

A New Approach to Calculate a Robust Variogram for Volume Variance Calculations and Kriging

(deriving the Z-variogram from the Y-variogram)

Brandon Wilde and Clayton V. Deutsch

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

An important problem in Mining Geostatistics and dealing with multiscale data is inference of a reliable small-scale variogram of the original variable (Z). Robust alternatives such as the pairwise relative variogram, correlogram and normal scores variograms converge to incorrect values. Our goal is to establish a robust calculation scheme where the values converges to the correct result, that is, there are no known biases in the variogram that will be used for subsequent volume variance and kriging steps.

Introduction

The variogram of original grades is notoriously unstable. Clustered data and outliers are the main source of instability. Relative variograms and correlograms address these issues in the definition of the calculated statistic. Normal-score and median indicator variograms address these issues by a transformation of the data. Neither a redefinition of the spatial correlation measure nor a non-linear transformation of the data is theoretically acceptable. The stationary covariance or stationary variogram is required for volume variance calculations and estimation of the original variable.

Our concern with robust measures (relative variograms, correlograms, normal score variograms and median indicator variograms) is that they often show short scale structure that is very different from the original variable variogram. Of course, in presence of sparse and clustered data, we have no access to the original variable variogram and it is better to have an approximation of the spatial structure than no structure at all. This justification has sufficed for many years – robust measures are systematically used by many practitioners with no accounting for their bias. The goal of this paper is to use a robust measure, but to adapt a theoretical correction to ensure that it converges to the correct stationary covariance/variogram.

Basic Idea

The normal scores (Y) variogram is very robust because the Gaussian transform (1) ensures that there are no outliers, and (2) removes the proportional effect, which mitigates the effect of clustered data (see Figure 1). The Y-variogram, however, is biased with respect to the original variable (Z) variogram. The bias is known in the case of the lognormal distribution, see the paper of Manchuk and others in last year's CCG report. The key idea of this note is to infer the Y-variogram and then back transform it to an unbiased Z-variogram. This back transformation of the variogram is referred to as 'mapping' (see Figure 2). The back transformation is non-trivial because we must consider each bivariate distribution and perform the back transformation with Monte Carlo Simulation.

The first step is to transform the original variable by the normal score transform. Variograms are then calculated using the normal score data. Accurate models of these variograms are then created using either a variogram fitting algorithm or the guess and check method of creating a variogram model. It is important to create as accurate a model as possible. Once the normal scores variogram has been correctly modeled, it is used to ‘map’ a variogram of the original variable. This mapping is performed one lag distance (h) at a time until each lag is mapped. The mapping process begins by first determining the correlation, rho (ρ), at a given ‘h’. The correlation is obtained from the normal scores variogram by the following equation:

$$\rho = 1 - \gamma_y \quad (1)$$

A large number of paired samples are then taken from a bivariate Gaussian distribution characterized by ρ . This bivariate Gaussian distribution is sampled using Monte Carlo Simulation. Two values, Y_1 and Y_2 , are drawn randomly from the distribution. They are then given the appropriate correlation by the following equation:

$$Y_h = \rho \cdot Y_1 + \sqrt{1 - \rho^2} \cdot Y_2 \quad (2)$$

Y_h and Y_1 are kept as the values of interest as they have the correct correlation. These two values are then back transformed to Z_1 and Z_2 according to the initial normal scores transformation performed on the original variable. The squared difference of Z_1 and Z_2 is calculated for each of the paired samples. The average of these squared differences is taken to be the value of the original variable variogram for the specific ‘h’. This process is then repeated for each ‘h’ until the complete variogram is mapped. This variogram can then be fit and used for modeling purposes. It is unbiased and much more stable than the variogram calculated directly from the original variable.

Program

The vmodel program was revised to perform the calculations. See the parameters below

```

START OF PARAMETERS:
vmodelyz.var          -file for variogram output
2 40                  -number of directions and lags
 0.0 0.0 2           -azm, dip, lag distance
90.0 0.0 2           -azm, dip, lag distance
nscore.trn           -file with input transformation table
0.0 2.19             - minimum and maximum data value
1 0.0                - lower tail option and parameter
1 2.19               - upper tail option and parameter
1 0.087              -nst, nugget effect
1 0.913 0.0 0.0 0.0 -it,cc,ang1,ang2,ang3
                    69.4 69.4 69.4 -a_hmax, a_hmin, a_vert

```

The output of the program has also been modified. It outputs columns containing the lag number, lag distance, NS variogram value, mapped variogram value, standardized mapped variogram value, the number of directions calculated, rho, and the sample variance used to standardize the mapped variogram. Column labels were also added to the output. In order to plot this modified output, vargplt was also modified to include a parameter for the column number you want to plot.

This new program is called `vargpityz`. It has also become necessary to be able to fit the standardized mapped variogram in order to use it for kriging purposes. To do this, `varfit` was modified to allow the input of a column number. This modified fitting program is called `varfityz`. Using these three modified programs, you can successfully create and model a variogram to be used for kriging with the original variable.

