

# A Short Note on the Comparison of Kriging and the Average of Simulated Realizations

Brandon Wilde and Clayton V. Deutsch

Centre for Computational Geostatistics  
Department of Civil & Environmental Engineering  
University of Alberta

*The result of averaging many simulated realizations and kriging are theoretically the same; however, there are many practical reasons why they differ. Simulation in Gaussian units, departures from stationarity, inconsistencies between the normal scores and the original variable variogram, the type of kriging and different search parameters contribute to the difference. This short note develops some of those differences. This is important in the context of checking simulated realizations and comparing the results to kriging.*

## Background

There are two common ways to derive Gaussian simulation techniques: (1) as Monte Carlo drawing from a multivariate Gaussian distribution, and (2) as a scheme to correct the smoothness of kriging. Theoreticians prefer the first approach, practitioners prefer the second. Both lead to the same results. It is correct to interpret the simulated value at a location ( $y_{cs}(\mathbf{u})$ ) as the addition of the kriging estimate ( $y_{sk}(\mathbf{u})$ ) and a stochastic component ( $r(\mathbf{u})$ ):

$$y_{cs}(\mathbf{u}) = y_{sk}(\mathbf{u}) + r(\mathbf{u}) \quad (1)$$

Where  $y_{sk}(\mathbf{u})$  is calculated with the well known simple kriging equations and  $r(\mathbf{u})$  is drawn by Monte Carlo simulation from a distribution with a mean of zero and a variance equal to the kriging variance arising from the estimate of  $y_{sk}(\mathbf{u})$ . This interpretation holds regardless of what Gaussian simulation technique is used including sequential, matrix, spectral or turning bands.

Although simulation could be interpreted as little more than kriging, we know that simulation is more susceptible to biases. Kriged estimates are relatively insensitive to slight deviations in the variogram whereas simulated values are much more sensitive. This increased sensitivity comes from the kriging variance that is used for drawing the  $r$  values. Simulation places a greater reliance on stationarity since the global distribution intervenes at every location; once again from the  $r$  values. Moreover, simulation is much more sensitive to departures from multivariate Gaussianity. It is common for biases and other errors to appear in simulation, but not in kriging. For this reason, we advocate careful checking of simulated realizations (see previous CCG reports and papers such as the *Minimum Acceptance Criteria*). Once check is to compare the average of simulated realizations with estimates from kriging.

The comparison between the average of simulated realizations and kriging should be performed under the following conditions: (1) Gaussian units – the simulated values should be averaged in Gaussian units and the kriging should be performed with the Gaussian transformed (normal scores) data, (2) simple kriging – there is no theoretical basis for using ordinary kriging in most simulation algorithms, (3) a large search – up to the variogram range, and (4) with a very large

number of simulated realizations. Checking the average of simulated realizations to kriging should be performed even if it means multiple kriging runs and multiple simulation runs.

Comparing results in artificial Gaussian/normal units is not completely satisfactory. Practitioners know that the non-linear Gaussian transformation could mask biases and problems. Another very good check on simulated realizations is to compare (1) the average of many simulated realizations in original units with (2) the expected value from MG kriging. The short note on `PostMG` (see last year's CCG report) provides the details on how MG works – based on G. Verly's Ph.D. thesis. The approach of MG is to perform kriging in Gaussian units and then back transform quantiles of the conditional Gaussian distribution. The back transformed quantiles should give the same average as multiple realizations (in original units). This check is not completely satisfying since the same assumptions go into simulation as MG kriging; the check is a consistency/software check rather than a comparison of different methods.

Going one step further, we could check the average of simulated realizations against a conventional kriging run. This is recommended, but there are many reasons why the results could differ and yet both kriging and simulation are implemented in a reasonable way:

1. Inconsistency between the normal scores variogram and the original variable variogram. The previous paper in this report (308) showed how the normal scores and original variogram should be consistent with one another. The idea was to use the robust normal scores variogram to derive an original variable variogram. In practice, however, there can be inconsistencies between the modeled variograms that lead the average of simulated realizations to depart from kriging estimates.
2. Simple kriging (SK) is required in simulation to ensure histogram and variogram reproduction. Ordinary kriging (OK) is often applied in estimation for robustness, that is, to adapt to local departures from stationarity. The differences between SK and OK can be large – especially locally.
3. A large search with, perhaps, a multiple grid search strategy is often applied in simulation to ensure that the variogram is reproduced over all distances. The variogram will only be reproduced for distances and data that are included in the kriging. The search for data in simulation is large. Kriged models are often constructed with a restricted search strategy to avoid undue reliance on the decision of stationarity and to avoid excessive smoothing of the simulated realizations. The difference in the search between simulation and kriging can lead to large differences in the results.
4. The average of simulation will only compare to estimation if a very large number of realizations are drawn. The departures are surprisingly large even with 100 realizations; we must manage our expectations when we use a practical number of realizations (between 20 and 100).
5. Simulation depends more on properties of the random function model such as multivariate Gaussianity, stationarity and ergodicity. These are not decisions that a practitioner makes explicitly; however, they can have a large affect on the results. There are tests for Gaussianity. Trend analysis can indicate whether or not the decision of stationarity is reasonable. The range of the variogram relative to the size of the domain indicates the reasonableness of the implicit assumption of ergodicity.

The examples documented below illustrate some of the differences we could expect. We did not attempt to quantify the deviations we would accept in practice. This short note just shows some

of the differences. An area of future work is to develop acceptance criteria that could be implemented by software vendors and checked by practitioners.

## Examples

The first example is from a bench in a copper mine. The data is taken from blasthole data. Figure 1 shows the locations of the data. The blasthole data were resampled to use the data closest to a 50m grid. This was done to create more realistic kriging and simulation results. Figure 2 shows the distribution of the copper grades from the data set. Figure 3 shows the variograms calculated from the data. They are used to build the models we compare. Figure 4 shows the first models built. The first plot shows the kriging estimates created using simple kriging with 24 data. It was created using the normal scores variogram, a very large search radius, and a 50x56 grid. The second plot shows an average of 100 simulated realizations. For these realizations, sgsim did not transform the data, but used the already transformed data and the normal score variogram to create the realizations with the same grid as for kriging. The scatterplot is a comparison of the two previous plots, comparing the estimates generated by each method. Figure 5 is a comparison of the back transformed data. These figures were created using the data output in the creation of the previous plots. The first plot shows the back transformation of the estimates created by simple kriging previously. The back transformation was done using PostMG. The units are in the same units as the original copper grades. The second plot was created by back transforming all 100 realizations and then taking the average of those realizations. The first scatterplot compares these two plots. The last scatterplot shows what happens when this process is done incorrectly. It was generated by simultaneously using the wrong back transformation program for the kriged estimates and by averaging the simulated realizations before back transforming. It is important that these steps be done in the correct order when dealing with a non-linear transform. Figure 6 compares kriging and simulation using the untransformed data. The first plot is the simply kriged estimates. These estimates were created using the raw copper grades variogram, a very large search radius, and a grid 50x56. The second is the average of 100 simulated realizations where simulation did perform the data transformation and back transformation. The normal score variogram was used on the same grid as for kriging. The scatterplot compares the results of these two plots. Figure 7 compares another type of kriging, ordinary kriging, with simulation. In this case, ordinary kriging is limited to using 24 data. The first plot shows the estimates generated by ordinary kriging in original units with the original units variogram. A large search radius was used and the grid is 50x56. The second plot compares this kriging with simulation. The third plot shows the difference between the two methods. Values greater than zero occur when OK generates an estimate greater than the simulation average and vice versa. The last plot compares the difference with the estimate generated by OK. Figure 8 shows a comparison of ordinary kriging with 12 data to simulation. The first plot shows the estimates generated by ordinary kriging in original units with the original units variogram. The second plot compares this kriging with simulation. The third plot shows the difference between the two methods. Values greater than zero occur when OK generates an estimate greater than the simulation average and vice versa. The last plot compares the difference with the estimate generated by OK. Figure 9 is a comparison of ordinary kriging with 6 data and simulation. The first plot shows the estimates generated by ordinary kriging in original units with the original units variogram. The second plot compares this kriging with simulation. The third plot shows the difference between the two methods. Values greater than zero occur when OK generates an estimate greater than the simulation average and vice versa. The last plot compares the difference with the estimate generated by OK.

