Paper 310, CCG Annual Report 14, 2012 (© 2012)

Standardized Pairwise Relative Variogram as a Robust Estimator of Spatial Structure

Mahshid Babakhani and Clayton V. Deutsch

A requirement in geostatistical modeling is to find an appropriate and stable variogram. The traditional variogram is sensitive to outliers and sparse data. A number of robust alternatives have been proposed. Although the correlogram, indicator variogram and normal scores variograms sometimes work, the pairwise relative variogram is a remarkably robust estimator of the true underlying spatial structure. The variogram is adjusted by the squared mean of each data pair. A challenge in the interpretation of the pairwise relative variogram is that the sill of the pairwise relative variogram has not been well documented. The goal of this paper is to understand the sill of the pairwise relative variogram in order to standardize it for improved interpretation.

Introduction

Knowing the sill of the variogram is important for interpretation. The sill of the variogram is the variance, which is the variogram value that corresponds to zero correlation (Gringarten & Deutsch, 2001). It helps in understanding features such as trends, cyclicity, geometric anisotropy and zonal anisotropy. The sill of the pairwise relative variogram depends on the mean, variance and shape of the data distribution and is not easily predictable. A pairwise relative variogram is calculated using a method similar to the traditional variogram. The difference between them is the denominator, which serves to reduce the influence of outliers and clustered data (Isaaks, 1989). The definition of the pairwise relative variogram is as follows:

$$\gamma_{PR}(h) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{z(\mathbf{u})-z(\mathbf{u+h})}{\left(\frac{z(\mathbf{u})+z(\mathbf{u+h})}{2}\right)}\right]^2 \tag{1}$$

Based on experience, the pairwise relative variogram is more stable than the traditional variogram. However, the sill of this variogram estimator is unknown. In this paper, a meaningful relationship for the sill of the pairwise relative variogram is established. It will be shown that the sill of the pairwise relative variogram for standard (μ=1, σ=1) lognormal data is a constant; therefore, a robust approach to various calculation and interpretation is to 1) transform to a standard lognormal distribution, 2) compute a standard pairwise relative variogram, and 3) interpret with established pairwise relative variogram. This approach will be documented with some real data.

Dependence of the sill of the pairwise relative variogram

The mean, variance and shape of the data distribution are the three factors that the sill of the pairwise relative variogram depends on. To see these dependence more, in a lognormal distribution, with different values of mean and variance, the sill of the pairwise relative variogram is computed. An interesting result is the dependence of the sill of the pairwise variogram on the ratio of the square root of variance to the mean not the mean or variance individually. This ratio is defined as the coefficient of variation (CV). As a result, the sill of the pairwise relative variogram in lognormal distribution depends on the shape of data distribution. In Figure 1 a pairwise relative variogram of a standard lognormal distribution with the CV of one is shown. It can be seen that the approximate value of the sill is 0.44. Figure 2 illustrates the relationship between the sill of the pairwise relative variogram and different values of CV for a lognormal distribution.

Other distributions were considered to help understand more about this relationship between the sill of the pairwise relative variogram and CV. The uniform distribution is completely different in shape with the lognormal distribution; however, the sill of the pairwise relative variogram for uniform distribution has close behavior to the lognormal distribution (see Figure 3).

From the definition of the pairwise relative variogram, outliers do not have much effect on this kind of variogram. This is shown with a corrupted lognormal distribution. First it is taken the 0.01% of the values and set the Z value to 1000, then the mean and variance and CV are recomputed, the same thing is done with the 0.02% of data. Adding more outliers makes the CV increase and it changes the shape of the lognormal distribution. That is the reason why with the different values of CV the sill of the lognormal distribution remains same. Figure 4 illustrates that the outliers have no effect on the sill of the pairwise relative variogram.

Positively and negatively skewed distributions are also examined. According to Figure 5, it seems that high values have more effect than the low values.

Standardization Procedure

An important goal in a geostatistical study is to figure out a stable and standard variogram for an appropriate interpretation. The pairwise relative variogram is a very stable variogram estimator. The important thing is to find a way to standardize the sill of the pairwise relative variogram. Determining a general relationship for the sill given data of arbitrary shape, mean, variance and outliers would be intractable. There is another option that is simpler and more reliable. The data are transformed into a *standard* lognormal distribution with mean and variance of one. The sill of the pairwise variogram for this standard lognormal distribution is 0.44, so the pairwise relative variogram is standardized by dividing it by the constant 0.44:

$$\gamma_{PR,S} = \frac{\gamma_{PR}}{0.44} \tag{2}$$

This standard pairwise variogram is easy to compute and helps the researchers for a more clear interpretation.

