Variograms to Guide Spatial Sampling for Kriging

Margaret A. Oliver, Soil Research Centre, University of Reading, Reading, RG6 6DW United Kingdom. m.a.oliver@reading.ac.uk
Ruth Kerry, Brigham Young University, Provo, Utah, USA.

Abstract: Detailed information on soil to manage polluted or agricultural sites is often prohibitively expensive to obtain. To sample adequately, the approximate scale of spatial variation needs to be known. If soil data are available, variograms can be computed and used to determine the kriging errors for several grid intervals and an interval selected to meet a specific tolerable error. In the absence of prior knowledge, if soil properties appear related to ancillary data such as aerial photographs, elevation or apparent electrical conductivity (EC_a), individual or multivariate variograms of such data may indicate the scale of variation in the soil. If the scale of variation indicates too few data to compute a reliable variogram conventionally, it can be estimated by residual maximum likelihood or a standardized variogram from ancillary data can be used.

Keywords: variogram, residual maximum likelihood (REML), standardized variogram

1. Introduction

Soil properties can vary at markedly different spatial scales within sites of interest, such as fields. The variation comprises that over short distances of a few metres and over longer distances of tens or hundreds of metres. For most environmental and agricultural management, it is variation over tens or hundreds of metres that managers want to resolve and we can regard the short-range variation as ‘noise’ or a sampling effect. Many soil attributes have to be determined from samples taken in the field, therefore there is a need to predict accurately at places where there are no data. Kriging provides a sound basis for prediction leading to accurate digital mapping for managing soil attributes (Oliver, 2010). The accuracy of kriged and other interpolated predictions, however, depends on the quality of sample information to compute accurate variograms and availability of spatially dependent data from which to predict (Webster and Oliver, 2007). This means that sampling should be at an interval that is well within the correlation range of spatial variation. Therefore, it is essential that the spatial scales of variation in the properties of most importance for environmental and agricultural management are used to guide sampling.

Sampling on a grid is often used because it provides an even cover of values and minimizes the maximum estimation variance (or error) for a given grid interval and it is efficient for sample collection in the field. If variograms of soil properties from previous surveys exist for an area with a similar soil parent material, they can be used with the kriging equations to determine an optimal grid interval. If the scale of variation is large, the sampling intervals recommended by this method will also be large and there may be too few data from which to compute a reliable variogram by the usual method of moments estimator. Webster and Oliver (1992) showed that at least 100 data are required to compute a reliable variogram in this way from isotropic data. However, Kerry and Oliver (2007) have shown that a variogram estimated by residual maximum likelihood (REML) can provide more accurate predictions with fewer data than one estimated conventionally. For some soil properties, the variation might be evident in remote and proximally sensed imagery. Variograms computed from such ancillary data can be used to determine the approximate scale of spatial variation. A standardized variogram based on ancillary data or existing variograms of soil properties can also be used to krige spatially dependent (Kerry and Oliver, 2008). We illustrate these methods with a case study in England.

2. Materials and Methods

Matheron's method of moments (MoM) estimator to compute the variogram is given by

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2m(\mathbf{h})}\sum_{i=1}^{m(\mathbf{h})}\{z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})\}^2\,, \qquad (1)$$

where $z(\mathbf{x}_i)$ and $z(\mathbf{x}_i+\mathbf{h})$ are the actual values of Z at places $\mathbf{x}_i$ and $\mathbf{x}_i+\mathbf{h}$, and $m(\mathbf{h})$ is the number of paired comparisons at lag $\mathbf{h}$. The parameters of the model fitted to the experimental variogram can be used with the data for prediction at points or over blocks. Kriged predictions are a weighted average of the data, $z(\mathbf{x}_1)$, $z(\mathbf{x}_2)$, …, $z(\mathbf{x}_n)$, at the unknown point or block, B,

$$\hat{Z}(B) = \sum_{i=1}^{n}\lambda_i z(\mathbf{x}_i), \qquad (2)$$

where n usually represents the data points within the local neighbourhood and λ_i are the weights. To ensure that the estimate is unbiased the weights are made to sum to one. The estimation variance of $\hat{Z}(B)$ is

$$\mathrm{var}\left[\hat{Z}(B)\right] = \mathrm{E}\left[\left\{\hat{Z}(B) - Z(B)\right\}^2\right] = 2\sum_{i=1}^{n}\lambda_i\bar{\gamma}(\mathbf{x}_i, B) - \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\gamma(\mathbf{x}_i, \mathbf{x}_j) - \bar{\gamma}(B, B), \qquad (3)$$

where $\bar{\gamma}(\mathbf{x}_i, B)$ is the average semivariance between data point $\mathbf{x}_i$ and the target block B, and $\bar{\gamma}(B, B)$ is the average semivariance within B, the within block variance. The kriging error is the square root of this.

