A Short Note on Why Geostatisticians Use the Variogram

Jason A. McLennan and Clayton V. Deutsch

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

Covariances and correlations were used before the variogram in spatial statistics literature. Kriging and simulation proceeds with the covariance counterpart to variogram models. Nevertheless, the pioneers of Geostatistics preferred the variogram. This short note provides a brief explanation in terms of robustness with respect to trends.

Random Function Formalism

Consider the random function (RF) $Z(\mathbf{u})$ that consists of the set of random variables (RVs) $\{Z(\mathbf{u}_l)$, for all $\mathbf{u}_l \in \mathbf{D}\}$. The RF is commonly assumed multivariate Gaussian after a univariate Gaussian transformation of the $Z(\mathbf{u})$ variable. The parameters of the multivariate distribution are inferred from the available sample data $z(\mathbf{u}_s)$, $s = 1, \ldots, n$. At an unsampled location $\mathbf{u}_0$, the conditional distribution is normal with mean and variance equal to the simple kriging estimate $z^*_{\text{SK}}(\mathbf{u}_0)$ and variance $\sigma^2_{\text{SK}}(\mathbf{u}_0)$. The simple kriging estimate is a weighted linear combination of the surrounding n sample data:

$$z^*_{\text{SK}}(\mathbf{u}_0) = \sum_{s=1}^{n} \lambda_{\text{SK}}(\mathbf{u}_s) \cdot z(\mathbf{u}_s) \tag{1}$$

The weights $\lambda(\mathbf{u}_s)$ that minimize the expected error variance of the estimate are given by the following system of equations:

$$\sum_{s'=1}^{n} \lambda_{\text{SK}}(\mathbf{u}_{s'}) \cdot C_Z(\mathbf{u}_s - \mathbf{u}_{s'}) = C_Z(\mathbf{u}_s - \mathbf{u}_0) \quad s = 1,\ldots,n \tag{2}$$

And the minimum error variance or kriging variance is:

$$\sigma^2_{\text{SK}}(\mathbf{u}_0) = C(\mathbf{0}) - \sum_{s=1}^{n} \lambda_{\text{SK}}(\mathbf{u}_s) \cdot C_Z(\mathbf{u}_s - \mathbf{u}_0) \tag{3}$$

The covariances $C_Z(\mathbf{u}_{s'} - \mathbf{u}_s)$ and $C_Z(\mathbf{u}_s - \mathbf{u}_0)$, or just $C_Z(\mathbf{h})$ where $\mathbf{h}$ is a lag vector, are then required to solve the kriging equations and establish the local $\mathbf{u}_0$ distribution of uncertainty from which simulation is performed. In order to calculate $C_Z(\mathbf{h})$, it is actually the variogram function $\gamma_Z(\mathbf{h})$ that is calculated, interpreted, and modeled. $\gamma_Z(\mathbf{h})$ is calculated from all $N(\mathbf{h})$ pairs of scattered sample data $(z(\mathbf{u}_s), z(\mathbf{u}_s + \mathbf{h}))$ approximately separated by the lag $\mathbf{h}$:

$$\gamma_Z(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{s=1}^{N(\mathbf{h})} \left(z(\mathbf{u}_s) - z(\mathbf{u}_s + \mathbf{h}) \right)^2 \tag{4}$$

An assumption of second-order stationarity then allows the calculation of $C_Z(\mathbf{h})$ through:

$$\gamma_Z(\mathbf{h}) = C(\mathbf{0}) - C_Z(\mathbf{h}) \tag{5}$$

The kriging equations can then be solved and simulation can follow.

Why the Variogram?

Like the variogram, $C_Z(\mathbf{h})$ is calculated from all pairs of scattered sample data ($z(\mathbf{u}_s)$, $z(\mathbf{u}_s + \mathbf{h})$) approximately separated by the lag vector $\mathbf{h}$:

$$C_Z(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{s=1}^{N(\mathbf{h})} \left(z(\mathbf{u}_s) \cdot z(\mathbf{u}_s + \mathbf{h}) \right) - m_{-\mathbf{h}} m_{+\mathbf{h}} \tag{6}$$

where the $m_{-\mathbf{h}}$ and $m_{-\mathbf{h}}$ are the means of the $N(\mathbf{h})$ $z(\mathbf{u}_s)$ tail and $N(\mathbf{h})$ $z(\mathbf{u}_s + \mathbf{h})$ head values, respectively. Notice the $C_Z(\mathbf{h})$ covariance function involves and is sensitive to the mean while the variogram acts as a filter on the mean m. This is the most important reason why the variogram (4) is used over the covariance (6) to quantify and model spatial correlation for estimation [1]. This property of the variogram is desirable in geostatistical applications where a decision of second-order stationarity is implied for an inherently non-stationary geological RF $Z(\mathbf{u})$.

A small 2D example is set up to see how the variogram effectively filters the mean from a non-stationary field. Figure 1 shows the construction of a non-stationary porosity RF $Z(\mathbf{u})$ over a 100 x 100m field. The porosity residuals (left) are unconditionally simulated using a spherical variogram with zero nugget effect and 20m isotropic range. The locally varying porosity model (middle) is a linear increasing function of the Y coordinate vector. The porosity variable is constructed by adding the mean and residual models. Some sample data are then extracted from the resulting porosity field at a 10m spacing for the subsequent calculation of spatial correlation.