Implementation

The progress of calculating the standard pairwise relative variogram is simple. The lognormal algorithm is implemented in the FORTRAN program `lognorm_new`, which was developed from the `lognorm` program (Deutsch, 1998). The parameter file of `lognorm_new` is shown in Table 1. In this program the mean and standard deviation can be chosen according to what measure of CV is needed for the variable. The output file includes the whole data from the original data file plus an additional column of the transformed data.

```
                    Parameters for lognormal new
                    ****************************

START OF PARAMETERS:
../data/cluster.dat        - file with data
3                          - columns for variable
1  1                       - mean and standard deviation you want
lognormal.out              - file for output
```

Table 1: Parameters file for calculating lognormal distribution

The standardization part of the pairwise relative variogram is implemented in the FORTRAN program `gamv2004_pair`, which was developed from `gamv2004` program (C.V. Deutsch 2004, D.F. Machuca-Mory 2007) .In the original program the number 6 is defined for calculating the pairwise relative variogram. In the new program (-6) is the number for calculating the standardized pairwise relative variogram. Therefore, to write the progress in a script, first the executable *nscore.exe* transforms the data into normal score then the executable *lognorm_new.exe* transforms those data into the lognormal distribution with arbitrary mean and standard deviation that they are chosen to be one. Finally executable *gamv2004_new.exe* calculates the standard pairwise relative variogram.

A 3-D data set of oil sands data is used to calculate the standardized pairwise relative variogram. This data set consists of 281 drillhole of 5808 data. The Bitumen variable is used in this study. Figure 6 shows the location map of the original bitumen and histogram of the lognormal bitumen. Figure 7 demonstrates the difference between the semivariogram calculated from original bitumen variable and pairwise relative variogram from lognormal of that variable. It can be seen that the semivariogram from the original variable of bitumen is so unstable. By using the `gamv2004_pair` the standardized pairwise relative variogram is calculated and it is shown in Figure 8 . The two variograms in this figure illustrate the standardized pairwise relative variogram in both vertical and horizontal directions.

Another example is calculating the pairwise relative variogram of porosity in the dataset of a small data set, 2DWellData.dat contains of 62 data. The variable in this study is porosity. Figure 9 is showing the location map of the porosity data and the histogram of the standard lognormal of porosity. Figure 10 shows two variograms one of them is the semivariogram from original porosity and the other one is the standardized pairwise relative variogram from lognormal of porosity.

Conclusions

The proposed technique for calculating the standard pairwise relative variogram is simple, as it does not require the data to follow any particular distribution. The contribution of this research is the reliance on the transformation of the data into standard lognormal distribution and the calculation of the standard pairwise relative variogram. This technique has some shortcomings: 1) Mixing the population is one of the pitfalls in this approach. It means that in a population with two domains and two different distributions, this technique considers only one population and transforms all the data into a lognormal distribution. 2) Another pitfall is the tolerance in the calculation of the pairwise relative variogram.

References

Deutsch, C. V. and A. G. Journel. (1998). *GSLIB: Geostatistical Software Library and User's Guide. Applied Geostatistics Series.* New York: Oxforf University press, New York.

Gringarten, E., & Deutsch, C.V. (2001). Teacher's Aide Variogram Interpretation and Modeling. *Mathematical Geology*.

Isaaks, E. H. and Srivastava, R.M. (1989). *An introduction to applied geostatistics.* New York: Oxford University Press.

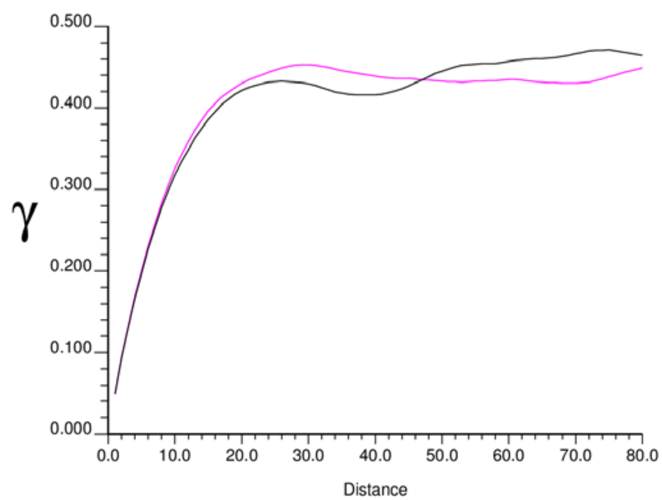

Figure 1: Two directional pairwise relative variogram for a standard lognormal distribution. The violet line is in the x direction, the black line is in the y direction.