McBratney et al. (1981) showed how the variogram and kriging equations could be used to determine an optimal sampling interval for prediction by kriging before obtaining new data from a survey. The kriging weights, and also the kriging variances or errors, depend on the configuration of the sampling points in relation to the target point or block and on the variogram and not depend on the observed values at these points. Therefore if we have a variogram function from a previous survey we can determine the kriging errors for any grid size before sampling.

The experimental multivariate variogram **(**Bourgault and Marcotte, 1991) was computed from aerial photograph data by the standard formula adapted for the multivariate case:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2m(\mathbf{h})}\sum_{i=1}^{m(\mathbf{h})}\{\mathbf{z}(\mathbf{x}_i) - \mathbf{z}(\mathbf{x}_i + \mathbf{h})\}^{\mathrm{T}}\mathbf{M}\{\mathbf{z}(\mathbf{x}_i) - \mathbf{z}(\mathbf{x}_i + \mathbf{h})\}, \qquad (4)$$

where $\mathrm{z}(\mathbf{x}_i)$ and $\mathrm{z}(\mathbf{x}_i + \mathrm{h})$ are the vectors of observations at $\mathbf{x}_i$ and $\mathbf{x}_i + \mathbf{h}$, T is the transpose and $\mathbf{M}$ is a $p \times p$ positive-definite symmetric matrix defining the relations between the variables.

Pardo-Igúzquiza (1998) suggested that a reliable variogram could be computed from a 'few dozen' data by maximum likelihood or residual maximum likelihood (REML). Kerry and Oliver (2007) examined this idea further and suggested that 50 to 60 data might suffice (this paper also provides the detail on the theory of the method).

3. Results

The model parameters of a variogram computed from loss on ignition (LOI) data of a field in Wallingford, Oxfordshire, England were used to determine the kriging errors over blocks of various sizes and for a range of grid intervals. The kriging errors are plotted against grid spacing Fig. 1a and a suitable sampling interval would be 120 m for a tolerable error of 0.5%.

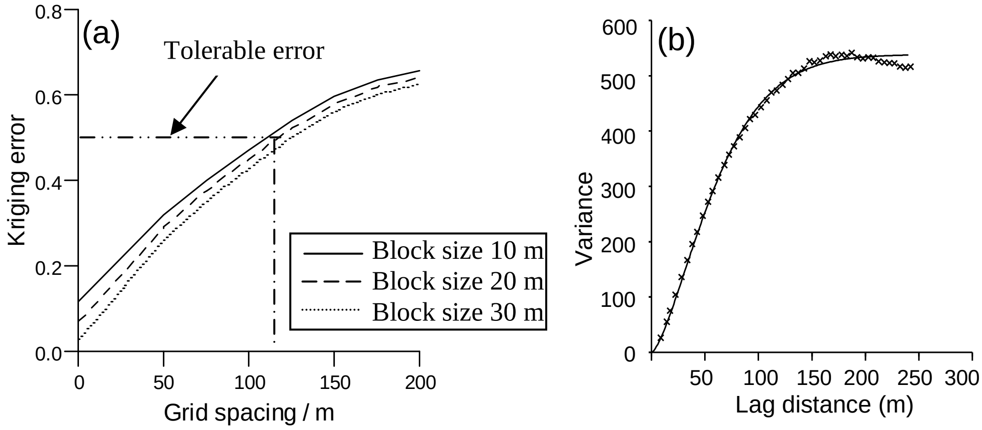

Figure 1: (a) Graph of kriging error against grid spacing for loss on ignition and (b) multivariate variogram of aerial photograph data at Wallingford, Oxfordshire, England.

The multivariate variogram computed from the red, green and blue wavebands of an aerial photograph of bare soil at Wallingford was fitted by a stable exponential function with an approximate range of 205 m (Fig. 1b). Based on less than half the variogram range, this suggested a sampling interval of about 90 m. Figure 2a–c shows kriged maps based on the conventional variograms with the original data on a 30-m grid, the suggested interval of 90 m and for 50 sites based on a 120-m grid with additional samples at 60 m, respectively for LOI at Wallingford. Figure 2d,e shows the kriged maps of LOI based on the 90-m grid with a variogram estimated by REML, and based on a 120-m grid with 15 additional targeted samples estimated by the variogram in Fig. 1b standardized to a sill of unity. Figure 1c–e shows that additional samples at a shorter interval, a variogram estimated by REML or a standardized variogram improve estimates from sparse data.

4. Concluding remarks

The results show the importance of knowing the scale of spatial variation, of having data at distances shorter than half the range of correlation and of alternative methods of estimating the variogram.