Figure 2 shows the variogram and covariance functions using relations (4) and (6) in both the X (yellow) and Y (red) direction. Notice for relatively small $\mathbf{h}$ up to approximately one-third of the range, the variogram in both the X and Y direction is virtually the same. This shows that $\gamma_Z(\mathbf{h})$ is a filter on the short scale locally varying mean. In contrast, the covariance is significantly different at both the short and longer scales showing its sensitivity to the locally varying drift. The Y direction in particular shows more spatial continuity than the X direction.

Conclusion

The variogram is considered more robust in the presence of trends and departures from stationarity since it filters non-stationary or locally varying means. The covariance on the other hand is sensitive to the mean. The relatively robust virtue of the variogram is the reason for its popularity.

References

Isaaks, E.H. and Srivastava, R. M. *Spatial Continuity Measures for Probabilistic and Deterministic Geostatistics*. Mathematical Geology, 20 (4), 1988, pp. 313-341.

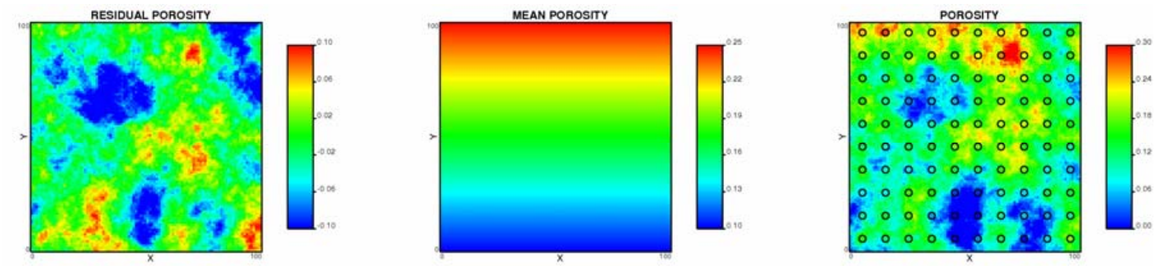

Figure 1: The construction of a non-stationary porosity random function showing the residuals (left), locally varying mean (middle), and resulting porosity variable (right). The dark circles represent sample data locations.

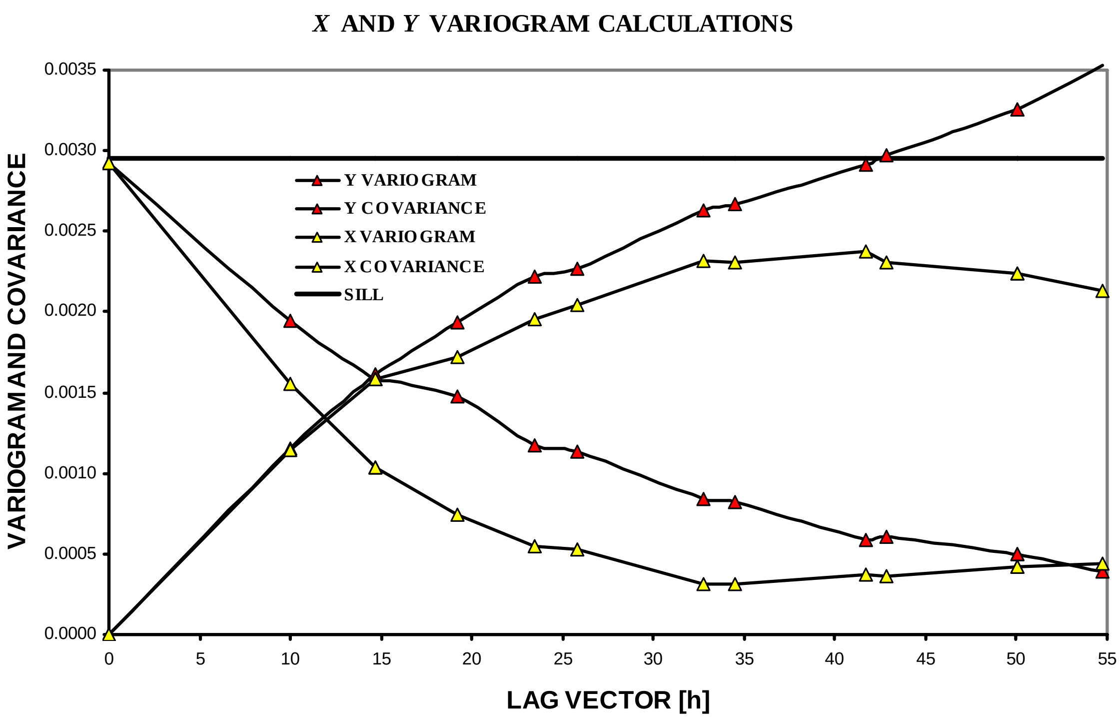

Figure 2: The *X* and *Y* direction experimental variograms for the porosity sample data set constructed in Figure 1.