Examples

A number of examples in the presence of varying data densities were considered. Figures 2-5 report the results from a slice taken from `mitch.dat`. There are 169 data in an area 2 x 4.5 km. Figure 3 demonstrates the difference between the variogram calculated from the original variable and from the normal scores of that variable. As is usually the case, the variogram calculated from the original variable is unstable. Figure 4 shows the fitted normal scores variogram and the original grades variogram mapped from the normal scores variogram. It was mapped using the new program, `vmodelyz` and plotted using `vargpityz`. Figure 5 shows the mapped variogram along with the fitting of it. Also shown are all four variograms together, that is, the mapped variogram, the fitted mapped variogram, the fitted original grade variogram (hard to see, is right along the sill) and the fitted normal scores variogram. This last variogram is the one with the lowest nugget and longest range. Figure 6 compares the two variograms by using them to kriging with otherwise identical parameters. The kriging done with the unmapped variogram is very erratic and has many artifacts due to the variograms instability. The kriging done with the mapped variogram is much cleaner and contains less artifacts because the variogram is more stable.

The data used for Figures 6-8 provide a good check as they come from closely spaced blast hole data. The variogram calculated from the original grades is stable and easily modeled due to the abundance and close spacing of data so it is very similar to the normal scores variogram. If the mapped variogram is similar to the variogram calculated from the copper grades, we know our program is correct. Figure 7 shows the similarity of the variograms calculated from the original grades and the normal scores. The normal scores variogram does have a slightly longer range and lower nugget than the other, which is expected. Figure 8 shows the fitted normal score variogram and the original grades variogram mapped from it. Figure 9 shows the fitted and mapped variogram. It also shows all four variograms plotted together. They are all very similar. As in the previous example, the fitted normal score variogram has the lowest nugget effect. The mapped variogram, which should be very similar to the fitted original grades variogram in this case, shows slightly lower variography at greater distances.

The data used for Figures 9-11 are from `red.dat`. The data are located by Northing and Elevation coordinates. They are along an azimuth of zero with various elevations. It is a 2-D dataset with no east-west component. Figure 10 shows the variograms calculated directly from the data. The variogram from the gold grades is shown first along with the variogram calculated from the normal scores. Figure 11 shows the fitted normal scores variogram which was used to map the variogram shown next to it. Figure 12 shows the mapped variogram overlain by its fitted variogram. It also shows all four variograms together. From this figure we see that the mapped variogram falls between the variogram calculated from the gold grades above it and the variogram calculated from the normal scores below it. This is to be expected.

Figures 12-14 are from `amoco.dat`. It contains porosity information. In this case, the variogram calculated from the porosity has lower variability than the variogram calculated from the normal scores. Figure 13 shows the variograms calculated directly from the data. Figure 14 shows the

fitted and mapped variograms. Figure 15 shows the fitting of the mapped variogram and all 4 variograms plotted together. It is interesting that the variability is lower when calculated from the porosity values.

Figures 15-17 are from dallas.dat, some lead data. Figure 16 shows the variograms calculated from the data. Figure 17 shows the fitted variogram for the calculated normal scores variogram used to map the other variogram. Figure 18 shows the fitting of the mapped variogram and all four variograms plotted together. In this case, the mapped variogram demonstrates greater variography than the variogram calculated from the lead data.

Conclusions

The proposed approach is simple and yet practically very important. Avoiding bias in presence of sparse data is increasingly important as managers and stakeholders expect increasing accuracy from numerical models.

References

- Deutsch, C.V. and Journel, A.G., 1998. *GSLIB: Geostatistical Software Library: and User's Guide*. Oxford University Press, New York, 2nd Ed.