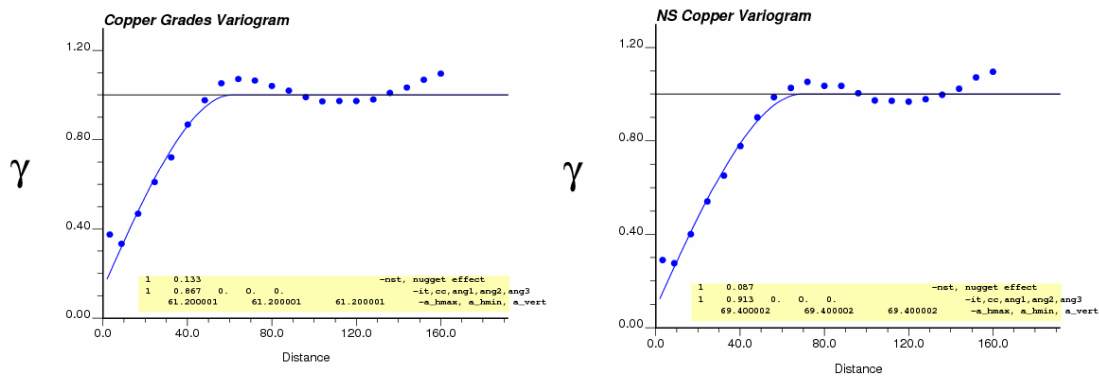
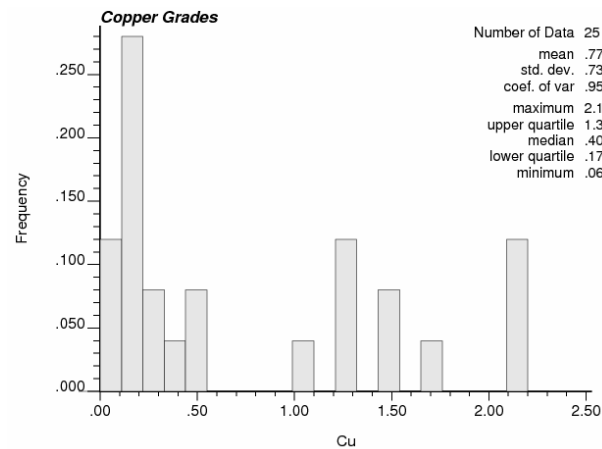
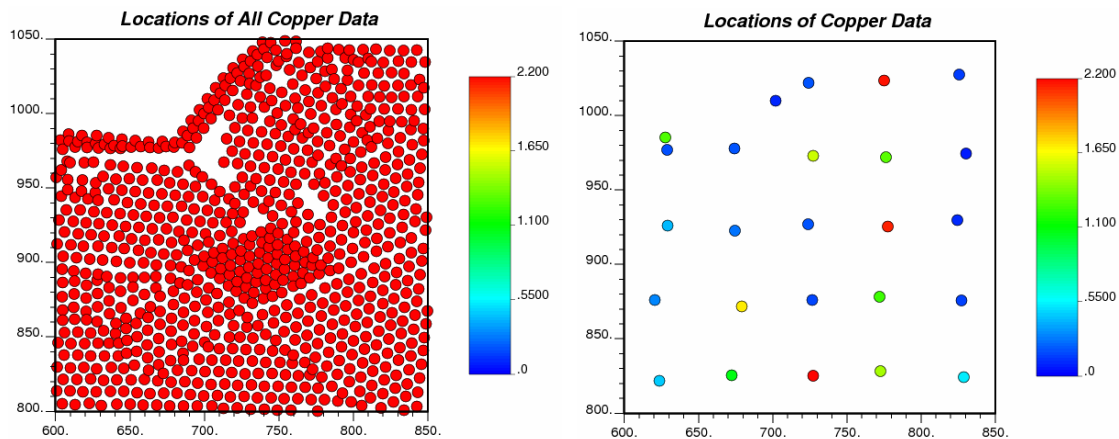
The second example is from a 5m thick slice from mitch.dat. All of the data fall between 850 and 855m elevation. Figure 10 shows the locations of these 169 data and Figure 11 shows the distribution of copper grades from this data set. Figure 12 shows the variograms calculated from the copper grades and normal scores. Figure 13 compares simple kriging and simulation using the normal scores. The first plot is of the simply kriged normal score transformed grades. The normal score variogram is used with a very large search radius and 24 data. The second plot is the average of 100 simulated realizations. The sgsim program did not transform the data itself, but used the already normal score transformed data. The normal score variogram was used in the calculations with a very large search radius and 24 data. The scatterplot compares the two methods. There is fairly good correlation between the two. Figure 14 is a comparison of the back transformation of each method. The first plot was created by using PostMG to back transform the results from simple kriging to get back into the original units. The second plot was created by back transforming and then averaging the 100 realizations created by simulation. The scatterplot compares the two back transformed methods of creating a map of estimates. Due to the instability of the variogram calculated from the copper grades, the additional comparisons are not done.

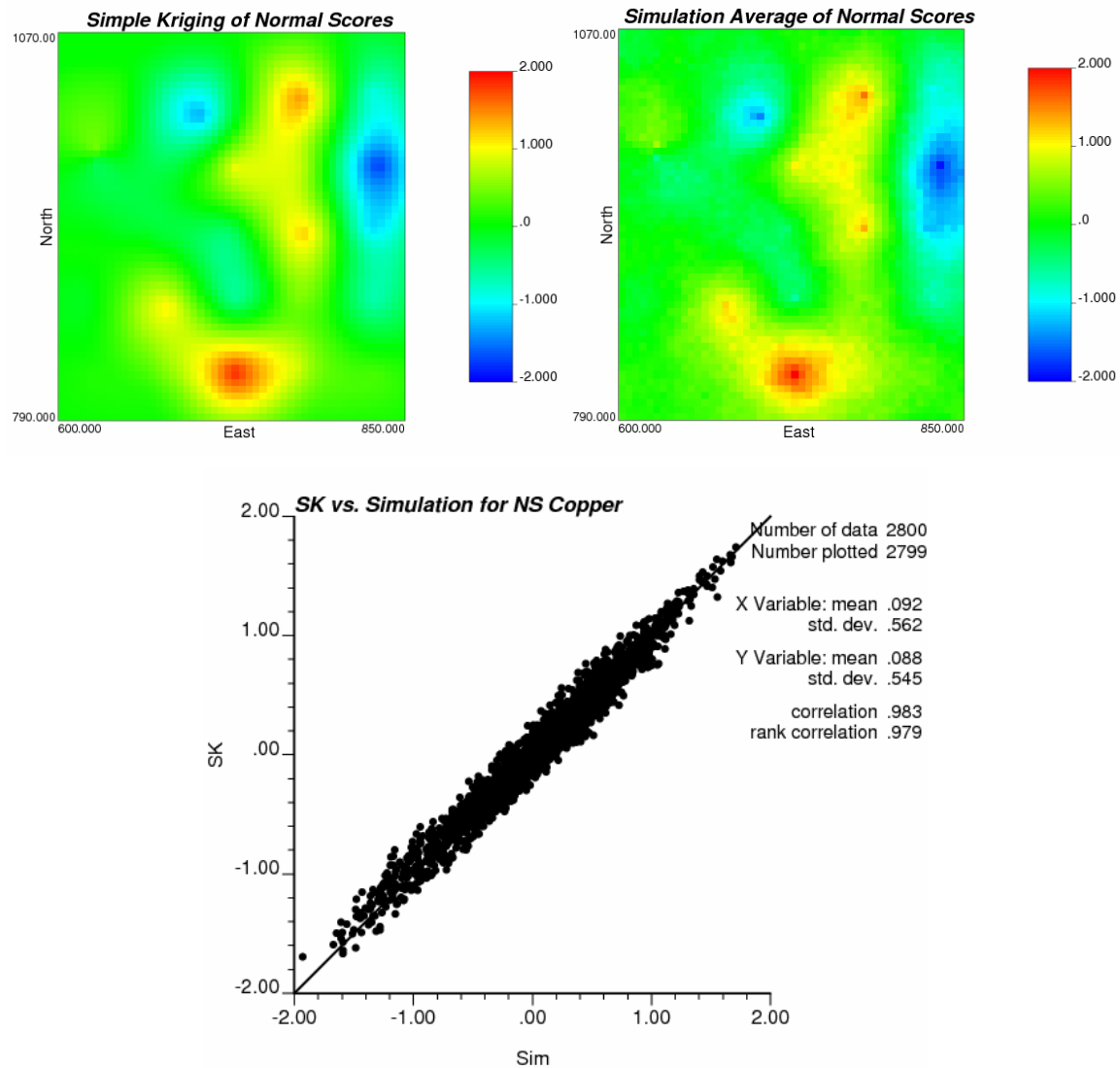
The last example comes from a set of data called red.dat. Figure 15 shows the locations of the data and the data distribution. Figure 16 shows the variograms calculated from the gold grades and from the normal scores. They are isotropic. Figure 17 is a comparison of simple kriging and simulation using the normal scores. These plots were created using the normal score transformed data and variogram. The first is the simple kriging estimates using a large search radius and 24 data. The second is the average of 100 simulated realizations also using a large search radius and 24 data. Sgsim did not transform the data. The scatterplot compares the results of these two methods. Figure 18 is a comparison of the back transformation of the two modeling methods. The simple kriging was back transformed using PostMG and those results are plotted. The 100 simulated realizations were back transformed and then their average was calculated and it is shown. The scatterplot compares the results of these back transformations. Figure 19 is a comparison of simple kriging and simulation using the gold grades directly. In the first plot, kriging was performed with a very large search radius, 24 data, and the variogram calculated from the raw grades. Simulation was done on the raw grades with a very large search radius and 24 data. The normal score variogram was used as simulation transforms the data internally. The plot is the average of 100 realizations. The scatterplot compares these two methods. Figure 20 compares ordinary kriging and simulation. The number of data used to generate estimates is 24. The first plot is ordinary kriging done on the raw data with a very large search radius using the variogram calculated from the actual grades. The adjacent scatterplot compares this type of kriging with the simulation average computed from the raw grades. The third plot is of the difference between ordinary kriging and simulation and the fourth plot is a comparison of the differences and the ordinary kriging estimate. Figure 21 is a comparison of ordinary kriging using 12 data and simulation. The first plot shows the results of this kriging. The variogram calculated from the actual grades is used. The second plot compares this kriging with the average of 100 simulated realizations created using the actual grades. The third plot is the difference between the methods and the fourth plot compares the differences with the ordinary kriging estimates. Figure 22 is a comparison of ordinary kriging using six data and simulation. The first plot is the results of doing ordinary kriging with a very large search radius and 6 data. The variogram calculated from the actual grades is used. The second plot compares this kriging with the average of 100 simulated realizations. The third plot shows the differences between the two methods. The fourth plot compares the differences with the ordinary kriging estimates. These are all of the examples done for this paper.

