sim}(u)\right\} = \sigma_{sk}^2(u) \tag{31}$$

So, the conditional variance is correct! Now, we must check the covariance between two conditionally simulated points. Consider two updated points *u* and *u'*. Then the updated simulated values at those locations are:

$$\begin{aligned} y_{sim}^{new}(u) &= y_{sk}(u) + y_{sim}^{old}(u) - y_{sk,old}(u) \\ y_{sim}^{new}(u') &= y_{sk}(u') + y_{sim}^{old}(u') - y_{sk,old}(u') \end{aligned} \tag{32}$$

Recalling that $Cov(x, y) = E\{[x - m_x][y - m_y]\}$, the covariance between the values at *u* and *u'* is:

$$\begin{aligned} Cov\left\{y_{sim}^{new}(u), y_{sim}^{new}(u')\right\} = E\Big\{&\left[\left(y_{sk}(u) + y_{sim}^{old}(u) - y_{sk,old}(u)\right) - y_{sk}(u)\right] \\ &\left[\left(y_{sk}(u') + y_{sim}^{old}(u') - y_{sk,old}(u')\right) - y_{sk}(u')\right]\Big\} \end{aligned} \tag{33}$$

$$Cov\left\{y_{sim}^{new}(u), y_{sim}^{new}(u')\right\} = E\left\{\left[y_{sim}^{old}(u) - y_{sk,old}(u)\right]\left[y_{sim}^{old}(u') - y_{sk,old}(u')\right]\right\} \tag{34}$$

$$= E\left\{y_{sim}^{old}(u)y_{sim}^{old}(u') - y_{sk,old}(u)\,y_{sim}^{old}(u') - y_{sim}^{old}(u)y_{sk,old}(u') + y_{sk,old}(u)y_{sk,old}(u')\right\} \tag{35}$$

$$\begin{aligned} &= C(h) - \sum_{\alpha=1}^{k} \lambda_\alpha(u) E\left\{y_{sim}^{old}(u_\alpha)y_{sim}^{old}(u')\right\} - \sum_{\alpha=1}^{k} \lambda_\alpha(u) E\left\{y_{sim}^{old}(u_\alpha)y_{sim}^{old}(u)\right\} \\ &+ \sum_{\alpha=1}^{k}\sum_{\beta=1}^{k} \lambda_\alpha(u)\lambda_\beta(u) E\left\{y_{sim}^{old}(u_\alpha)y_{sim}^{old}(u_\beta)\right\} \end{aligned} \tag{36}$$

$$\begin{aligned} &= C(h) - \sum_{\alpha=1}^{k} \lambda_\alpha(u) C(u' - u_\alpha) - \sum_{\alpha=1}^{k} \lambda_\alpha(u) C(u - u_\alpha) \\ &+ \sum_{\alpha=1}^{k}\sum_{\beta=1}^{k} \lambda_\alpha(u)\lambda_\beta(u') C(u_\alpha - u_\beta) \end{aligned} \tag{37}$$

Now, recall the kriging system of equations:

$$\sum_{\beta=1}^{n}\lambda_\beta(u')C(u_\alpha-u_\beta)=C(u'-u_\alpha) \qquad \alpha=1,\ldots,k$$
$$\sum_{\beta=1}^{n}\lambda_\beta(u)C(u_\alpha-u_\beta)=C(u-u_\alpha) \qquad \alpha=1,\ldots,k \tag{38}$$

So, (36) becomes:

$$=C(h)-\sum_{\alpha=1}^{k}\lambda_\alpha(u)\sum_{\beta=1}^{k}\lambda_\beta(u')C(u_\alpha-u_\beta)-\sum_{\alpha=1}^{k}\lambda_\alpha(u')\sum_{\beta=1}^{k}\lambda_\beta(u)C(u_\alpha-u_\beta)$$
$$+\sum_{\alpha=1}^{k}\sum_{\beta=1}^{k}\lambda_\alpha(u)\lambda_\beta(u')C(u_\alpha-u_\beta) \tag{39}$$

Therefore:

$$Cov\left\{y_{sim}^{new}(u),\, y_{sim}^{new}(u')\right\}=C(h)-\sum_{\alpha=1}^{k}\lambda_\alpha(u')\sum_{\beta=1}^{k}\lambda_\beta(u)C(u_\alpha-u_\beta) \tag{40}$$

Thus, the conditional covariance between any two updated simulated values is correct. So we have seen that this method of updating simulated realizations gives the right expected value, the right variance and the right conditional covariance for any two locations *u* and *u'*.