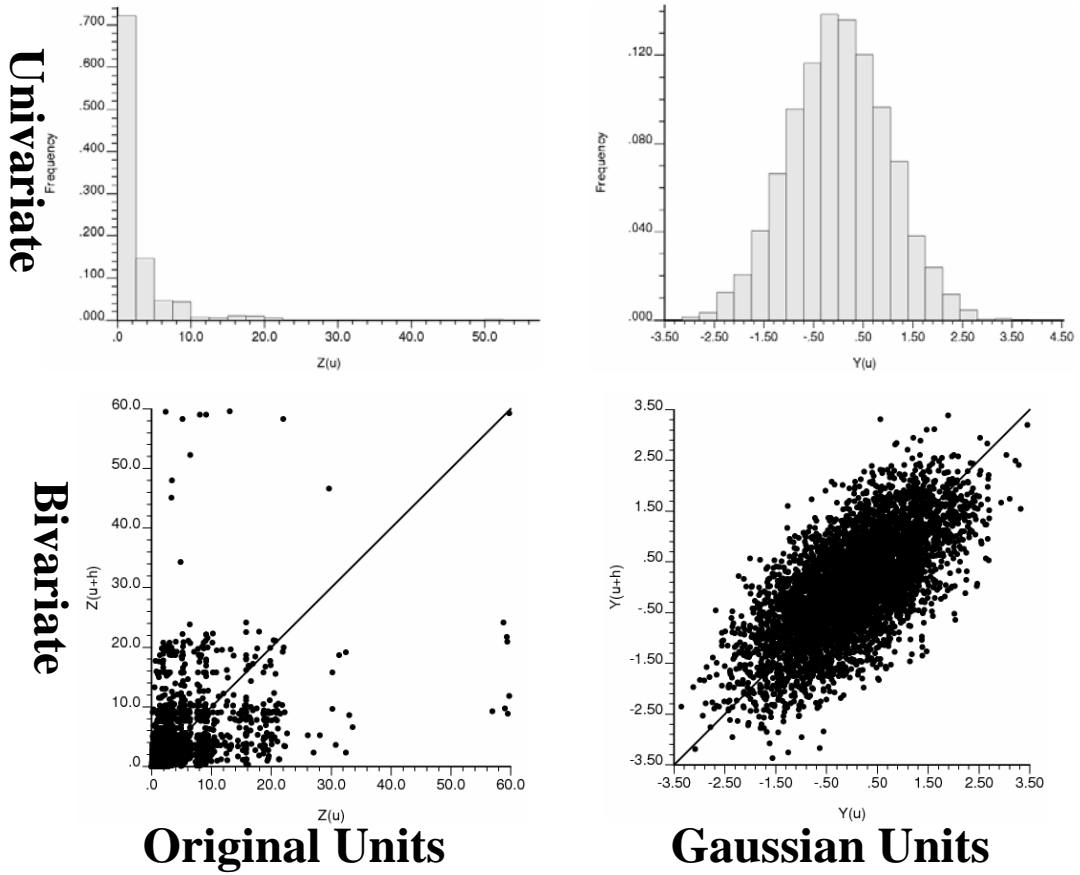


Figure 1: Motivation for calculating variograms with normal scores. The poor correlation of the original units gives an unstable variogram.

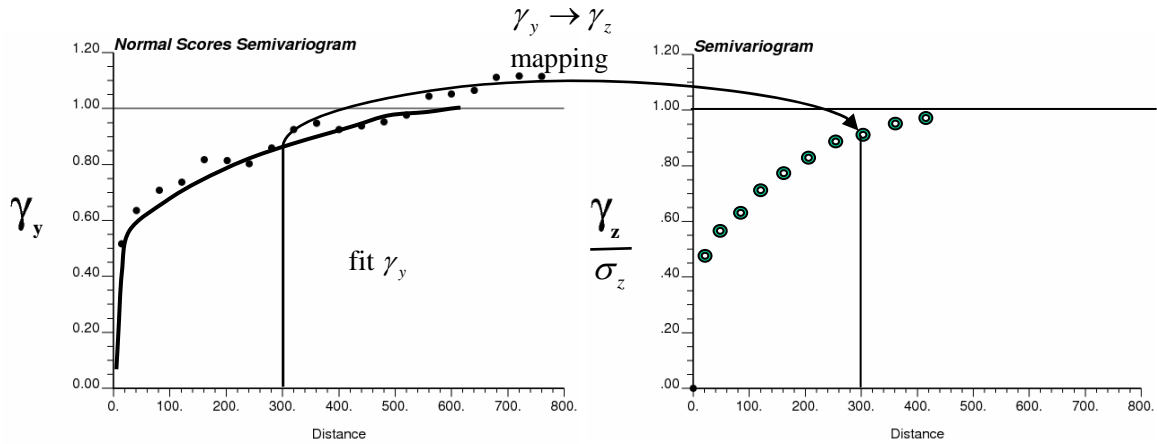


Figure 2: The mapping process.

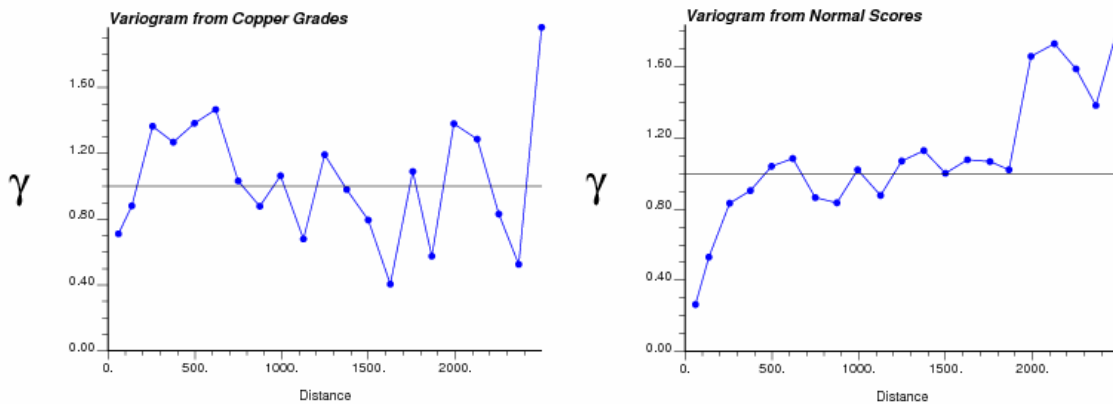


Figure 3: The left hand variogram calculated directly from the copper grades is unstable and difficult to model. This is characteristic of most variograms calculated from untransformed variables. The right hand variogram is calculated from the normal score transformed copper grades. It is much more stable and easy to model.