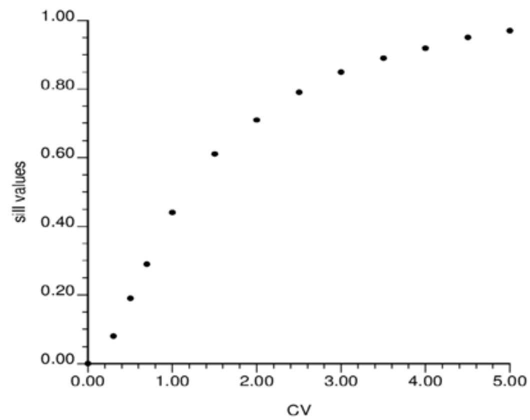

Figure 2: The relation between CV and the sill of the pairwise relative variogram for lognormal distribution

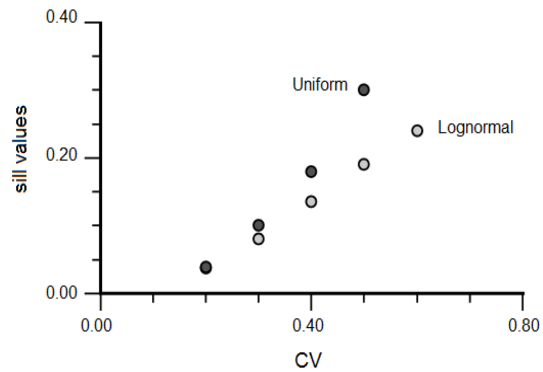

Figure 3: The light grey dots are the sill of the pairwise variogram for the lognormal distribution and the dark ones are for the uniform distribution

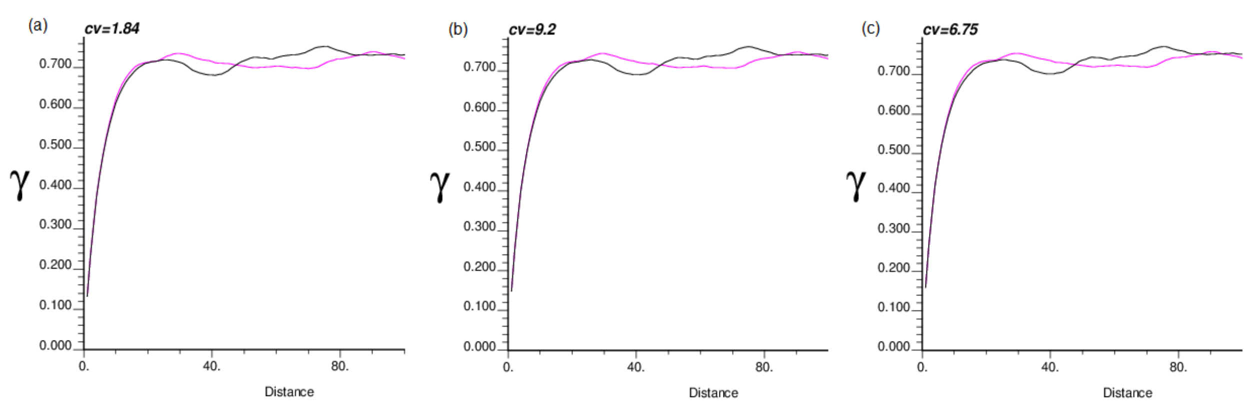

Figure 4: figure 4(a) is the lognormal distribution with the CV of 1.84; Figures 4(b) and 4(c) are with the outliers.

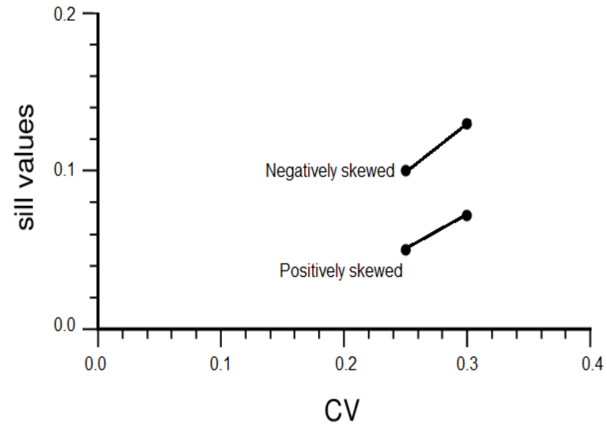

Figure 5: The sill of the negatively skewed and positively skewed lognormal distribution.