## Conclusions

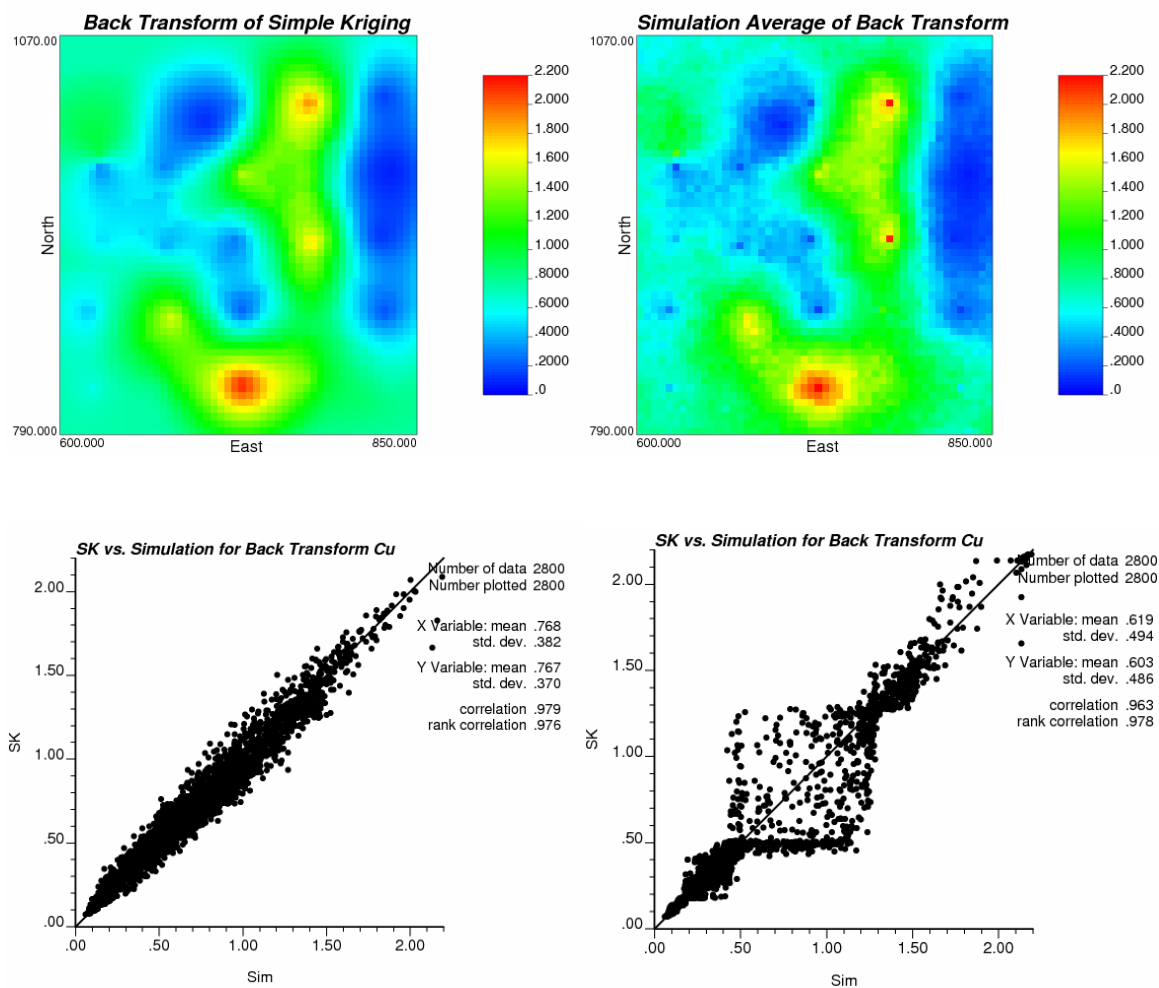
We can see from these examples that it is incorrect to assume that the average of simulated realizations is the same as kriging, especially ordinary kriging. The different variograms, type of kriging and search strategy cause the difference. These differences are expected and not a sign of problems.

There are three recommended modes of checking: (1) check the average of simulation in Gaussian units with kriging in Gaussian units, (2) check the average of back transformed simulated values with the results of PostMG, and (3) check the average of back transformed simulated values with conventional kriging practice. We expect a very close agreement in the first two cases. The closeness of the third check will depend on many implementation considerations discussed above. Future work consists of quantifying the differences we can expect and accept.