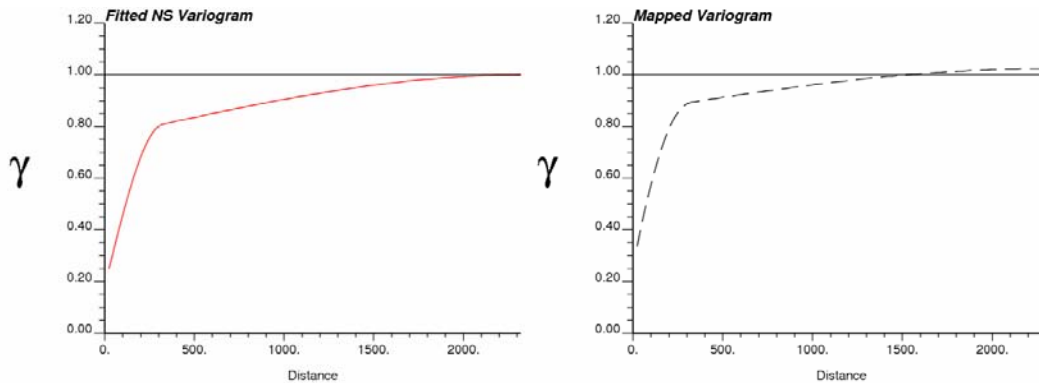


Figure 4: The left hand variogram has been fitted to the previously calculated normal score variogram (Figure 3b). It was then used by the new program, vmodelyz, to map the right hand variogram.

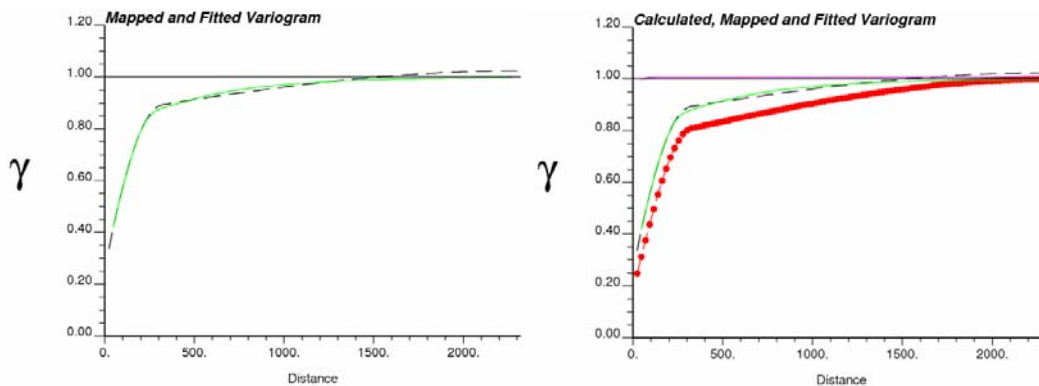


Figure 5: The mapped variogram is then fitted to be used for kriging purposes. The right hand plot shows all four variograms together. The variogram fitted from the variogram in Figure 3a is right along the sill. The mapped and fitted variograms overlay each other. The variogram fit from Figure 3b is the one with the lowest nugget effect and longest range.

