Precision Agric (2008) 9:173–175
DOI 10.1007/s11119-008-9056-2

SHORT DISCUSSION

Variogram or semivariogram? Understanding the variances in a variogram

Martin Bachmaier · Matthias Backes

Published online: 24 February 2008

The theoretical variogram and the confusion in the literature

The definition of the theoretical variogram, γ, is based on regionalized random variables $Z(\vec{x})$ and $Z(\vec{x}+\vec{h})$ where $\vec{x}$ and $\vec{x}+\vec{h}$ represent the spatial positions separated by a vector $\vec{h}$:

$$\gamma(\vec{h}) = \frac{1}{2}\mathrm{E}\left[[Z(\vec{x}+\vec{h}) - Z(\vec{x})]^2\right] = \frac{1}{2}\mathrm{Var}\left[Z(\vec{x}+\vec{h}) - Z(\vec{x})\right]. \tag{1}$$

The $Z(\vec{x})$ and $Z(\vec{x}+\vec{h})$ denote random variables. According to the intrinsic hypothesis, $\gamma(\vec{h})$ is assumed to depend only on the separation vector, the lag $\vec{h}$, but not on the location $\vec{x}$. Further, the increments $Z(\vec{x}+\vec{h}) - Z(\vec{x})$ are assumed to have no drift: $\mathrm{E}[Z(\vec{x}+\vec{h}) - Z(\vec{x})] = 0$ for all $\vec{h}$ and all $\vec{x}$; otherwise the last identity in Eq. 1 would not hold.

Concerning the terminology, there is great confusion in the geostatistical literature. Some authors call the function γ a 'variogram' (Wackernagel 2003; Worboys 1995; Gneiting et al. 2001), several authors call it a 'semivariogram' (Journel and Huijbregts 1978; Cressie 1991; Goovaerts 1997; Burrough and McDonnell 1998; Olea 1999; Stein 1999; Gringarten and Deutsch 2001), stating that a semivariogram is half a variogram, and others use the terms variogram and semivariogram synonymously (Isaaks and Srivastava 1989; Webster and Oliver 2007). To explain what is depicted in a variogram, authors of geostatistical books and articles often take refuge in phrases such as "spatial variability" or terms like "semivariogram value" or "semivariance" without saying of what. Evidently there is great uncertainty with regard to terminology and the interpretation of variograms.

The confusion concerning the prefix "semi" has arisen because Matheron (1965) in his seminal thesis had in mind the variance of differences, $\mathrm{Var}[Z(\vec{x}+\vec{h}) - Z(\vec{x})]$, but the

M. Bachmaier (✉)
Department of Life Science Engineering, Crop Production Engineering,
Technische Universität München, Am Staudengarten 2,
85354 Freising-Weihenstephan, Germany
e-mail: bachmai@wzw.tum.de

M. Backes
Institut für Kartographie und Geoinformation, Rheinische Friedrich-Wilhelms-Universität Bonn,
Meckenheimer Allee 172, 53115 Bonn, Germany
e-mail: matthiasbackes@web.de

quantity we want in practice is the half ("semi") of this, because this gives the "variance per point when the points are considered in pairs" (Webster and Oliver 2007, p. 54; cf. also the expression in Eq. 3). Thus, $\gamma(\vec{h})$ can be interpreted as the variance of the variable (e.g. of the yield data) at the given separation vector $\vec{h}$, which means that we consider only pairs that are spatially separated by the lag $\vec{h}$. It should not be called a semivariance since this term originates from the variance of the differences, which is not the actual quantity of interest. And if it were, one should not compute the half of it, but the whole variance. No one says that the semiheight of his or her body is 86 cm.

Understanding the empirical variances in a variogram

The empirical variance of measured values z_i can be computed in two different ways:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_i - \bar{z})^2 = \frac{1}{2}\cdot\frac{1}{n(n-1)}\sum_{\text{all } i\neq j}(z_j - z_i)^2, \tag{2}$$

where $n(n-1)$ is the number of pairs in the sum. This latter could be halved, because it suffices to consider only all pairs with $i < j$ since $(z_i - z_j)^2 = (z_j - z_i)^2$.

The variances in an experimental variogram, $\hat{\gamma}(\vec{h})$, arise by restricting the latter expression in Eq. 2 to pairs of measured values, $z_i = z(\vec{x}_i)$ and $z_j = z(\vec{x}_i + \vec{h})$, that are separated by a spatial vector $\vec{h}$:

$$\hat{\gamma}(\vec{h}) = \frac{1}{2}\cdot\frac{1}{N(\vec{h})}\sum_{i=1}^{N(\vec{h})}\left[z(\vec{x}_i + \vec{h}) - z(\vec{x}_i)\right]^2, \tag{3}$$

where $z(\vec{x}_i + \vec{h})$ and $z(\vec{x}_i)$ are the measured values of Z at $N(\vec{h})$ pairs of comparisons, $\vec{x}_i + \vec{h}$ and $\vec{x}_i$, separated by the vector $\vec{h}$. Thus, $\hat{\gamma}(\vec{h})$ should simply be called the (empirical) variance of the measured values, e.g. of the yield data, at the given separation vector $\vec{h}$. When referring to isotropic variation, $\hat{\gamma}(h)$ denotes this variance at a given separating distance $h = ||\vec{h}||$. We do not need the term semivariance unless we want to cite references where it is used. But then it should be added that it is the semivariance of the difference of random variables or measured values.

Finally, Eqs. 2 and 3 also show that the empirical variance of all measured values, s^2, can be computed as a weighted mean of the variogram variances $\hat{\gamma}(\vec{h})$, where the weighting is according to the number of pairs, $N(\vec{h})$, in $\hat{\gamma}(\vec{h})$ (Bachmaier 2007, Eq. 23).

References

Bachmaier, M. (2007). Using a robust variogram to find an adequate butterfly neighborhood size for one-step yield mapping using robust fitting paraboloid cones. *Precision Agriculture, 8*, 75–93.

Burrough, P. A., & McDonnell, R. A. (1998). *Principles of geographical information systems. Spatial information systems and geostatistics*. Oxford, New York: Oxford University Press.

Cressie, N. A. C. (1991). *Statistics for spatial data*. New York: Wiley.

Gneiting, T., Sasvári, Z., & Schlather, M. (2001). Analogies and correspondences between variograms and covariance functions. *Advances in Applied Probability, 33*, 617–630.

Goovaerts, P. (1997). *Geostatistics for natural resource evaluation*. Oxford, New York: Oxford University Press.

Gringarten, E., & Deutsch, C. V. (2001). Teacher's aide—variogram interpretation and modeling. *Mathematical Geology, 33*, 507–534.

Isaaks, E. H., & Srivastava, R. M. (1989). *An introduction to applied geostatistics*. Oxford, New York: Oxford University Press.
Journel, A. G., & Huijbregts, C. J. (1978). *Mining geostatistics*. London: Academic Press.
Matheron, G. (1965). *Les Variables régionalisées et leur Estimation*. Paris: Masson.
Olea, R. A. (1999). *Geostatistics for engineers and earth scientists*. Boston: Kluwer Academic Publishers.
Stein, M. L. (1999). *Interpolation of spatial data—some theory for kriging*. New York: Springer Verlag.
Wackernagel, H. (2003). *Multivariate geostatistics*. Berlin, Heidelberg: Springer Verlag.
Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists*. New York: Wiley.
Worboys, M. F. (1995). *GIS—a computing perspective*. London: Taylor & Francis Ltd.