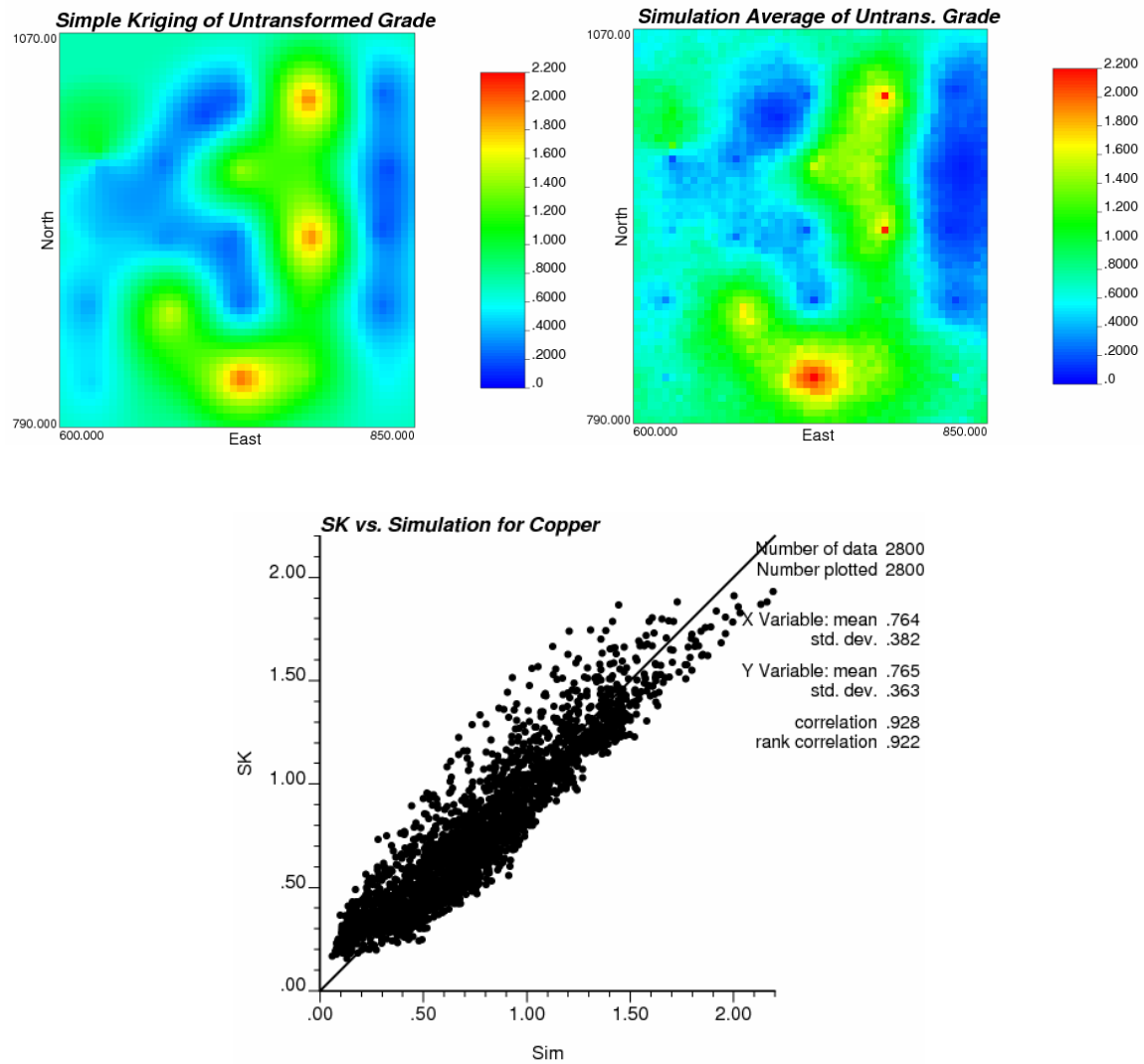


**Figure 4:** These figures were all created using the normal score transformed copper data.

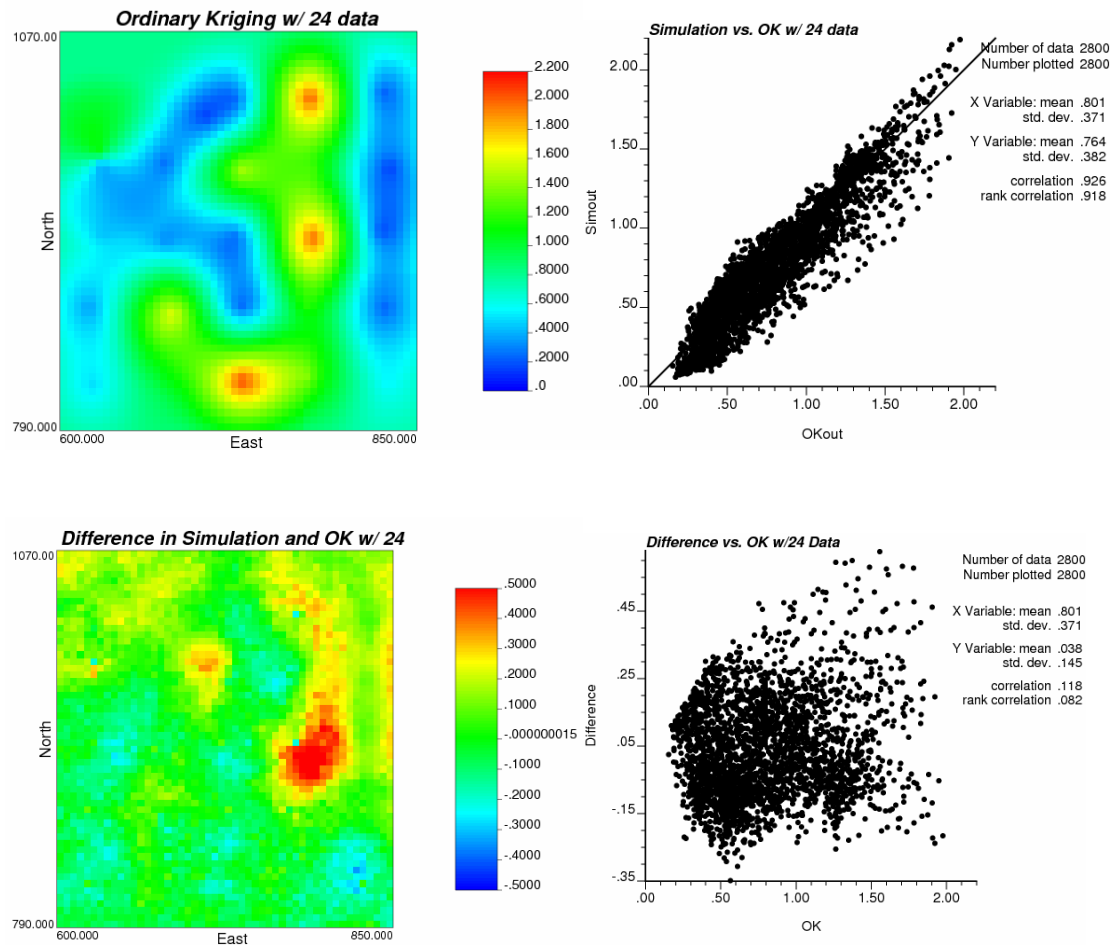


**Figure 5:** Comparison of Back Transformed Data

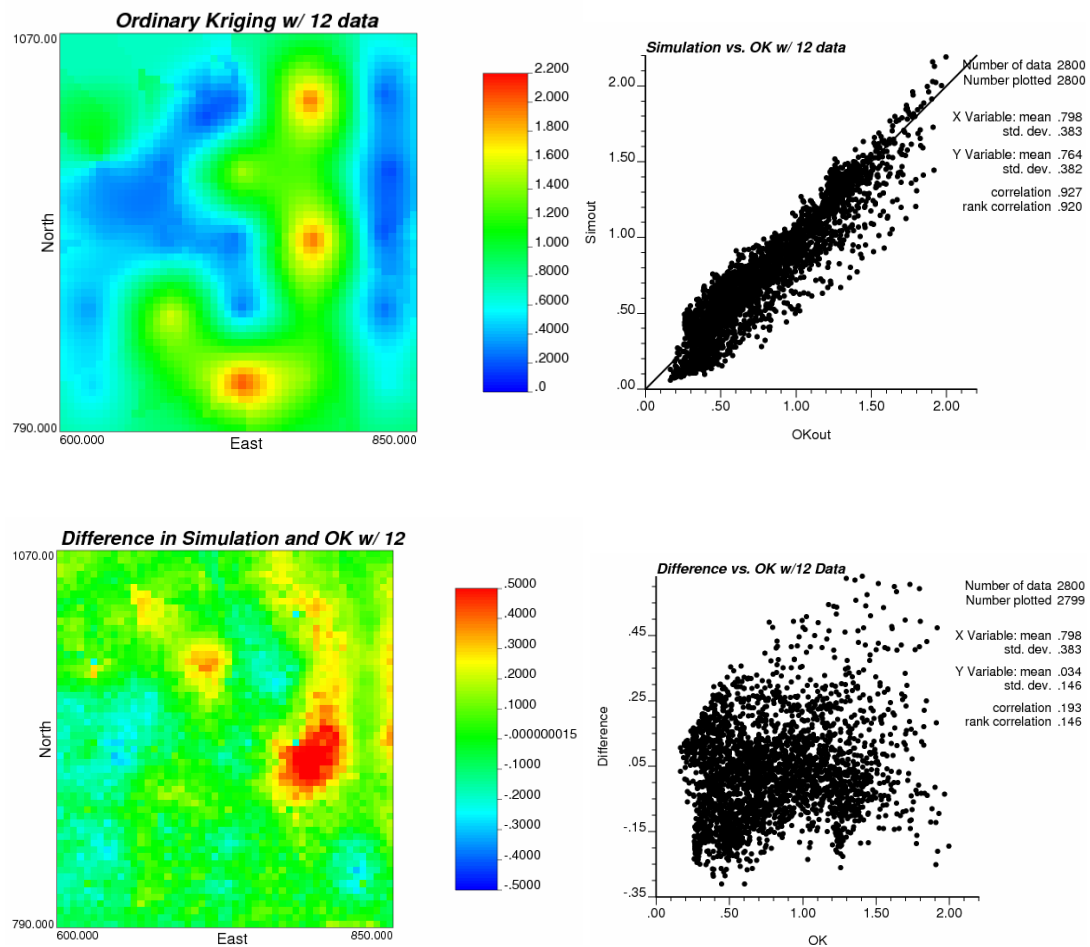




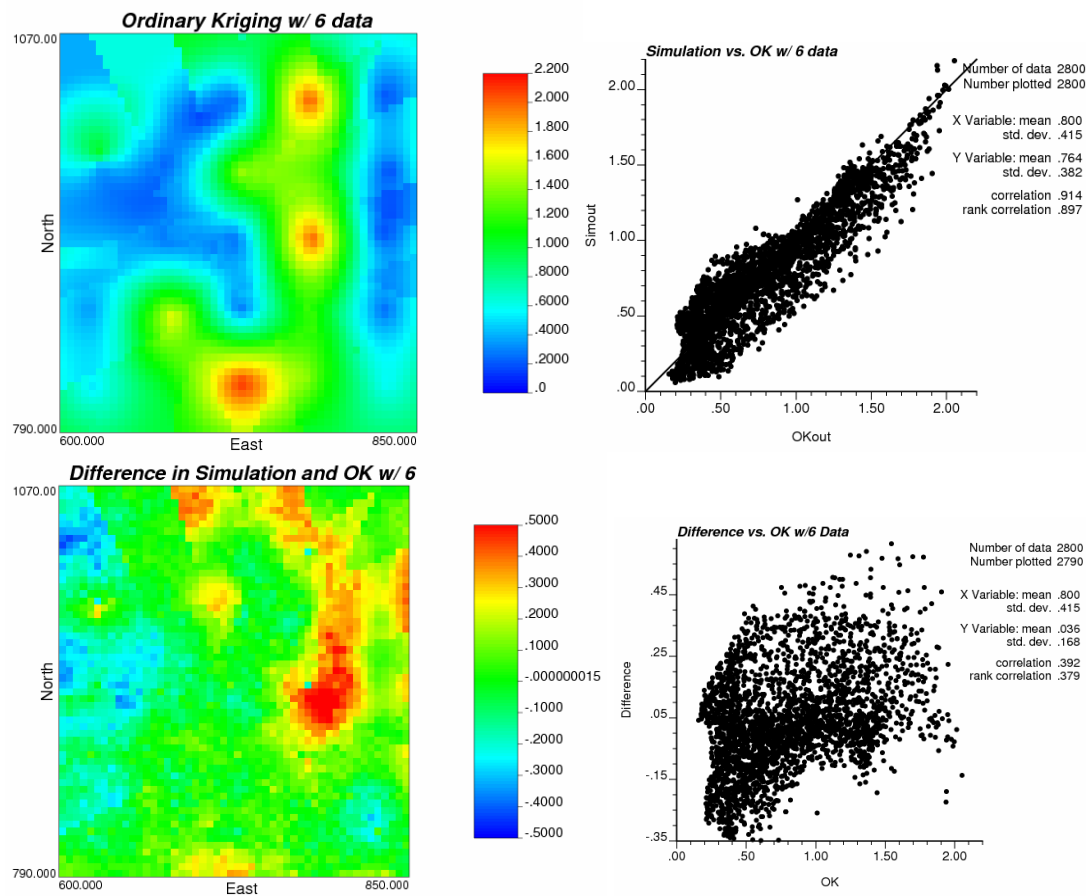
**Figure 6:** These models were built using the original untransformed data.