Implementation

A GSLIB-style program called UPDATE_SIM was created which calculates Δ_i at each data location. Then, a kriged estimate of Δ_i^* is calculated at each location in the domain of interest. Finally, the simulated realizations based upon the old data and the estimate of Δ_i^* are added together to arrive at an updated set of realizations.

In order to illustrate the algorithm, a simple 1D example is considered first. 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_{sk} at each location and for all three realizations. The kriged estimate is constructed by simple kriging using the same variogram as the original simulation and a mean of 0.0. Figure 5 shows the simulated realizations after they have been updated to incorporate the new data point. The updated realizations are created by adding the old realization to the kriged estimate Δ_{sk}. Note that the realizations are unchanged at distances of greater than 20 units from the new data point since that is the range of the variogram.

Next, a more complicated 2D example is considered. Figure 6 shows to location maps of samples from the Jura data set. The left location map shows 259 old data that are assumed to have been collected in the past. The right location map shows 100 new data that are assumed to have been collected recently. A set of simulated realizations have already been constructed using the 259 old data and it would be useful to update those realizations rather than reconstructing them from scratch. Figure 7 shows the omnidirectional variograms for the new and old data. Both variograms were modeled using a single-structure spherical variogram with a nugget effect of 0.15 and a range of 1.1 distance units. Although the fit for the new data variogram is not as good as that for the old data, the fit still appears reasonable given that its sample size is much smaller.

On the top row of Figure 8, the realizations created using only the 259 old data are displayed. The 100 new data are then used to update the old realizations using the program UPDATE_SIM. The program calculates the difference between all 359 (new and old) data locations and the old simulated realizations. The differences are zero at the old data locations and not zero at all the new data locations (note that this is done for every realization). The differences are kriged and the kriged realizations are added to the old realizations to create updated realizations. The updated realizations for the Jura data are shown in the middle row of Figure 8. The updated realizations honor all the old and new data and are unchanged at locations that are outside the range of correlation from new data.

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.

Conclusions and Future Work

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 research 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.

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.

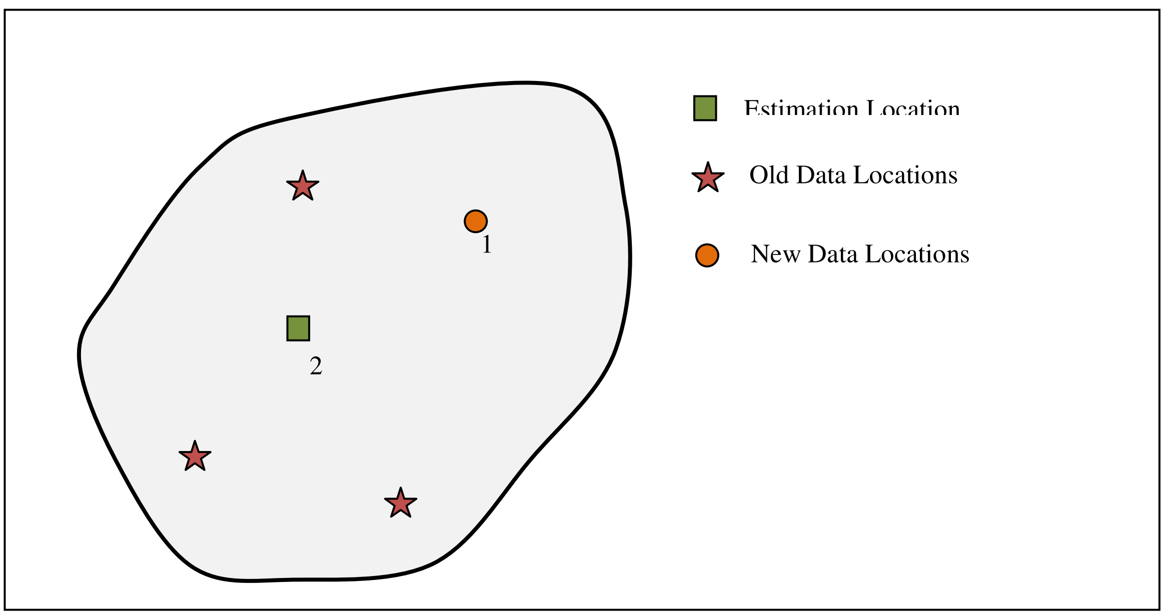

Figure 1: Three old samples, one new sample and the location for estimation within a domain of interest