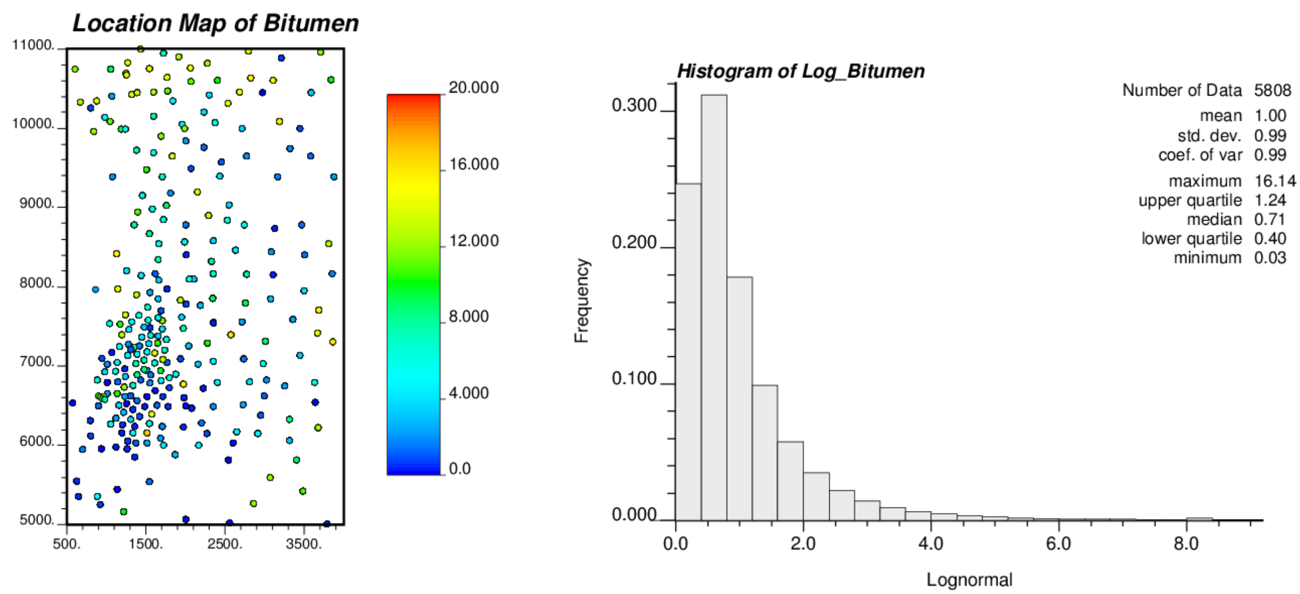

Figure 6: location map and histogram of bitumen data

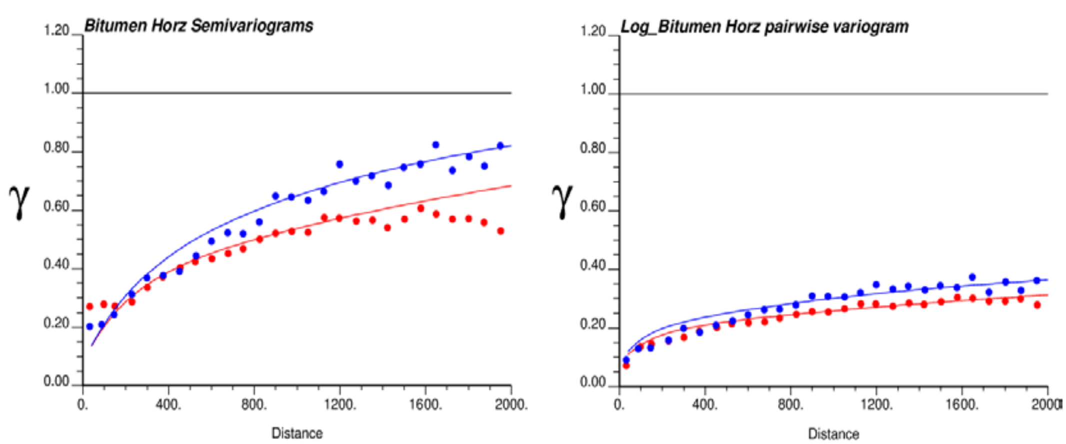

Figure 7: the variogram is semivariogram from the original variable and the right pairwise variogram is from lognormal of the bitumen.