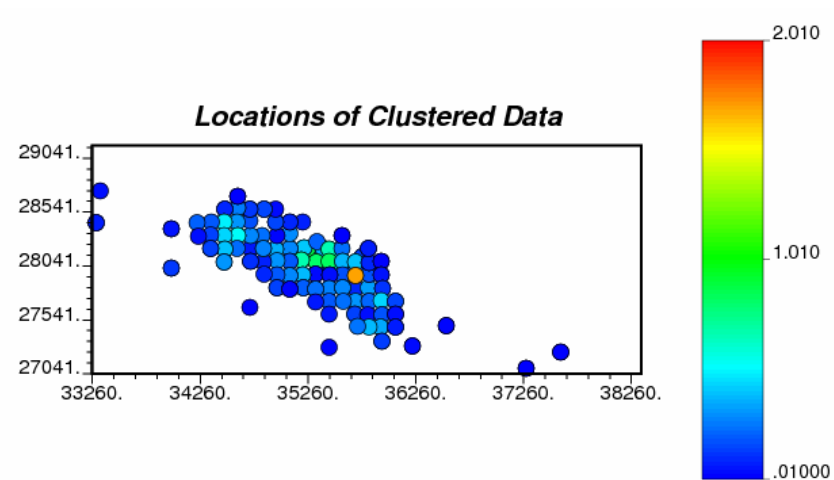
**Figure 7:** A comparison of Ordinary Kriging with 24 data and simulation.



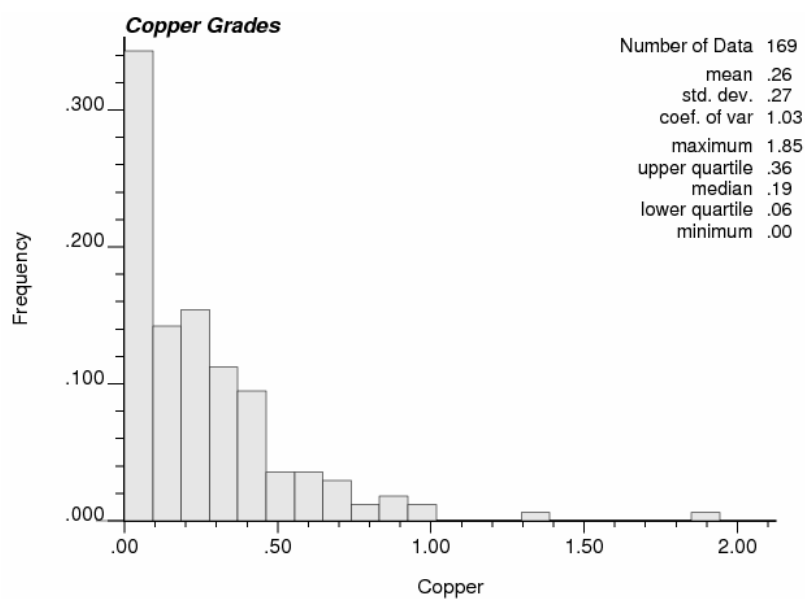
**Figure 8:** A comparison of ordinary kriging with 12 data and simulation.



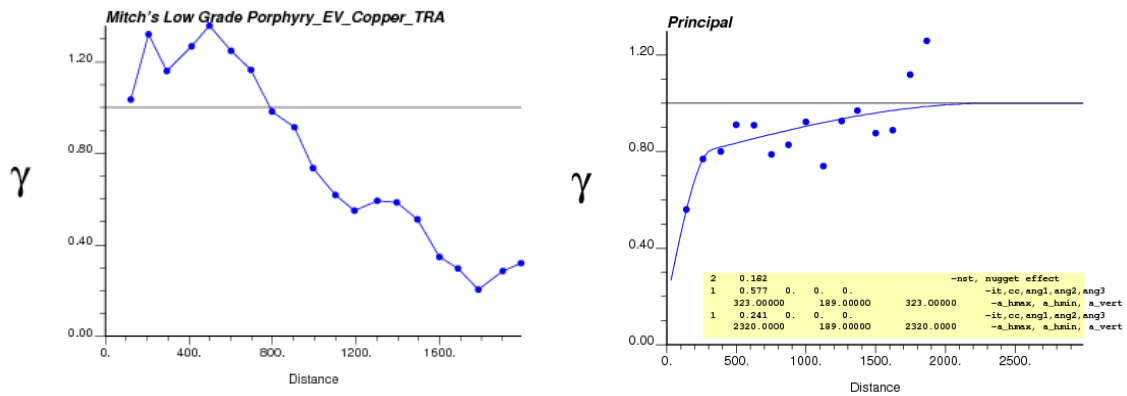
**Figure 9:** A comparison of ordinary kriging with 6 data and simulation.



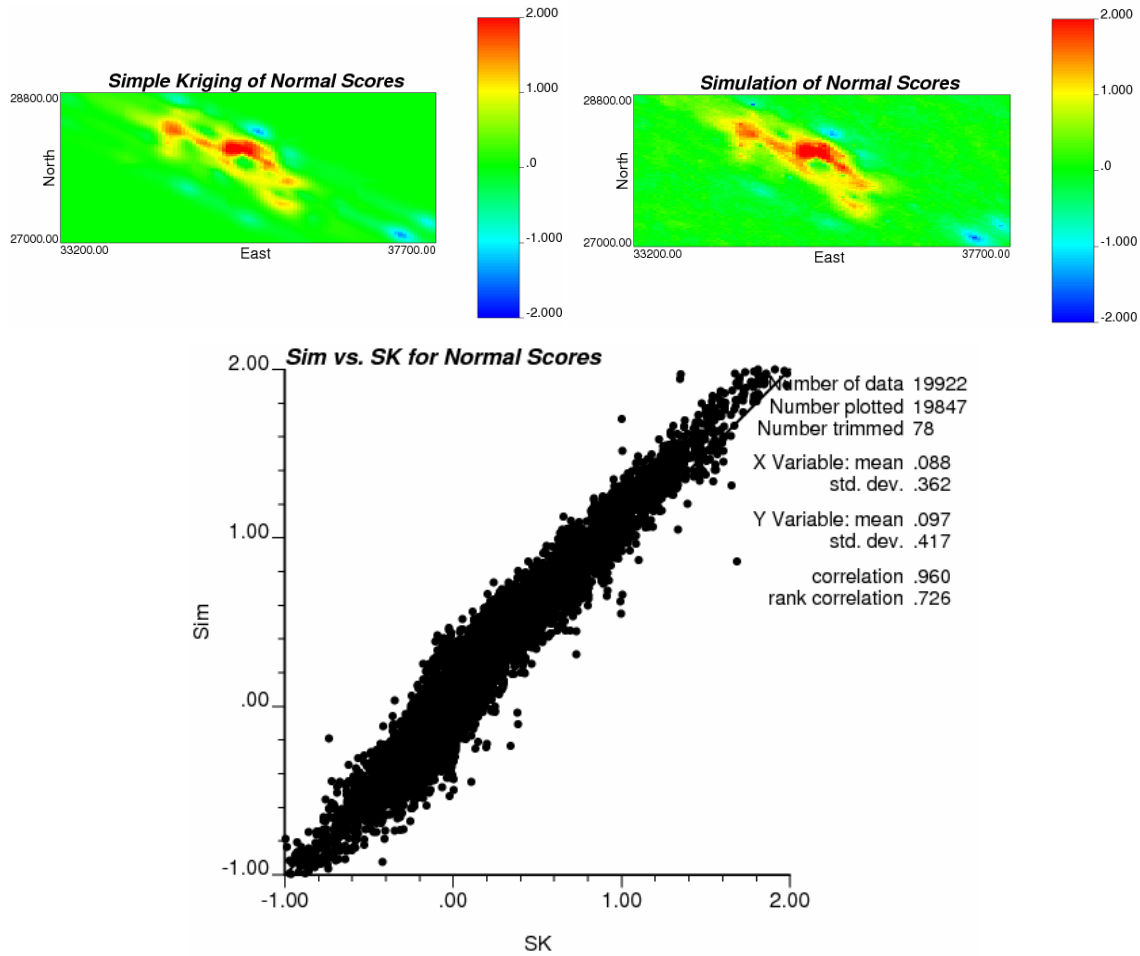
**Figure 10:** The locations of the data in the 5m slice.



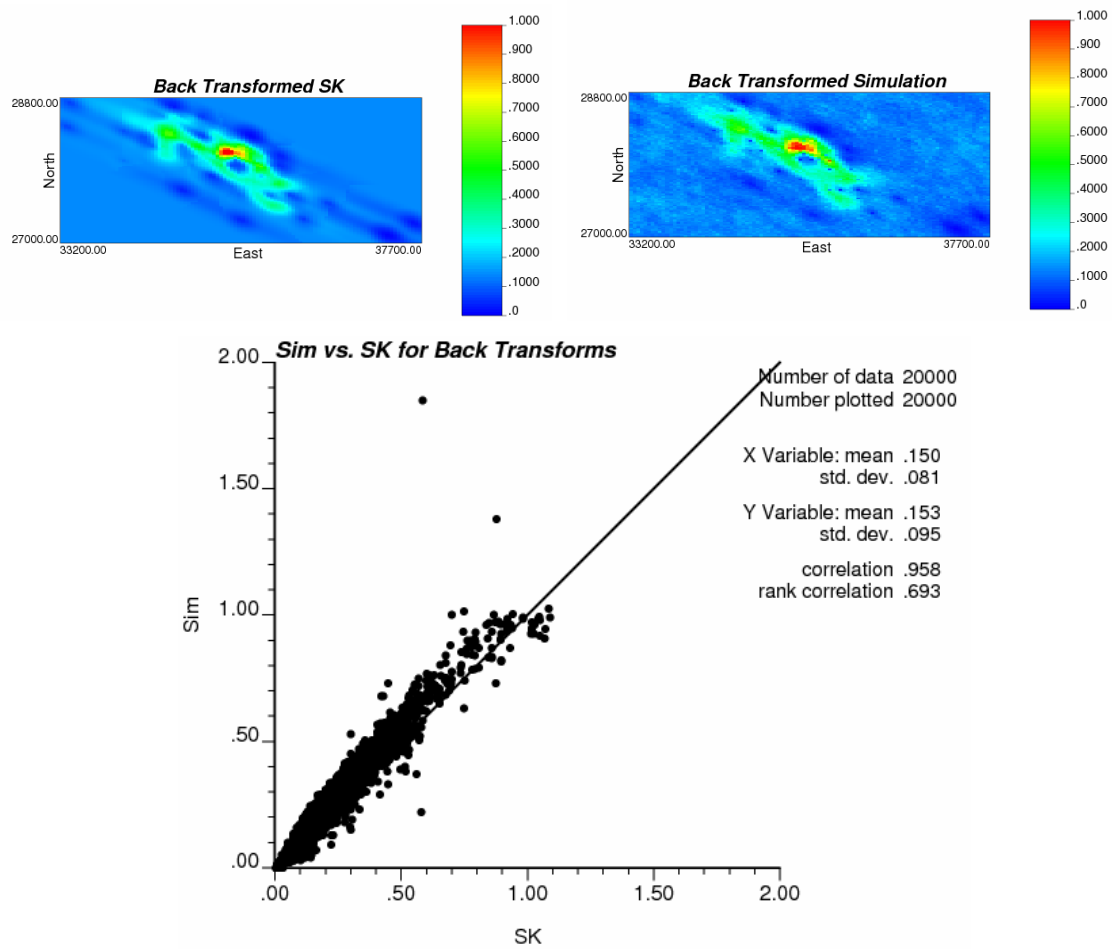
**Figure 11:** The distribution of copper data.