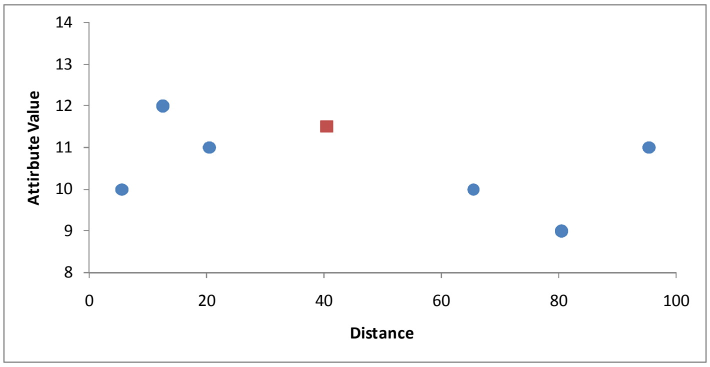

Figure 2: Six old samples (shown as blue circles) and one new sample (shown as a red square) along a line

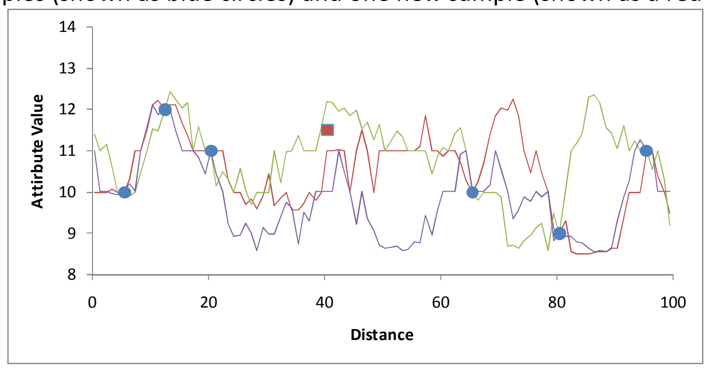

Figure 3: Three simulated realizations, based upon the old data points (the new data point was excluded from the sequential Gaussian simulation).

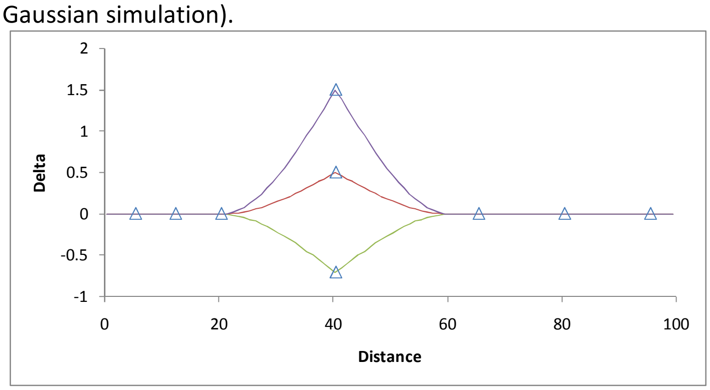

Figure 4: Three kriged estimates of the difference between new data point and the simulated realization ($\Delta_{sk} = \sum_{\alpha=1}^{n} \lambda_\alpha [y(u_\alpha) - y_{old}(u_\alpha)]$).

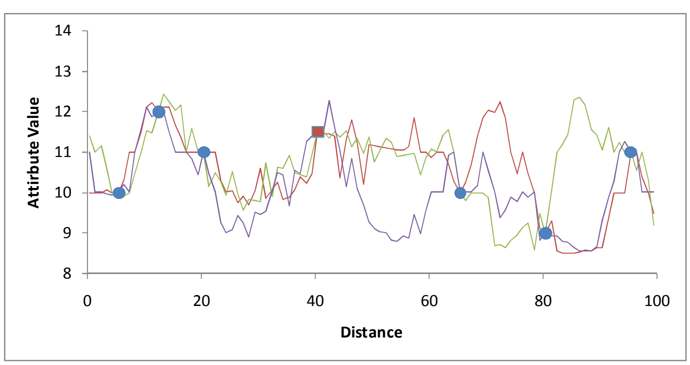

Figure 5: The updated simulated realizations, created by adding the kriged Δ_{sk} and the old simulated realizations.