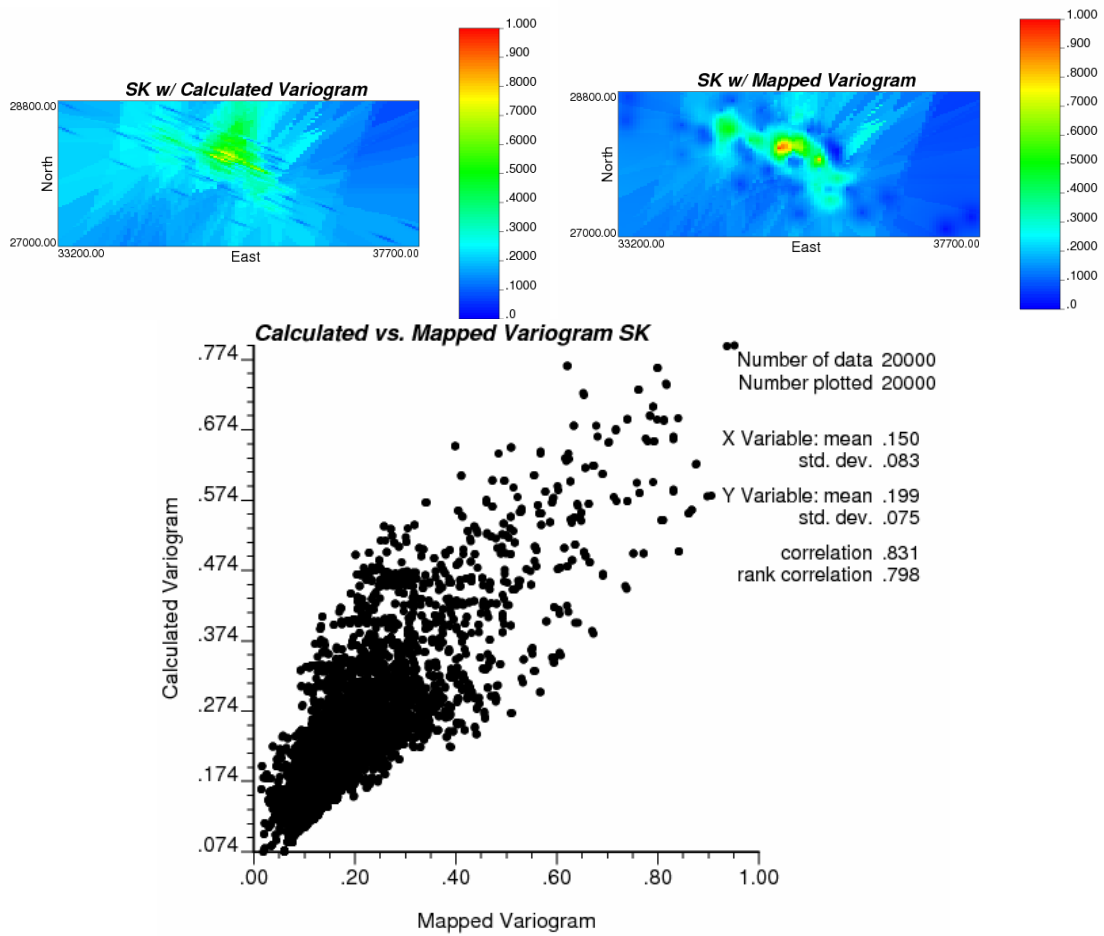


Figure 6: Simple Kriging using the variogram calculated from the copper grades, using the variogram mapped with the modified software, and a comparison of the two.

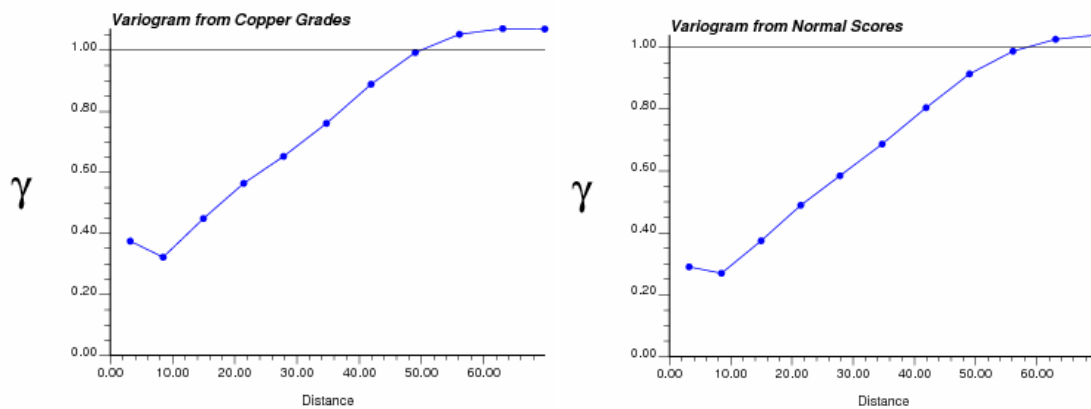


Figure 7: The variograms calculated from the original grades and the normal scores respectively.

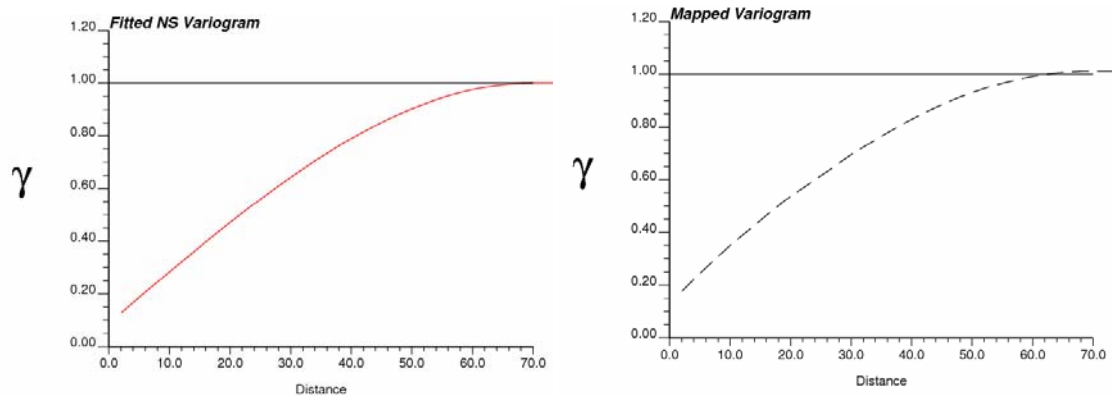


Figure 8: The variogram fitted to the one in Figure 7b and the one mapped from it using the new program.