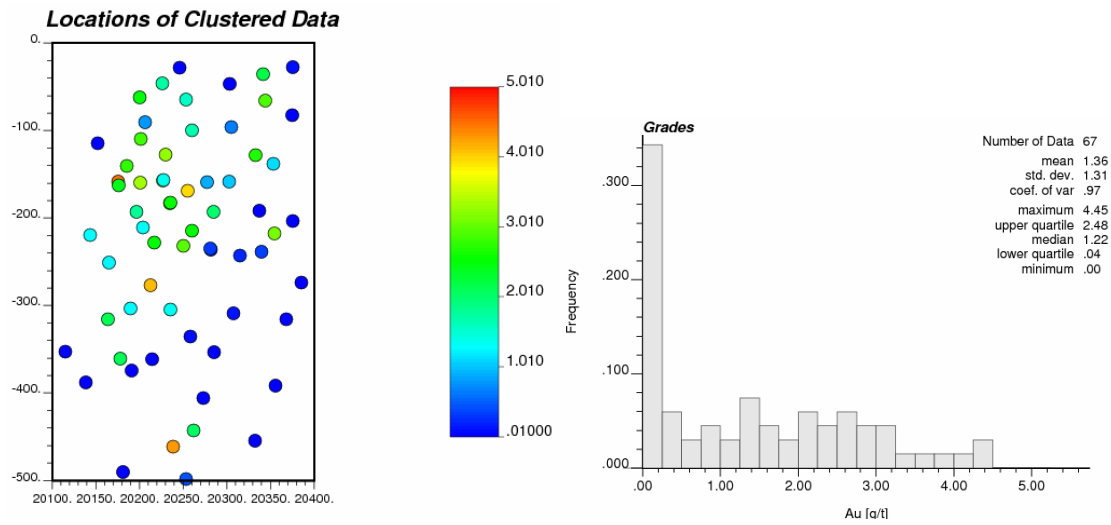
**Figure 12:** The anisotropic variogram calculated from the copper grades and normal scores respectively.



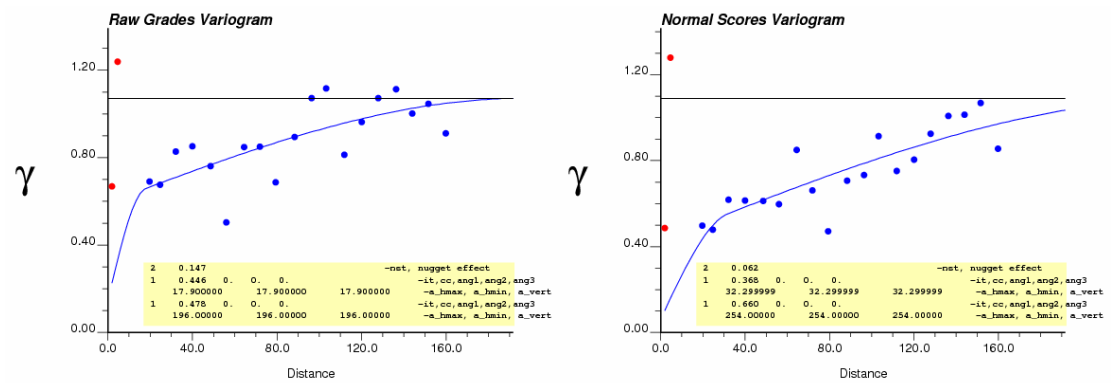
**Figure 13:** A comparison of simple kriging and simulation using the normal scores.