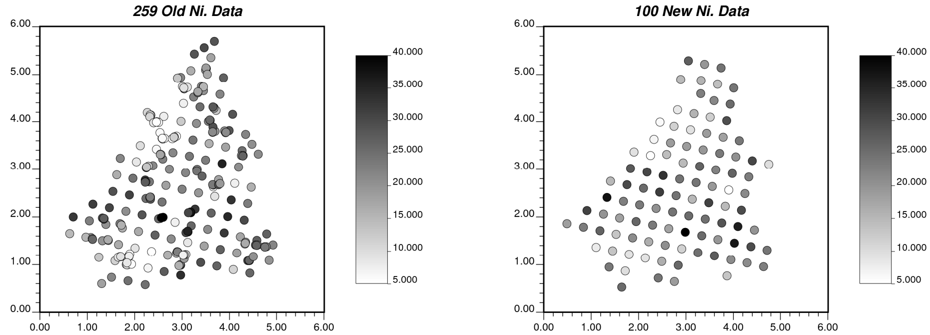

Figure 6: The Jura dataset. 259 old nickel concentration samples (ppm) and 100 new nickel samples.

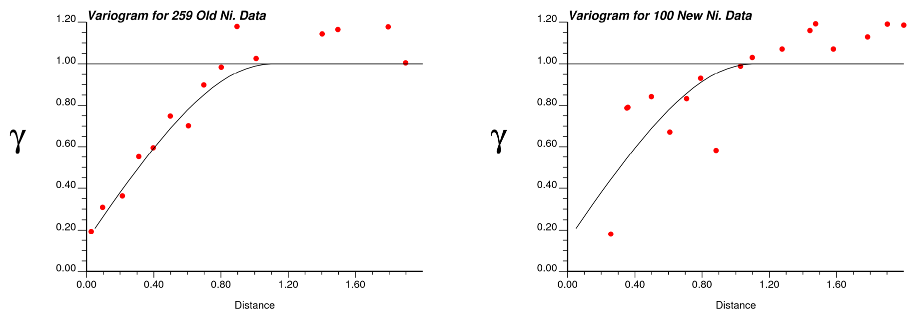

Figure 7: Omnidirectional variograms for the 259 old data on the left and the 100 new data on the right.

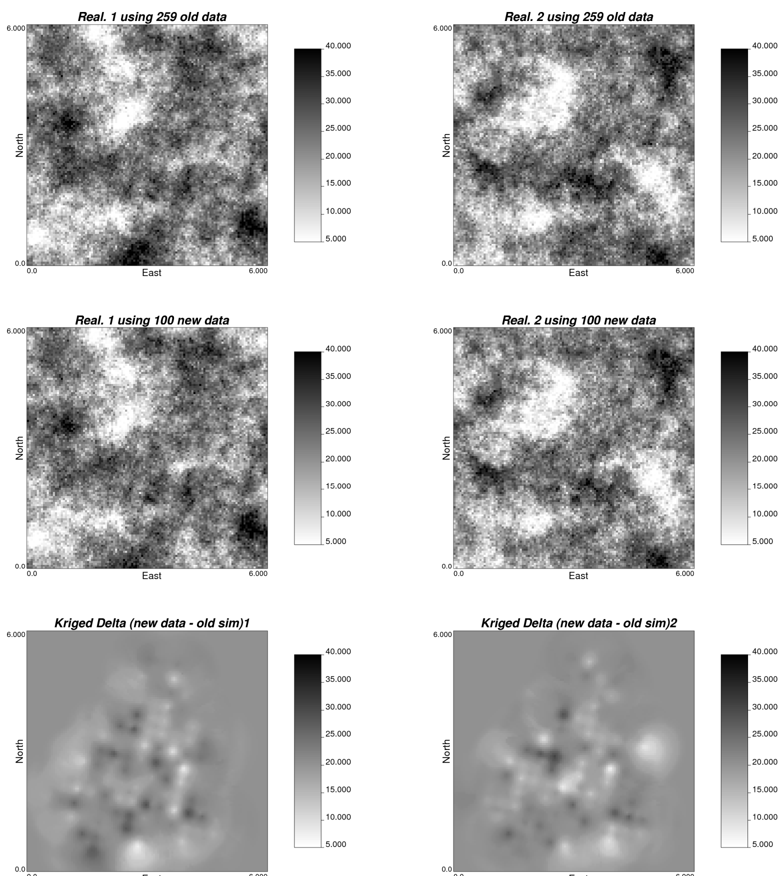

Figure 8: Top row: two simulated realizations of nickel content (ppm) created using 259 old data. Middle row: two updated realizations that incorporate the 100 new data in addition to the 259 old data. Bottom row: The kriged difference between the new data values and the old realizations