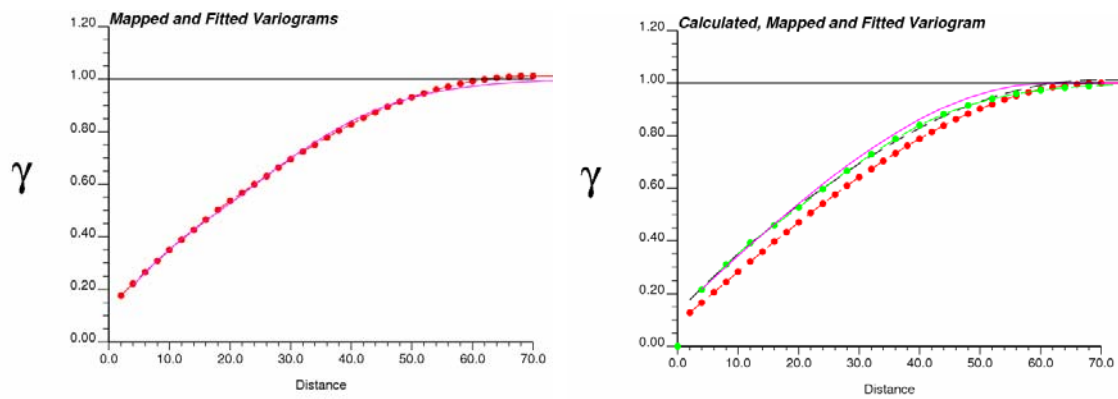


Figure 9: The mapped variogram along with the one fitted to it. All four variograms shown together.

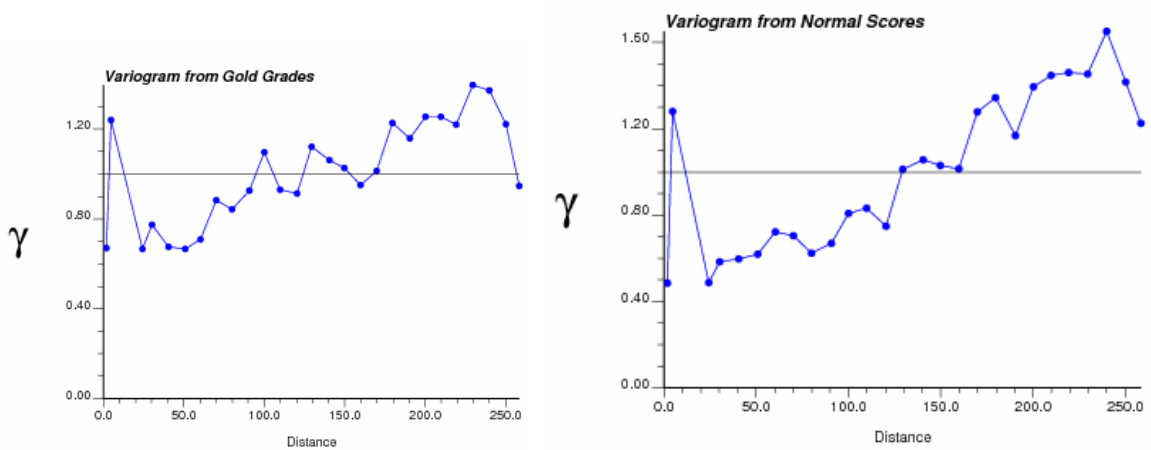


Figure 10: The variograms calculated from the gold grades and the normal scores respectively.

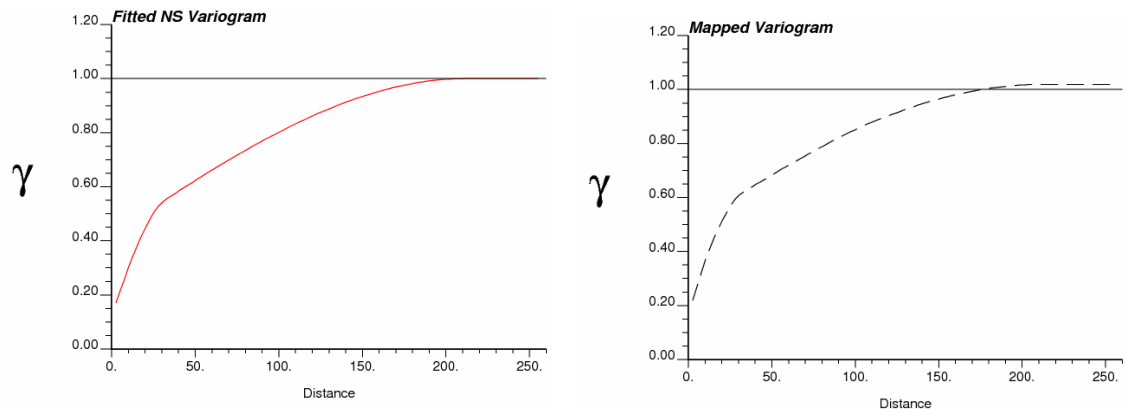


Figure 11: The variogram fitted to the one shown in Figure 10b and then variogram mapped from the first.

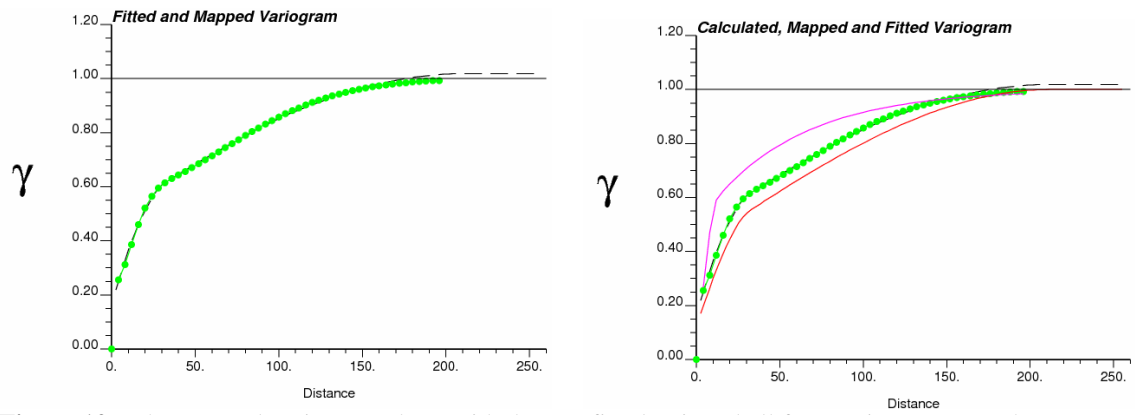


Figure 12: The mapped variogram along with the one fitted to it and all four variograms together.