**Figure 14:** A comparison of the back transformations for each method.

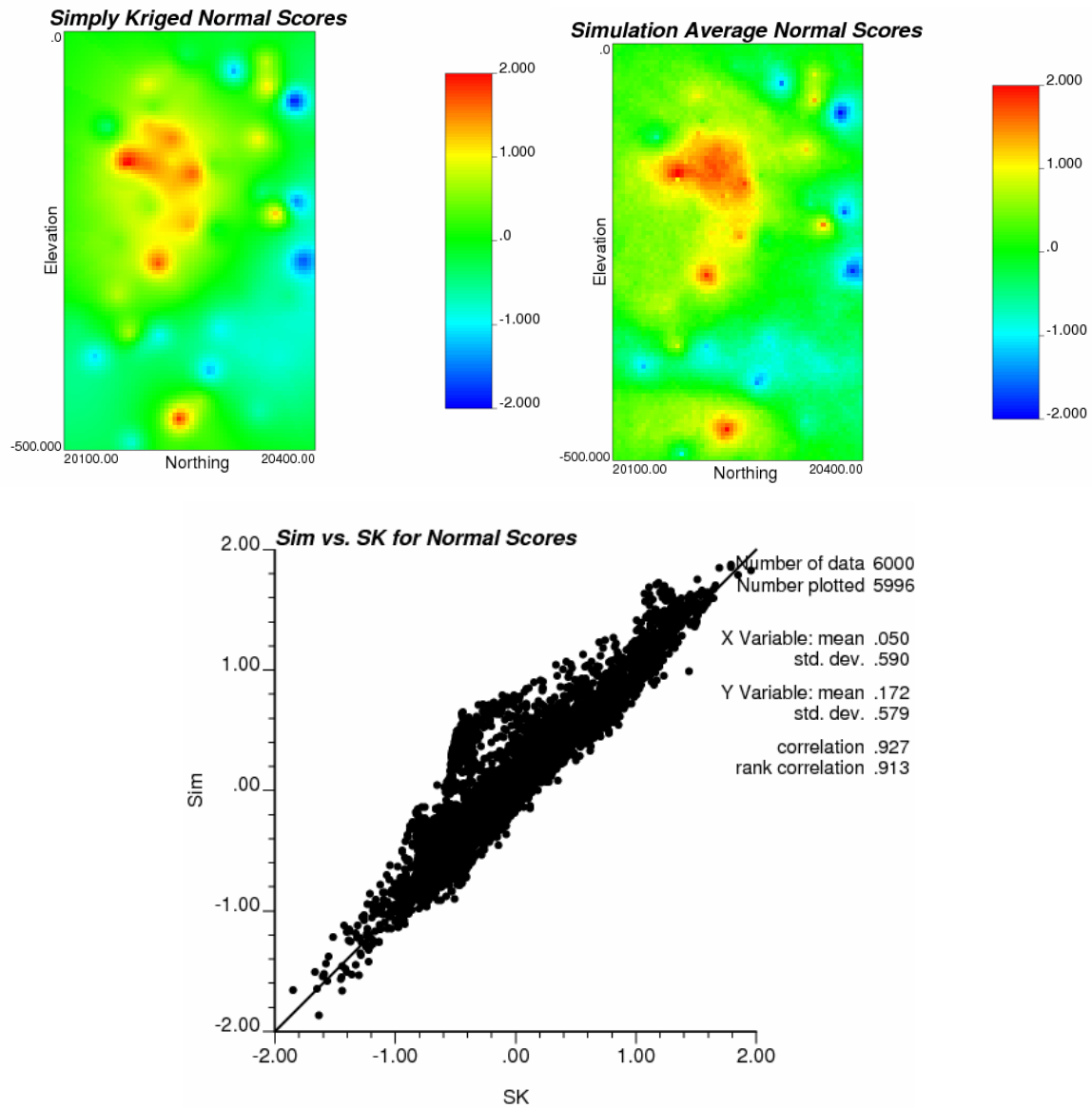


**Figure 15:** Data locations and distribution.

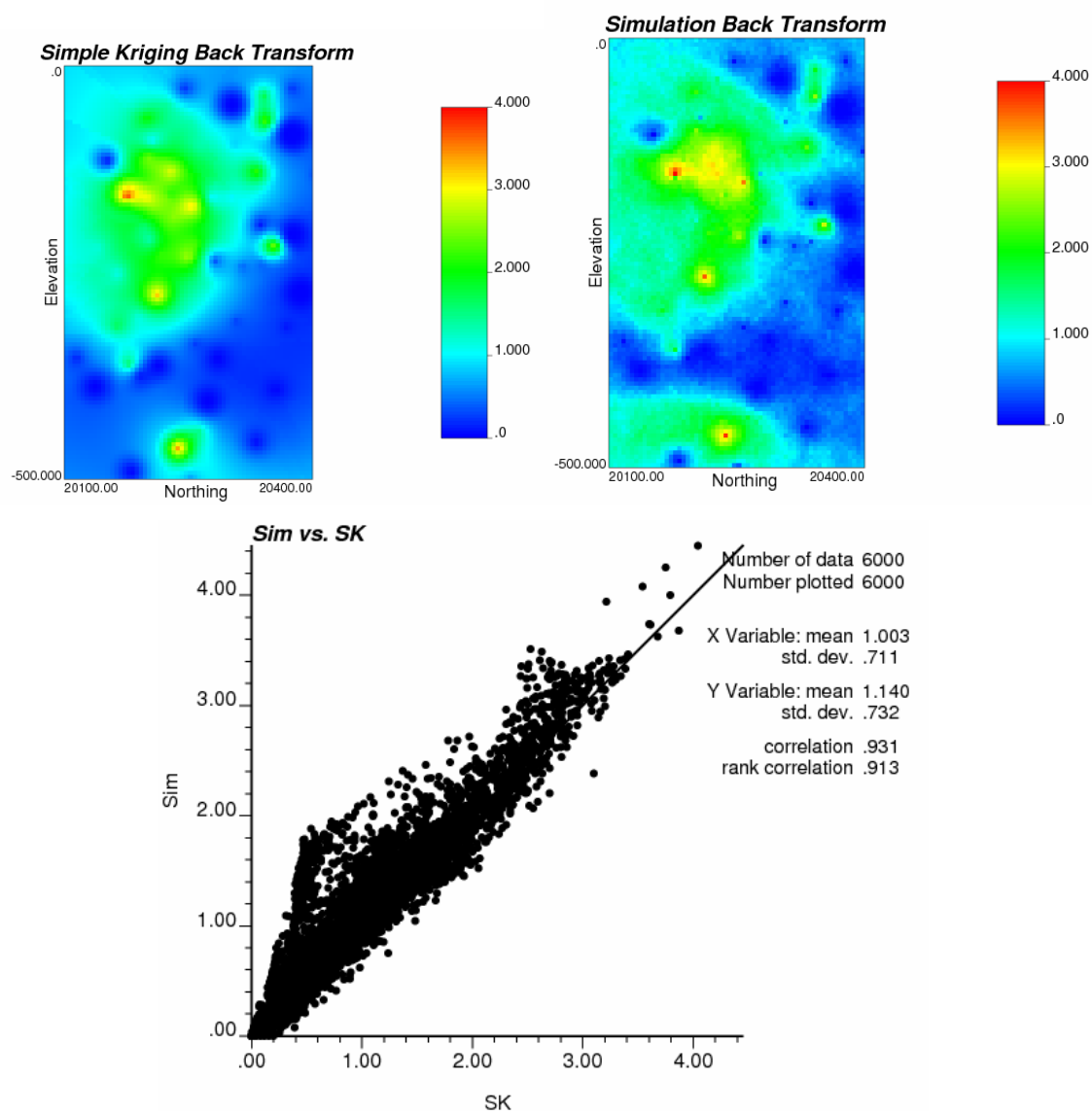


**Figure 16:** Variograms calculated from the gold grades and from the normal scores.

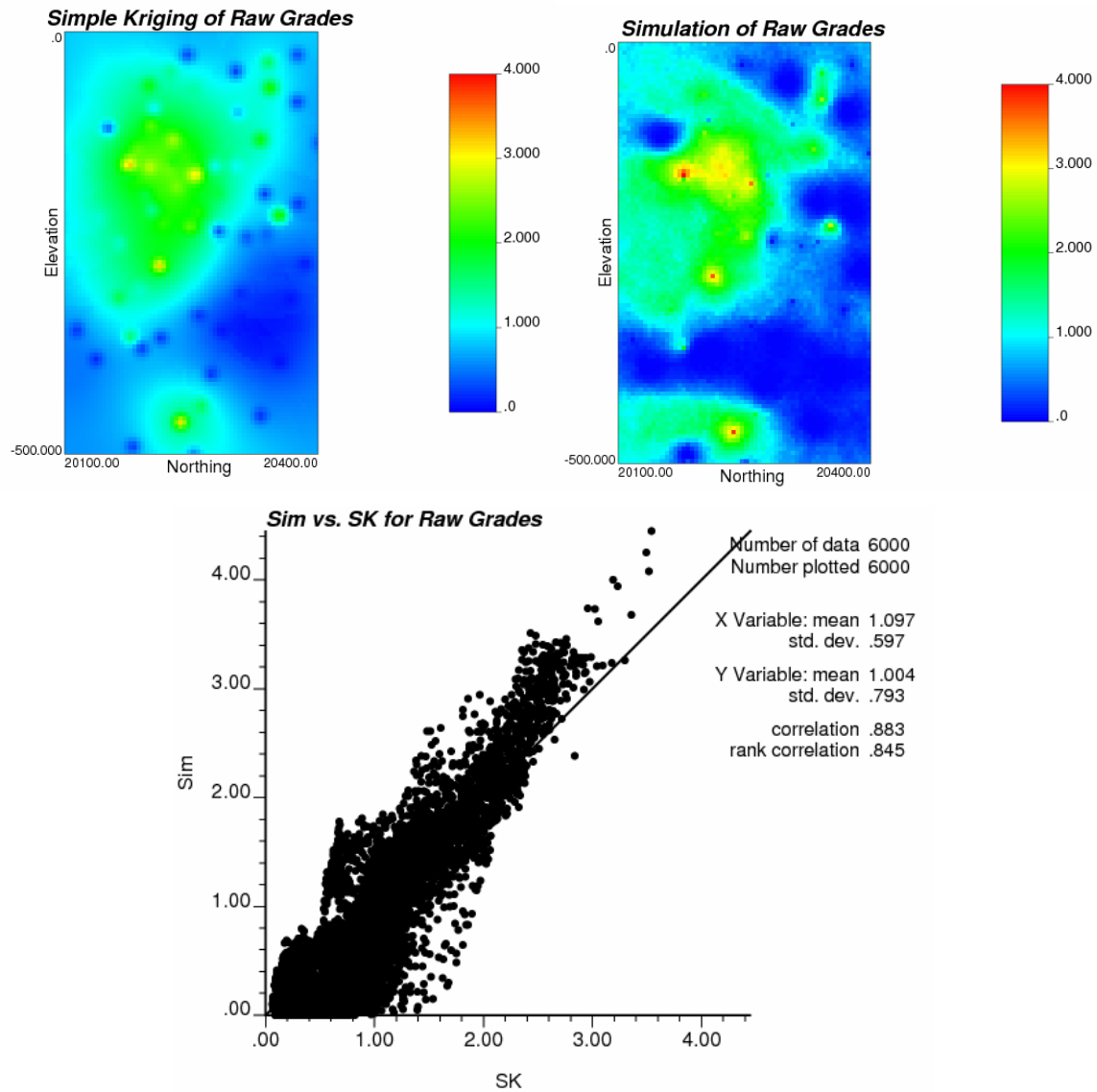




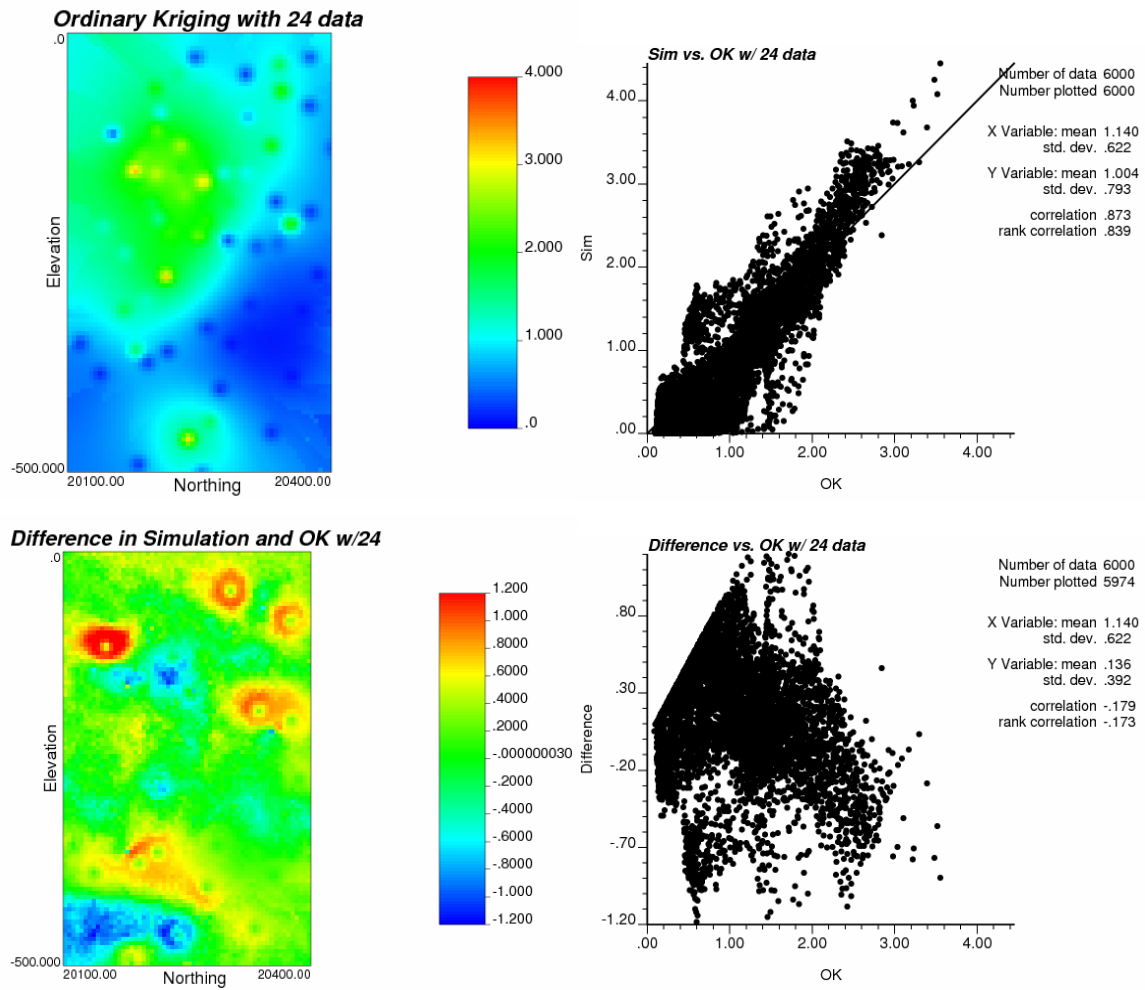
**Figure 17:** A comparison of simple kriging and simulation using the normal scores.