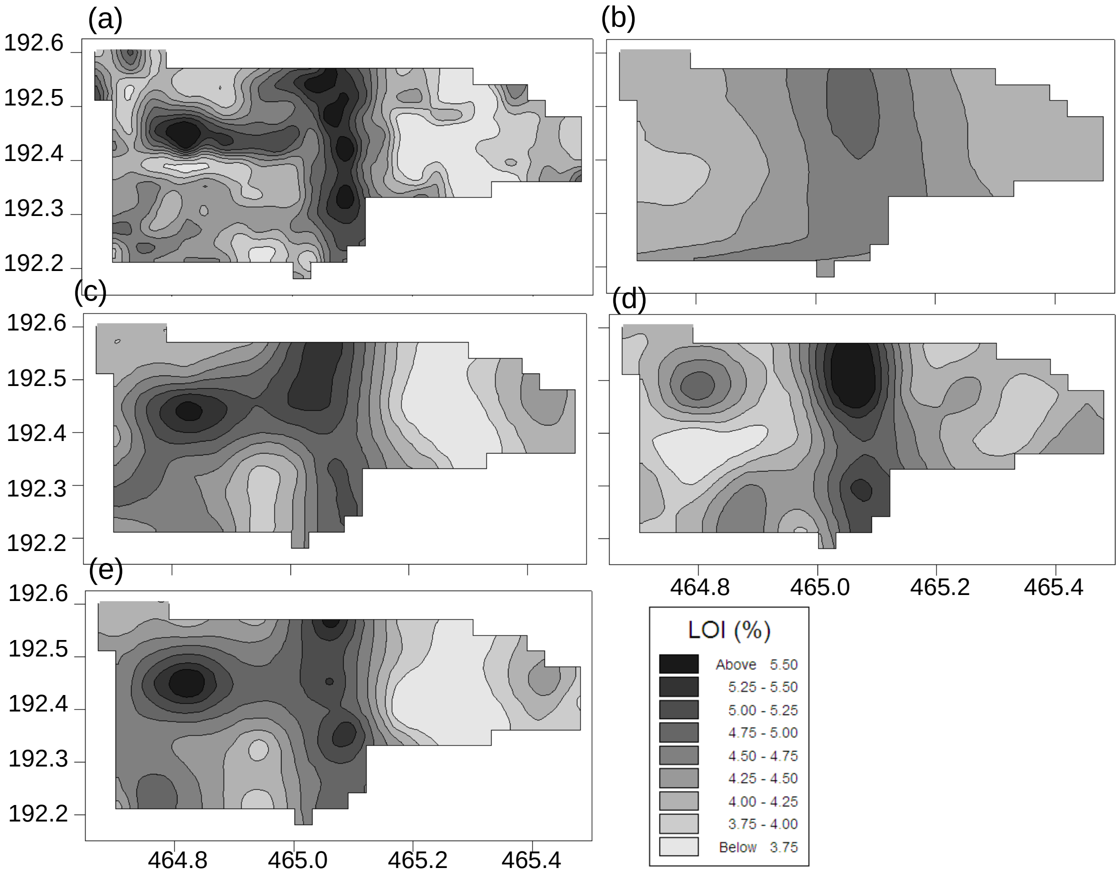

Figure 2: Kriged maps of LOI for Wallingford: with variograms estimated by MoM (a) 30-m grid (296 data), (b) 90-m grid (36 data), (c) 120-m grid + samples at 60-m (50 data); (e) 90-m grid with variogram estimated by REML and (d) 120-m grid + 15 targeted samples with standardized variogram.

References

Bourgault, G., Marcotte, D. (1991) Multivariable variogram and its application to the linear model of coregionalization. *Mathematical Geology*, 23, 899–928.

Kerry, R., Oliver, M. A. (2007) Sampling requirements for variograms of soil properties computed by the method of moments and residual maximum likelihood. *Geoderma*, 140, 383–396.

Kerry, R., Oliver, M. A. (2008) Determining nugget:sill ratios of standardized variograms from aerial photographs to krige sparse soil data. *Precision Agriculture*, 9, 33–56.

McBratney, A. B., Webster, R., Burgess, T. M. (1981) The design of optimal sampling schemes for local estimation and mapping of regionalized variables. I. Theory and method. *Computers & Geosciences*, 7, 331–334.

Oliver, M.A. (2010) An overview of geostatistics and precision agriculture, in: *Geostatistical Applications for Precision Agriculture*, M.A. Oliver (Ed.), Springer, 1–34.

Pardo-Igúzquiza, E. (1998) Maximum likelihood estimation of spatial covariance parameters. *Mathematical Geology*, 30, 95–107.

Webster, R., Oliver, M. A. (1992) Sample adequately to estimate variograms of soil properties. *Journal of Soil Science*, 43, 177–192.