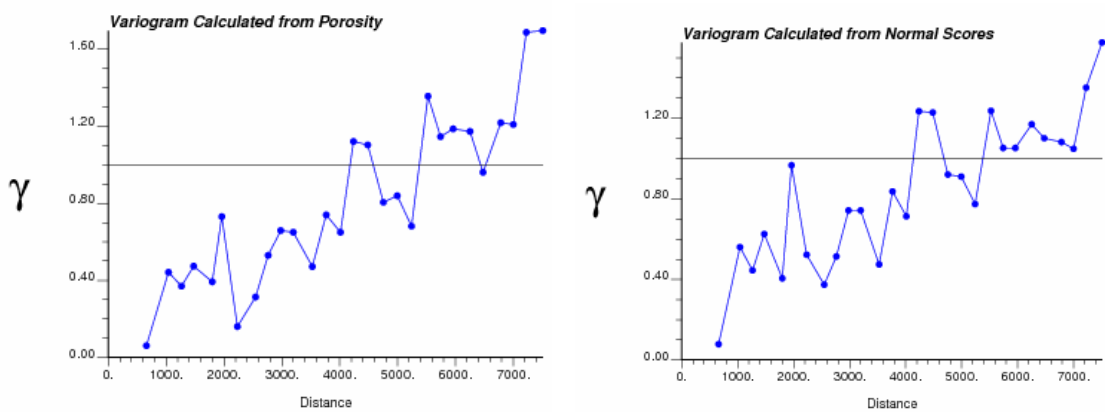


Figure 13: Variograms calculated from porosity and normal scores.

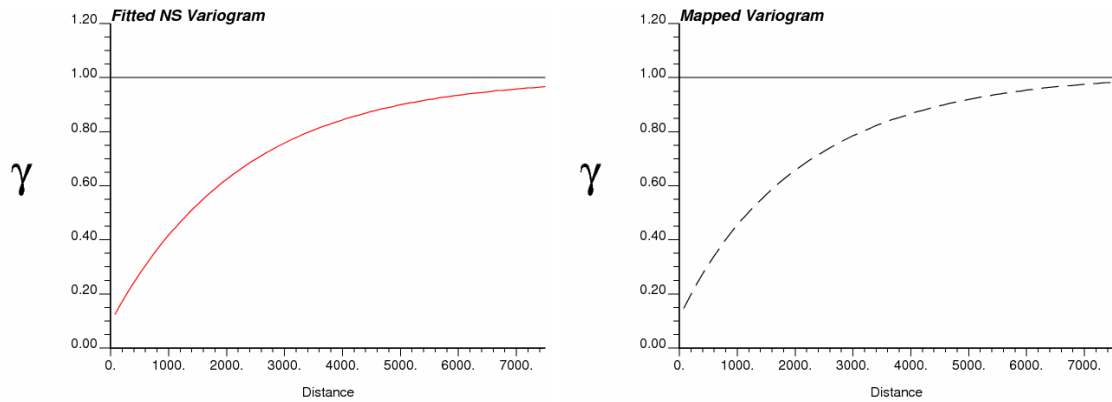


Figure 14: Variogram fitted to Figure 13b along with the variogram it was used to map.

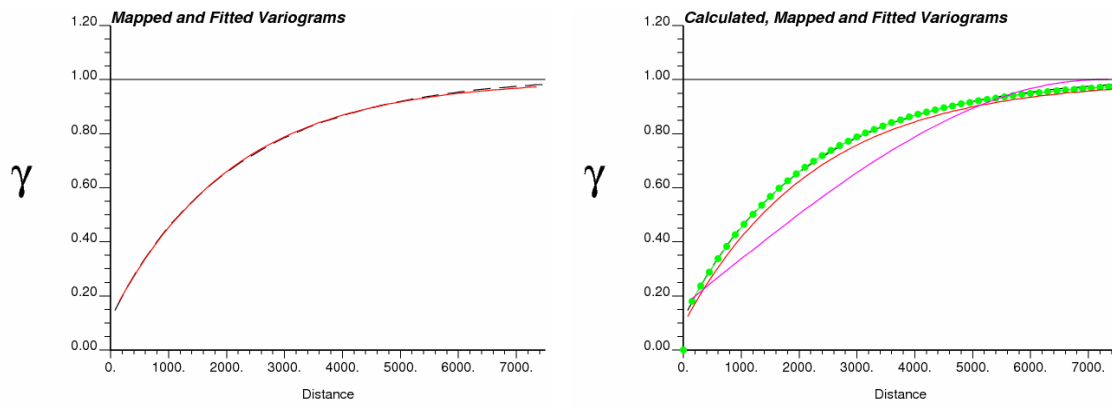


Figure 15: The mapped variogram with its fitted component and all four variograms plotted together.