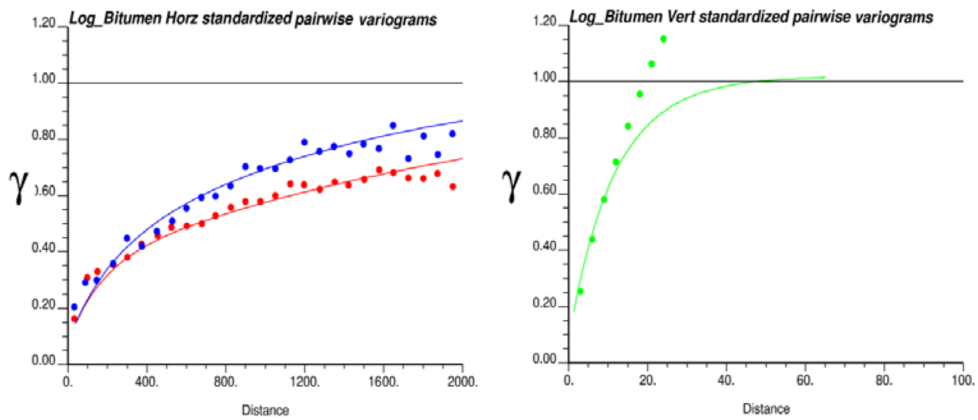

Figure 8: Standard pairwise variogram from lognormal bitumen in both vertical and horizontal directions.

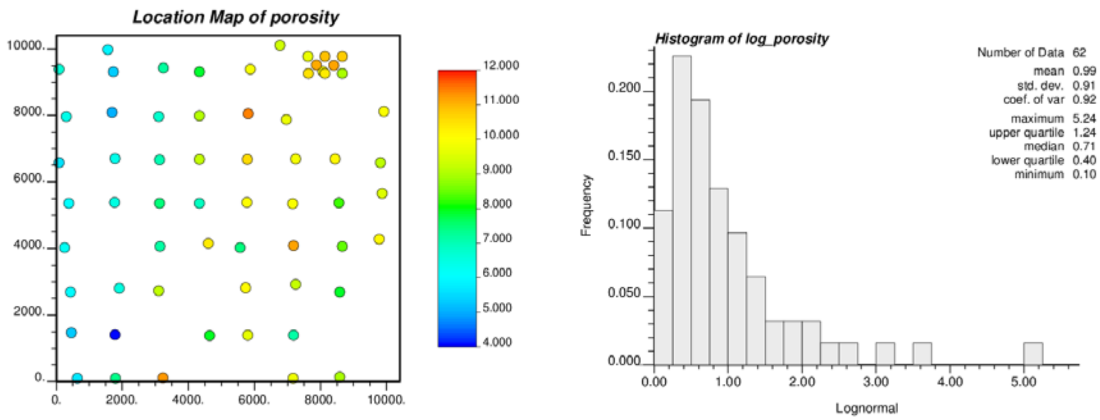

Figure 9: The location map of porosity data and the histogram of the standard lognormal of the porosity data

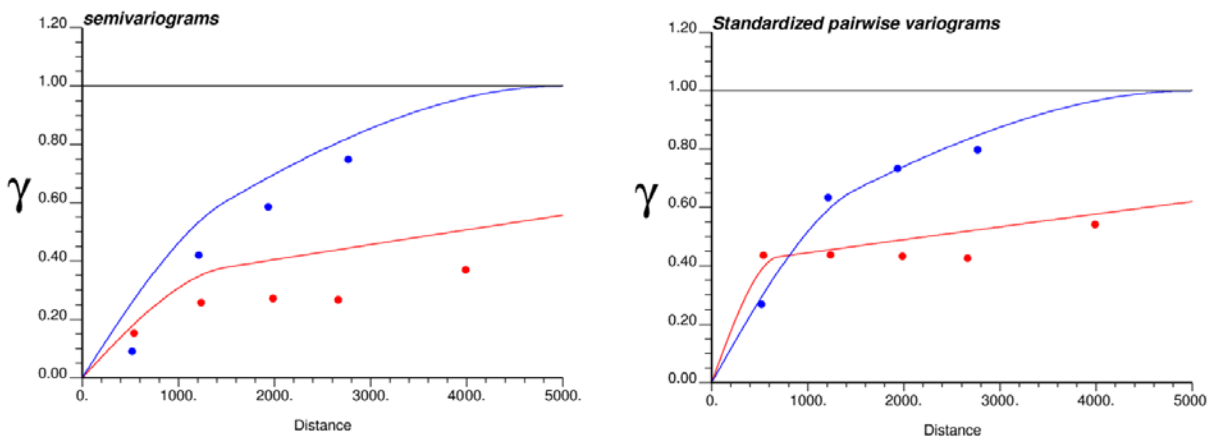

Figure 10: The left semivariogram is from original porosity and the right pairwise relative variogram is from lognormal of that variable