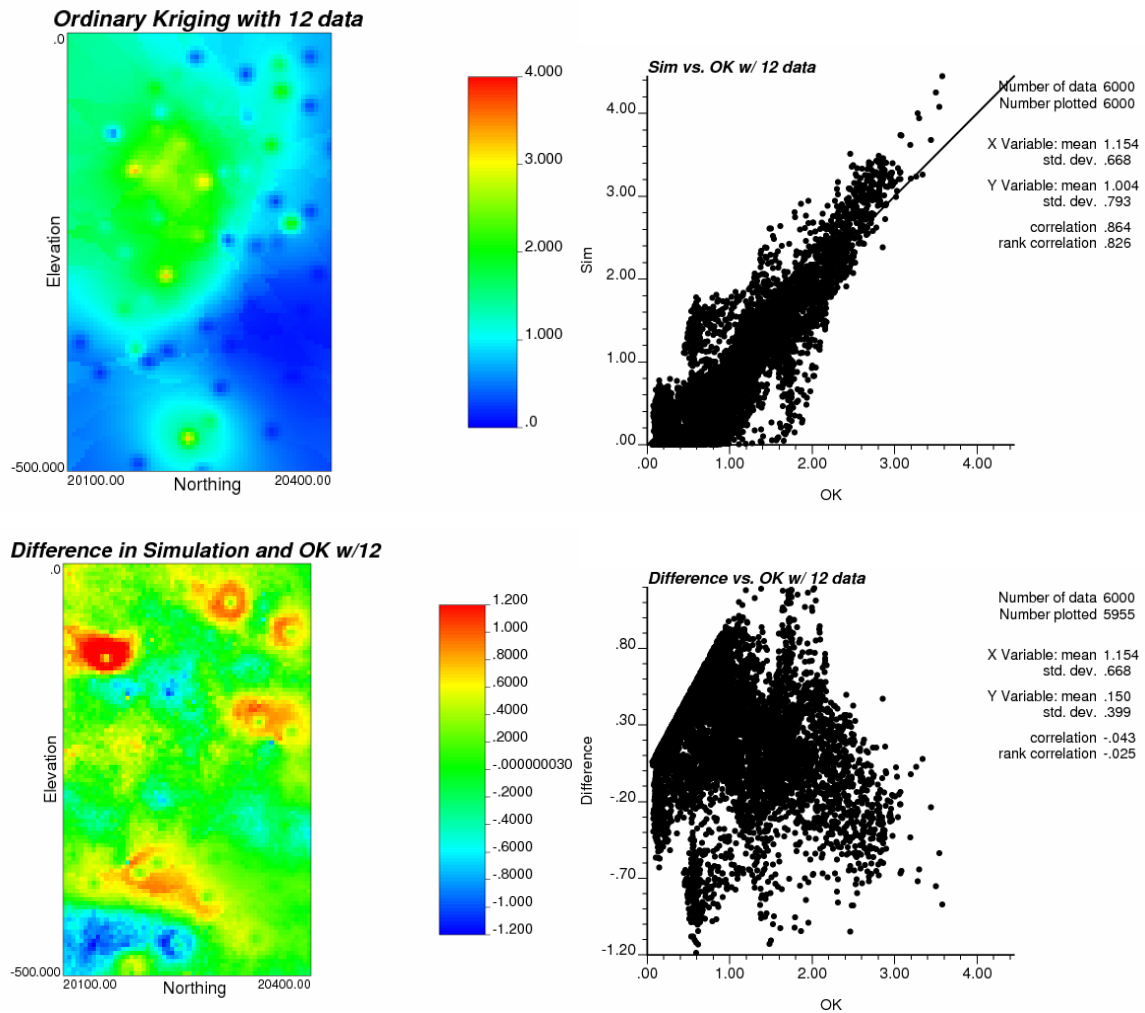
**Figure 18:** A comparison of the back transformations for each modeling method.



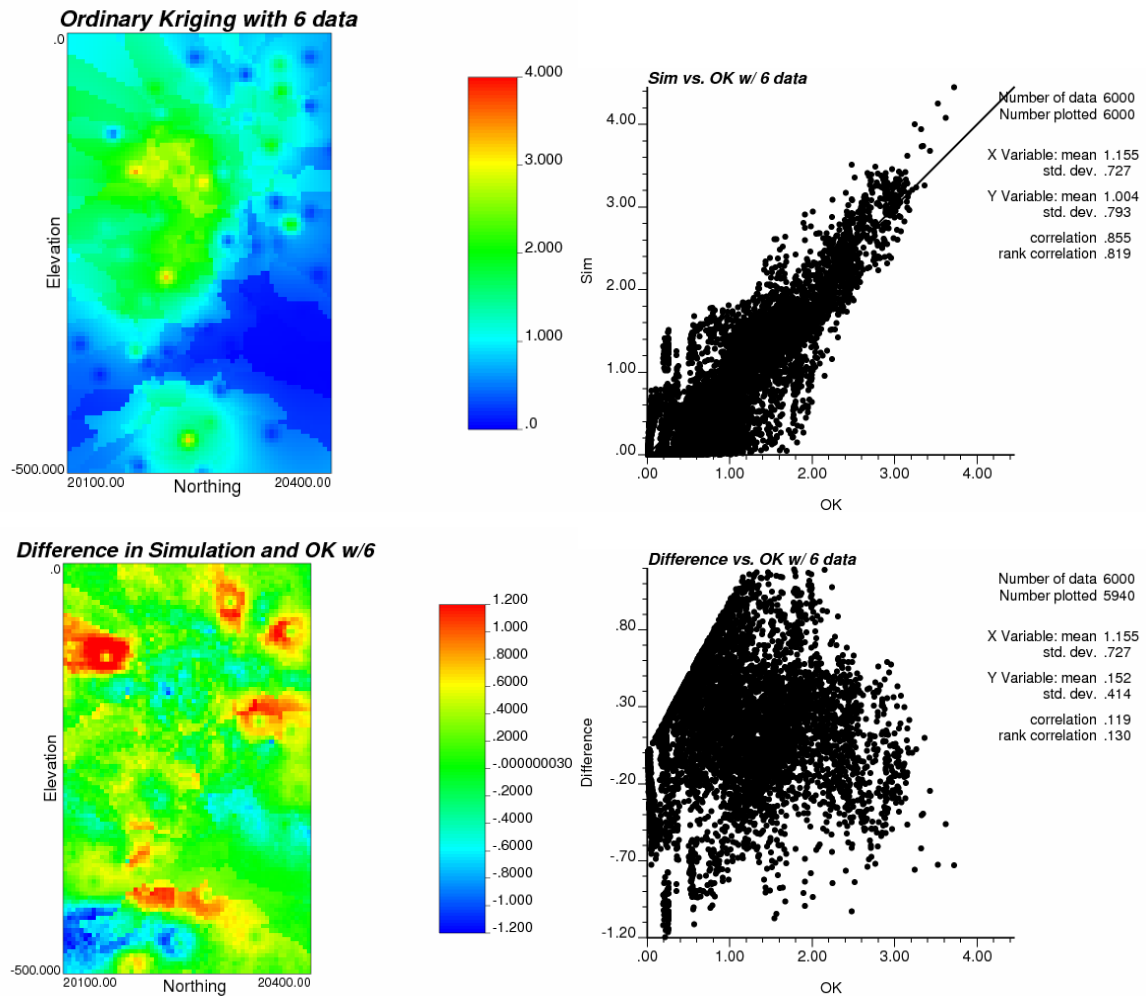
**Figure 19:** A comparison of simple kriging and simulation using the gold grades directly.



**Figure 20:** A comparison of ordinary kriging limited to 24 data and simulation.



**Figure 21:** A comparison of ordinary kriging with 12 data and simulation.



**Figure 22:** A comparison of ordinary kriging with 6 data and simulation.