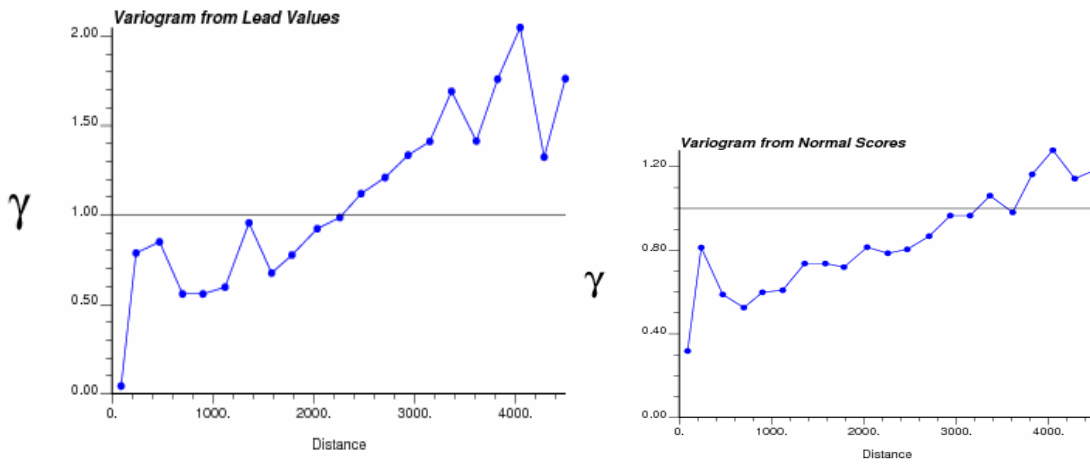


Figure 16: Variograms calculated from the data; lead values and normal scores.

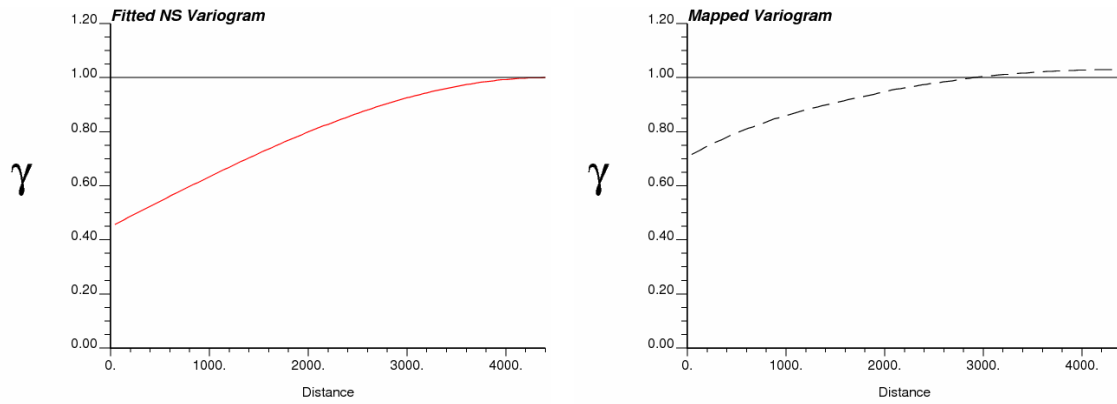


Figure 17: The variogram fitted from the previous normal scores and the variogram mapped from it.

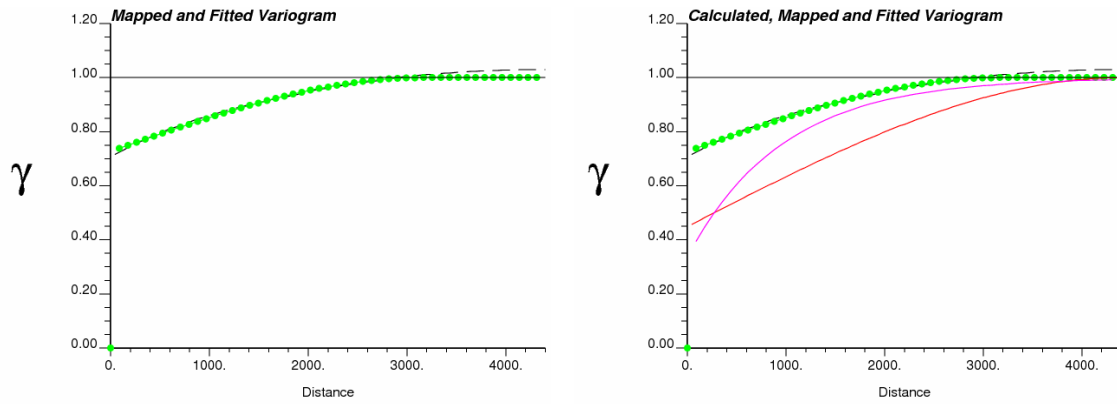


Figure 18: The mapped variogram with its fitting and all four variograms shown together.